State-Aware Adversarial Training for Utterance-Level Dialogue Generation

Yi Huang, Xiaoting Wu, Wei Hu , Junlan Feng , Chao Deng
JIUTIAN Team, China Mobile Research
{huangyi,wuxiaoting,huweiyjy,fengjunlan,dengchao}@chinamobile.com

Abstract

Dialogue generation is a challenging problem because it not only requires us to model the context in a conversation but also to exploit it to generate a coherent and fluent utterance. This paper, aiming for a specific topic of this field, proposes an adversarial training based framework for utterance-level dialogue generation. Technically, we train an encoder-decoder generator simultaneously with a discriminative classifier that make the utterance approximate to the state-aware inputs. Experiments on MultiWoZ 2.0 and MultiWoZ 2.1 datasets show that our method achieves advanced improvements on both automatic and human evaluations, and on the effectiveness of our framework facing low-resource. We further explore the effect of fine-grained augmentations for downstream dialogue state tracking (DST) tasks. Experimental results demonstrate the high-quality data generated by our proposed framework improves the performance over state-of-the-art models.

1 Introduction

Task-oriented dialogue systems (Young et al., 2013; Williams et al., 2016; Wu et al., 2020; Su et al., 2021) are designed to assist user in completing daily tasks, which involve reasoning over multiple dialogue turns. User goals expressed during conversation are important for the dialogue system and often encoded as a compact set of dialogue states, which is often expressed as a collection of slot-value pairs.

Nowadays generative conversational models are drawing an increasing amount of interest and becoming a more popular trend of task-oriented dialogue generation. Most existing generative conversational models (Shang et al., 2015; Vinyals and Le, 2015; Li et al., 2016; Yao, 2015; Luan et al., 2016; Zhang et al., 2019b) predict the next dialogue utterance given the dialogue history using the maximum likelihood estimation (MLE) objective,

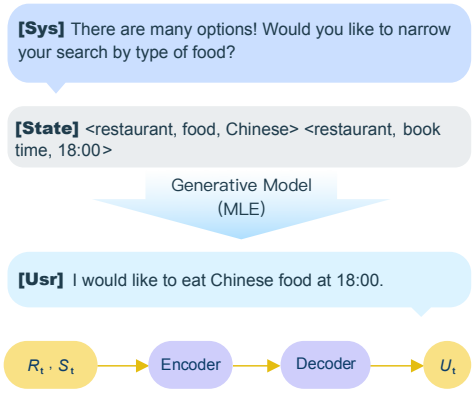


Figure 1: Dialogue generation via MLE training.

considering conversation history to learn to generate responses via optimizing the query-response pairs, as illustrated in Figure 1. Despite its success, this over-simplified training objective leads to problems: when generating dialogue responses from these models by iteratively sampling the next token, we do not have much control over attributes of the output text, such as the topic, the style, the sentiment, etc.

Solutions to these problems require answering a fundamental question: how to steer a powerful unconditioned dialogue model to generate content with desired attributes? Some existing studies have tackled this problem to control responses by using extended labels, however, these models still had some limitations (Wen et al., 2015; Li et al., 2016; Zhao et al., 2017; Huang et al., 2018; Zhou et al., 2018). One crucial issue was that they do not have explicit dialogue state guiding to guarantee that a controllable generation has a discriminability for a given condition.

Inspired by the success of adversarial training in computer vision (Denton et al., 2015) and natural language generation (Li et al., 2017), we delve into the challenge and propose our approach for state-aware dialogue generation with adversarial

Proceedings of the Towards Semi-Supervised and Reinforced Task-Oriented Dialog Systems (SereTOD), pages 62 - 74

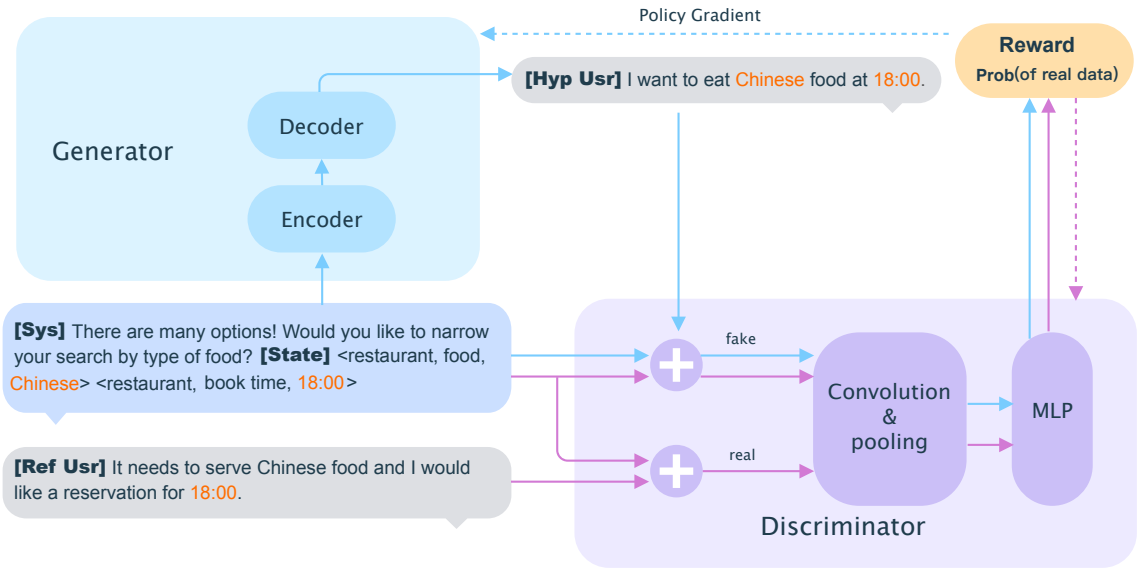


Figure 2: An overview of state-aware adversarial training. Different flow directions are marked with obviously distinguished arrows, blue and purple represent the training process of generator and discriminator, respectively. There are two cycles. The blue cycle is for generator learning, updating the model parameters of generator. The purple cycle is for discriminator learning, updating the discriminator model of periodic epoch. The learning of generator and discriminator is conducted in an alternate manner. Best viewed in color.

training. We focus on controlling the utterances by using dialogue state labels as conditions. We extend a framework of the generative adversarial network (Yu et al., 2017) for the task of generating conditional utterances on the basis of actual dialogue state constraints, alternatively training between a generator and a discriminator. The experimental results show that our proposed method has higher controllability for state-aware dialogue even though it has higher or comparable naturalness to existing methods, and improves the discriminability of generation. Furthermore, we investigate the effectiveness of our approach via downstream dialogue state tracking (DST) tasks. Experimental results demonstrate the high-quality data generated by our proposed framework improves the performance over state-of-the-art models.

The contributions of this paper are summarized as follows:

- We propose a novel adversarial training based framework for utterance-level dialogue generation, which generates more coherence and fluency utterances.
- For the downstream DST task, the high-quality data generated by our proposed framework improves the performance over state-of-the-art models.
- To our best knowledge, this is the first study of state-aware utterance generation via adversarial training with promising results.

2 Approach

In this section, we introduce the utterance-level dialogue generation of adversarial training. As shown in Figure 2, our framework consists of two main components: a generator and a discriminator. Different from the traditional generative dialogue model trained by MLE, we view the process of utterance generation as a sequence of actions that are taken according to a policy defined by the generator here. It generates controllable utterances based on input conditions, and the discriminator judges the quality of the utterances generated by the generator, feeding the reward back to the generator through policy gradient. The learning of generator and discriminator is carried out alternately.

2.1 Task Formulation

Let's denote a sequence of dialogue turns as a matrix $X_T = [R_1, U_1, \ldots, R_T, U_T]$, where U is the user utterance, R represents the system response and T denotes the number of turns. At each turn, user's goal can be regarded as a certain number

of domain-slot-value pairs (e.g., *(restaurant-area, west)*). The dialogue state tracking task is to track the value for each slot over X_t ($1 \leq t \leq T$). Belief states can be considered at two granularities: turn-level (S_t) and dialogue-level (B_t). S_t denotes the information introduced in the *t-th* turn and B_t represents the accumulated information from the first turn to the *t-th* turn. The task we focus on is to generate a user utterance U_t conditioned on the turn-level dialogue state S_t and corresponding system response R_t.

2.2 State-Aware Adversarial Training

To generate more human-like user utterances, we propose using adversarial training for generation: the generator is guided by the discriminator to produce utterances that are indistinguishable from the original dialogues and consistent with the belief state condition. The discriminator is trained on the dataset consisting of the utterances of original dialogues and the utterances generated by the generator. The learning of generator and discriminator is conducted in an alternate manner, which is detailed in Algorithm 1.

Generator

The generator *G* defines the policy that generates a user utterance U_t from a given dialogue history R_t and a turn-level user goal S_t. It takes a form similar to SEQ2SEQ models, which consists of an encoder and a decoder. In this paper, the GRU-based and the T5-based generators are employed to approximate $P(U_t|R_t, S_t)$, where the concatenation of R_t and S_t is used as input to the encoder and U_t is set to be the target sequence to be generated by the decoder.

Discriminator

The discriminator *D* is a binary classifier that aims to determine whether the user utterance is generated or from the original dataset. In order to make sense of belief state condition, the concatenation of turn-level belief state and user utterance is used as input to the discriminator. We follow the setting in SeqGAN to have CNN as the backbone model for the discriminator. First, the input sequence is represented as $[U_t] \bigoplus [S_t]$, where each token is represented as a k-dimensional token embedding and $\bigoplus$ is the concatenation operator to build the input matrix. Second, a kernel applies a convolutional operation to a window size of words to produce a new feature map and a max-over-time pooling operation works. Finally the output vector of a fully connected layer is fed to a 2-class sigmoid activation, returning the probability of the input utterance generated by generator or come from the original dialogue.

Algorithm 1 State-aware adversarial training

Input: A dialogue dataset C ={ R, U, S }.
Output: The parameters θ of G; The parameters ϕ of D.

```
Randomly initialize θ and φ;
Pre-train G using cross-entropy loss on C;
Generate user utterances using the pre-trained G;
Pre-train D using generated user utterances as negative samples and utterances from original dialogue as positive samples;
for each epoch do
    for each generator step do
        Generate a user utterance U'_{1:L}=(u'_1,...,u'_L) using the current G, where L denotes the number of tokens;
        for t in 1 : L do
            Compute R_{u'_l} by Eq. (1);
        end for
        Update θ according to Eq. (3);
    end for
    for each discriminator step do
        Sample <R, U, S> from the dataset C;
        Concatenate S and U as a positive sample;
        Generate U' using the current G;
        Concatenate S and U' as a negative sample;
        Update φ according to Eq. (4);
    end for
end for
return θ and φ;
```

Adversarial Training

We cast the state-aware utterance generation as a reinforcement learning problem that back-propagate the error computed by the discriminator to the generator via the policy gradient algorithm. The generator can be seen as an agent whose parameters θ define a policy. At each time step, it takes an action by generating a token and gets a reward value from the discriminator by employing Monte-Carlo search. The estimated probability of being real by *D* is used to calculate the reward:

$$R_{u_l} = D_\phi(U_{1:l}|S), \tag{1}$$

where u_l is the *l-th* token in U, R_{u_l} represents the reward of token u_l and ϕ is the parameters of D.

The goal of the generator is minimize the negative expected reward of generated utterance using the REINFORCE algorithm (Williams, 1992):

$$J_G(\theta) = -E_{U\sim G}[D_\phi(U|S)], \tag{2}$$

where $U \sim G$ represents the utterance U is generated from G and θ is the parameters of G.

With the likelihood ratio trick (Williams, 1992), the gradient of θ can be derived as:

$$\begin{aligned}\nabla J_G(\theta) &= -E_{U\sim G}[D_\phi(U|S)] \cdot \nabla \log G_\theta(U|S) \\ &\approx -D_\phi(U|S) \cdot \nabla \log G_\theta(U|S),\end{aligned} \tag{3}$$

The goal of the discriminator is to distinguish whether a user utterance is from original dialogue or generated by the generator. It computes the probability that the user utterance is from original dialogue given the turn-level belief state. Therefore, its objective function is to minimize classification error rate:

$$\begin{aligned}\min_\phi & -E_{U\sim ground-truth} \log D_\phi(U|S) \\ & -E_{U\sim G} \log(1 - D_\phi(U|S)),\end{aligned} \tag{4}$$

where $D_\phi(U|S)$ is the probability of U that it comes from original dialogue, $U \sim ground-truth$ represents the utterance U is from the golden label.

3 Experiments and Analysis

3.1 Dataset

We take MultiWOZ 2.0 and MultiWOZ 2.1 as datasets for the experiments. MultiWOZ[1] series dataset is a fully-labeled collection of human-human written conversations spanning over multiple domains and topics. It contains 8438 multi-turn dialogues with on average 13.7 turns per dialogue. It has 30 *(domain, slot)* pairs and over 4,500 slot values. Compared to MultiWOZ 2.0, MultiWOZ 2.1 has fixed the noisy state annotations and combined user dialogue acts as well as multiple slot descriptions per dialogue state slot into the new version. To date, these two datasets are recognized as the most widely used benchmark datasets in the field of dialogue systems.

[1]https://github.com/budzianowski/multiwoz/tree/master/data

3.2 Implementation Details

For a fair comparison, we introduce tow instantiations for the proposed framework, denoted as GRU-based and T5-based, respectively.

GRU-based: The generator is an encoder-decoder text generation model consists of simple GRU network, and the network structure of the discriminator is CNN. The optimizer for the generator and discriminator is Adam (Kingma and Ba, 2014). The learning rates are 1e-3 and 1e-4 respectively. In the adversarial training phase, the parameters of the 5 epoch discriminators are updated after each update of the parameters of the generator.

T5-based: The generator is an encoder-decoder implementation on the basis of T5, which is a pre-trained model composed of transformers, and the network structure of the discriminator is CNN. The optimizer for generator and discriminator is AdamW (Loshchilov and Hutter, 2018). The learning rates are 2e-5 and 5e-5, respectively. In the adversarial training phase, the parameters of the 4 epoch discriminators are updated after each update of the parameters of the generator.

We implement all the benchmarks using Pytorch on servers equipped with Nvidia Tesla V100 GPUs, each with 32GB memory. Source codes of our work in this paper will be open-sourced on Github as soon as we clean our code.

3.3 Main Results and Evaluation

Automatic Evaluation

We measure the quality of generated utterances by BLEU scores (Papineni et al., 2002) and BERT-score (Zhang et al., 2019a). In this experiment, only the utterances of each turn of original dialogues are used as reference sentences for the calculation of BLEU instead of the entire dataset as reference sentences. This is because the generated utterance from the dialogue model only need to be relevant to the turn-level state and input utterance, not the full dataset.

Tables 1 are experiments on the full dataset. GRU-based and T5-based represent the results of training with MLE, +GAN represents the results of using adversarial training (ADV). From Table 1 we can see our adapted models surpass original MLE up to 1.09% in BLEU-5, indicating the effectiveness of the added adversarial training process. GRU-based+GAN and T5-based+GAN exceed corresponding MLE-baselines with the same trending, respectively. Based on our proposed framework,

Model	MutilWOZ2.0					MutilWOZ2.1				
	BLEU-2	BLEU-3	BLEU-4	BLEU-5	BERT-Score	BLEU-2	BLEU-3	BLEU-4	BLEU-5	BERT-Score
GRU-based	23.41%	16.43%	11.82%	**8.70%**	88.33%	25.30%	17.82%	12.54%	8.99%	88.63%
+GAN	**24.37%**	**17.03%**	**12.15%**	8.66%	**88.35%**	**26.38%**	**18.98%**	**13.64%**	**10.08%**	**88.83%**
T5-based	25.43%	19.55%	15.39%	12.35%	89.17%	25.92%	19.95%	15.70%	12.58%	90.11%
+GAN	**25.46%**	**19.58%**	**15.42%**	**12.39%**	**89.18%**	**26.31%**	**20.23%**	**15.87%**	**12.65%**	**90.12%**

Table 1: Automatic evaluation of two models trained by MLE and adversarial training on full datasets.

Model	MutilWOZ2.0					MutilWOZ2.1				
	BLEU-2	BLEU-3	BLEU-4	BLEU-5	BERT-Score	BLEU-2	BLEU-3	BLEU-4	BLEU-5	BERT-Score
GRU-based	17.20%	11.11%	7.52%	5.18%	87.09%	17.20%	11.11%	7.52%	5.18%	87.09%
+GAN	**17.77%**	**11.69%**	**7.95%**	**5.46%**	**87.28%**	**17.77%**	**11.69%**	**7.95%**	**5.46%**	**87.28%**
T5-based	21.43%	15.39%	11.42%	8.66%	88.11%	23.71%	17.48%	13.00%	9.87%	88.53%
+GAN	**21.45%**	**15.67%**	**11.76%**	**8.92%**	**88.16%**	**23.98%**	**17.72%**	**13.22%**	**10.03%**	**88.55%**

Table 2: Automatic evaluation of two models trained by MLE and adversarial training in low-resource scenario.

Table 1 shows that the effectiveness of ADV is consistent in two datasets of different metrics.

In order to further explore the performance of our framework in the case of low-resource scenario, 100 instances of full dialogues are randomly selected from the training dataset, and 50 instances of complete dialogues are randomly selected from the validation dataset. Tables 2 shows the performance of two models on MultiWOZ 2.0 and MultiWOZ 2.1 under low-resource settings. Predictably, various degrees of performance degradation occurs, especially on GRU-based model. On the other hand, the improvement under the same setting demonstrates the effectiveness of our framework facing low-resource.

Combining the experimental results of the above different settings, it can be observed that both the BLEU score and BERT-score of the results after adversarial training are better than MLE training.

Human Evaluation

We evaluate the generated data from two perspectives: *statement fluency* and turn-level *belief state correctness*. The *statement fluency* indicates whether the generated sentence is fluent and human-likely. The turn-level *belief state correctness* evaluates whether $<R_t, U_t'>$ is consistent with S_t'.

There are two corresponding evaluation metrics, *Sentence fluency* and *Slot accuracy*. (1) *Sentence fluency* represents whether the generated sentence conforms to the natural expression of human beings and is suitable as an answer to a question. (2) *Slot accuracy* represents whether the generated utterance contains the dialogue state of the input utterance.

Randomly select 100 instances generated by the models and invite 3 experts to evaluate the data for human evaluation. Table 3 shows human evaluation results for naturalness and controllability of GRU-based generator in MultiWOZ 2.0 dataset. Regarding the naturalness, models used adversarial learning produced a more acceptable utterance to the dialogue context. At the same time, the adversarial-explicit model achieved the best performance among the compared ones in terms of the controllability. A same trend occurs in the evaluation of T5-based one: the results show that ADV outperforms the MLE on almost all metrics and even strengthens the performance of human reference. Notably, the performance of T5-based generator even outperforms the results of the original data (corresponding **USER** line) in Table 4. In the original MultiWOZ dataset, there are label-missing errors for dialogue states. Specifically, the turn-level belief state is not reflected in the user utterance fully. Through data inspection, we find

	Sentence Fluency		Slot accuracy
	Mean score (1-5)	$\geq 3(\%)$	
USER	4.59	96.30%	80.70%
MLE	4.00	87.70%	53.30%
ADV	**4.16**	**92.00%**	**64.00%**

Table 3: Human evaluation of GRU-based generator. Sentences are scored on a scale of 1 to 5. The average value represents the average score, and $\geq 3(\%)$ represents the proportion of the sentence evaluation score greater than or equal to three points in all sentences.

	Sentence Fluency		Slot accuracy
	Mean score (1-5)	$\geq 3(\%)$	
USER	4.78	100.00%	71.23%
MLE	4.70	98.67%	76.50%
ADV	**4.81**	98.67%	**79.73%**

Table 4: Human evaluation of T5-based generator.

that the T5-based generator corrects the errors in the original dataset. This situation shows that the model can well use the belief state as a condition to generate the corresponding user utterance.

Experimental results demonstrate that our approach produces more interactive, relevant, and fluent utterances than standard SEQ2SEQ models trained using the MLE objective function. Beyond this, evaluation details for automatic and human ways are shown in the appendix.

3.4 Downstream Results

In this section, we conduct experiments on a suite of downstream DST tasks and present the results of applying utterance-level dialogue generation on DST data augmentation. The learning of dialogue state tracker is detailed in Algorithm 2. Here data augmentation is to generate a new user utterance U_t' conditioned on a modified S_t' derived from original turn-level belief state S_t. The modification strategy uses the value substitution method (Li et al., 2021). To overcome the de-generation and over-generation phenomenons, a data filter F is employed to filtering the generated candidates (Li et al., 2021). Then a novel sequence of dialogue turns X_t'=[R_1,U_1,...,R_t,U_t'] is formed by replacing the original user utterance U_t with U_t', and B_t' which is induced by B_t based on the difference between S_t and S_t' is the dialogue-level belief states of X_t'. We use the resulting set of <X_t',B_t'> to do DST data augmentation. In the following, two known typical DST models are selected for further experiments.

- **TRADE**: TRAnsferable Dialogue statE generator (TRADE) (Wu et al., 2019) generates dialogue states from utterances using a copy mechanism, facilitating knowledge transfer between domains. The prominent difference from previous one-domain DST models is that TRADE is based on a generation approach instead of a close-set classification approach.

- **TripPy**: TripPy (Heck et al., 2020) presents a new SOTA approach which makes use of various copy mechanisms to fill slots with values to avoid the use of value picklists altogether. This model has no need to maintain a list of candidate values. Instead, all values are extracted from the dialogue context on-the-fly.

We train each DST model on the mixing of MultiWOZ 2.0 training data and augmented data.

Algorithm 2 The DST data augmentation

Input: A dialogue dataset C, the randomly initialized Generator G, data filter F, belief state modification strategy π, the dialogue state tracker with parameters ρ.
Output: Trained tracker.

1: Train G using Algorithm 1 on dataset C;
2: Train F with cross-entropy loss on dataset C;
3: Modify turn-level belief state from S_t to $S_t^{'}$ according to π;
4: Obtain new data C' according to $S_t^{'}$ by the trained G;
5: Obtain new data $C_F^{'}$ by filtering de-generation and over-generation data;
6: Train the tracker on C+$C_F^{'}$, update ρ;
7: Save the tracker with ρ;

Trained DST models are then evaluated on original test set. Results are shown in Table 5. Both training with augmentation of TRADE and TripPy outperform the original models. The joint goal accuracy improves to 49.76% and 59.09% respectively, and TripPy reaches the state-of-the-art performance on MultiWOZ 2.0 dateset. This shows the effectiveness of our adversarial utterance-level dialogue generation for DST data augmentation.

Model	Joint Goal Accuracy(%)
TRADE	48.62
+ Augmentation	**49.76**
TripPy	58.01
+ Augmentation	**59.09**

Table 5: Joint goal accuracy results on MultiWOZ 2.0.

3.5 Analysis and Discussion

Case study from Table 6 has displayed three typical errors in our experiments of dialogue generation tasks: (1) The slot value is wrong. (2) The slot is missing. (3) The slot type is wrong. Specifically, **System+State** stands for inputs, and **USER** represents references. Last two columns are the output results of MLE training and ADV training. The results show that the above-mentioned errors can be alleviated via our proposed framework, and the model after adversarial training can better adapt to the input conditions in the dialogue, achieving more appropriate utterance-level dialogue generation.

System+State	USER	MLE	ADV
system: tr0275 will leave at 13:29 and arrive by 14:07 . would you like to buy tickets ? **state: train book people 3**	yes , could you make a booking **for 3 people** ?	yes , i 'll need the train for 7 people . can you book the reference number ?	yes , i need book for **3 people** please .
system: i love the fez club , their phone number is 01223519224 .can i help you with anything else today ? **state: train destination ely , train day Thursday**	i would love some information on a train to get me to **ely** on **thursday** .	i also i also need a train to train leaving on **thursday**. (miss ely)	i also need a train to **ely** to leave on **thursday**.
system: would you like me to book it for you ? **state: train book people 7**	yes , for **7 people** .	yes , please . i need 7 tickets .	yes , please book it for **7 people** .

Table 6: Case study. **System+State** stands for inputs, and **USER** represents references. Last two columns are the output results of **MLE** and **ADV**. The black-lined font is the dialogue state of the current turn. We expect the generated utterances exactly correspond to the inputs. The red font are typical errors generated by MLE model.

Model	Confusion matrix			Acc(%)
MLE	Pre. / Real	**N**	**P**	97.12
	N	594	32	
	P	4	622	
ADV	Pre. / Real	**N**	**P**	**81.39**
	N	397	229	
	P	4	622	

Table 7: Confusion matrix for MLE and ADV. The results show that the pre-trained discriminator will misjudge the fake text after adversarial training.

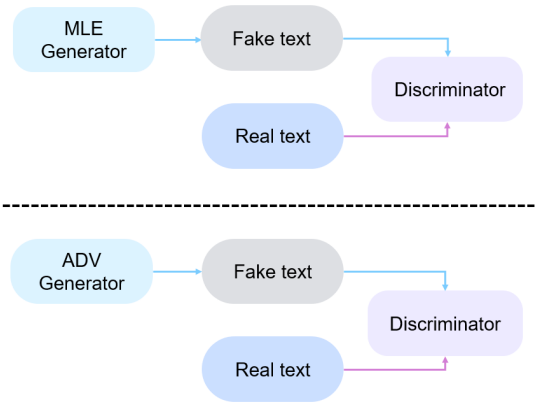
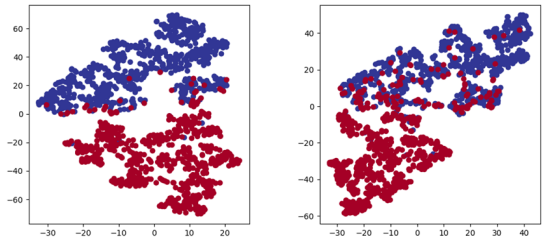


Figure 3: Contrast experiment for the quality judgment of generated utterances by two models.

To analyze the results quantitatively, we verify the effectiveness of utilizing adversarial training under the control variable method, that is, a same pre-trained discriminator is applied for both generators. We use the pre-trained discriminator to evaluate the utterances generated by the model of adversarial training (real text) and the utterances of original dialogues (fake text) as shown in Figure 3.

Table 7 shows the confusion matrix of predicting results via discriminator's classification, where negative samples (**N**) represent utterances generated by the generator and positive samples (**P**) represent utterances of original dialogues. The accuracy (Acc) is calculated as follow:



(a) Visualization of MLE. (b) Visualization of ADV.

Figure 4: Feature visualization of generative spaces for two models' comparison.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (5)$$

where TP and TN represent the number of correct predictions for real text and generated text, respectively. FP and FN represent the number of incorrect predictions for real text and generated text, respectively.

It can be seen the text generated by adversarial training makes it more difficult for the pre-trained discriminator to distinguish the authenticity of the inputs. The results further support this point of view can be seen in the confusion matrix. Comparing MLE model and ADV model, the discriminator's judgment result for real samples maintains, but the judgment on generated text differs. It can be confirmed the accuracy's drop of the discriminator is affected by the decline in the generator's ability to judge more realistic generated samples.

In order to present the classification results more intuitively, dimensionality reduction of the features after the convolutional layer in the discriminator is visualized using the t-SNE algorithm (Van der Maaten and Hinton, 2008). The visual features are shown in the Figure 4. The red dots represent the real text, and the blue dots represent the generated text. Comparing (a) and (b), it can be observed that boundary becomes unrecognizable and overlapping after adversarial training, which adds a layer of complexity to the discriminator and brings new challenges.

4 Related Work

The idea of generative adversarial networks (Goodfellow et al., 2014) has enjoyed great success in computer vision (Radford et al., 2015; Chen et al., 2016; Brock et al., 2018; Karras et al., 2020). Training is formalized as a game in which the generative model is trained to generate outputs to fool the discriminator; the technique has been successfully applied to image generation. However, to the best of our knowledge, this idea has not achieved comparable success in NLP. This is due to the fact that unlike in vision, text generation is discrete, which makes the error outputted from the discriminator hard to back-propagate to the generator. Some recent work has begun to address this issue: Lamb et al. (2016) propose providing the discriminator with the intermediate hidden vectors of the generator rather than its sequence outputs. Such a strategy makes the system differentiable and achieves promising results in tasks like character-level language modeling and handwriting generation. Yu et al. (2017) use policy gradient reinforcement learning to back-propagate the error from the discriminator, showing improvement in multiple generation tasks such as poem generation, speech language generation and music generation. Outside of sequence generation, Chen et al. (2018) apply the idea of adversarial training to sentiment analysis and Zhang et al. (2017) apply the idea to domain adaptation tasks. Cui et al. (2019) proposed Dual Adversarial Learning (DAL), which uses adversarial learning to mimic human judges and guides the system to generate natural responses. To improve the diversity of responses, Xu et al. (2018) proposed a Diversity-Promoting Generative Adversarial Network (DP-GAN). This method encourages the generation of highly diverse texts by assigning low rewards to repeated texts and high rewards to new texts, and a new discriminator structure is proposed to determine repeated texts.

Our work is related to recent work that formalizes sequence generation as an action-taking problem in reinforcement learning (Sutton and Barto, 2018). Ranzato et al. (2015) train RNN decoders in a SEQ2SEQ model using policy gradient to obtain competitive machine translation results. Bahdanau et al. (2016) take this a step further by training an actor-critic RL model for machine translation. Also related is recent work (Shen et al., 2015; Wiseman and Rush, 2016) to address the issues of exposure bias and loss evaluation mismatch in neural translation.

5 Conclusion

In this paper, we address the difficulty of utterance-level dialogue generation by proposing an adver-

sarial training based framework that can generate high-quality data to improve the downstream DST performance. Specifically, our method leverages an encoder-decoder framework in terms of an adversarial training paradigm, while taking advantage of dialogue state-aware semantic representation from the reinforced generator to construct the discriminator. The two-stage training process delivers more adversarial-balance for both after iterative interactions. Experimental results on MultiWoZ 2.0 and MultiWoZ 2.1 datasets demonstrate that the proposed framework significantly improves the performance over the state-of-the-art models. Future work includes more exploration into the design of generator-discriminator architect and improvement of more dialogue tasks.

Limitations

Our work pioneers in the adversarial training based framework for utterance-level dialogue generation, which trains an encoder-decoder generator simultaneously with a discriminative classifier that make the utterance approximate to the state-aware inputs. However, our paper may have following omissions and inadequacies.

- Our focused task is limited in turn-level belief state. DST of dialogue-level is beyond the scope of this article. We believe this situation will meet new challenges and we will explore more in the next work.
- The policy gradient reinforcement learning algorithm is used to optimizing the generator during adversarial training process, which slows down the training speed of T5-based generator.
- Though we list case study in our paper, we believe it needs more rethinking and comparison work into the internal mechanism in the future.

References

Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2016. An actor-critic algorithm for sequence prediction. *arXiv preprint arXiv:1607.07086*.

Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in neural information processing systems*, 29.

Xilun Chen, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. 2018. Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6:557–570.

Shaobo Cui, Rongzhong Lian, Di Jiang, Yuanfeng Song, Siqi Bao, and Yong Jiang. 2019. Dal: Dual adversarial learning for dialogue generation. *arXiv preprint arXiv:1906.09556*.

Emily L Denton, Soumith Chintala, Rob Fergus, et al. 2015. Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems*, 28.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Michael Heck, Carel van Niekerk, Nurul Lubis, Christian Geishauser, Hsien-Chin Lin, Marco Moresi, and Milica Gašić. 2020. Trippy: A triple copy strategy for value independent neural dialog state tracking. *arXiv preprint arXiv:2005.02877*.

Chenyang Huang, Osmar R Zaiane, Amine Trabelsi, and Nouha Dziri. 2018. Automatic dialogue generation with expressed emotions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 49–54.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Alex M Lamb, Anirudh Goyal ALIAS PARTH GOYAL, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. 2016. Professor forcing: A new algorithm for training recurrent networks. *Advances in neural information processing systems*, 29.

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*.

Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*.

Shiyang Li, Semih Yavuz, Kazuma Hashimoto, Jia Li, Tong Niu, Nazneen Rajani, Xifeng Yan, Yingbo Zhou, and Caiming Xiong. 2021. Coco: Controllable counterfactuals for evaluating dialogue state trackers. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.

Ilya Loshchilov and Frank Hutter. 2018. Fixing weight decay regularization in adam.

Yi Luan, Yangfeng Ji, and Mari Ostendorf. 2016. Lstm based conversation models. *arXiv preprint arXiv:1603.09457*.

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, Luke Metz, and Soumith Chintala. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2015. Minimum risk training for neural machine translation. *arXiv preprint arXiv:1512.02433*.

Yixuan Su, Lei Shu, Elman Mansimov, Arshit Gupta, Deng Cai, Yi-An Lai, and Yi Zhang. 2021. Multi-task pre-training for plug-and-play task-oriented dialogue system. *arXiv preprint arXiv:2109.14739*.

Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. 2015. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv preprint arXiv:1508.01745*.

Jason D Williams, Antoine Raux, and Matthew Henderson. 2016. The dialog state tracking challenge series: A review. *Dialogue & Discourse*, 7(3):4–33.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256.

Sam Wiseman and Alexander M Rush. 2016. Sequence-to-sequence learning as beam-search optimization. *arXiv preprint arXiv:1606.02960*.

Chien-Sheng Wu, Steven Hoi, Richard Socher, and Caiming Xiong. 2020. Tod-bert: Pre-trained natural language understanding for task-oriented dialogue. *arXiv preprint arXiv:2004.06871*.

Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. 2019. Transferable multi-domain state generator for task-oriented dialogue systems. *arXiv preprint arXiv:1905.08743*.

Jingjing Xu, Xuancheng Ren, Junyang Lin, and Xu Sun. 2018. Diversity-promoting gan: A cross-entropy based generative adversarial network for diversified text generation. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3940–3949.

Yao Yao. 2015. A review of corpus-based statistical models of language variation. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 11–15.

Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proceedings of the IEEE*, 101(5):1160–1179.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019a. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

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

Yuan Zhang, Regina Barzilay, and Tommi Jaakkola. 2017. Aspect-augmented adversarial networks for domain adaptation. *Transactions of the Association for Computational Linguistics*, 5:515–528.

Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. *arXiv preprint arXiv:1703.10960*.

Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

A BERT-Scroe

Experimental Setup : We randomly selected 100 samples from the data set to obtain the results of MLE and ADV. Furthermore, respectively calculate samples with USER's text the similarity degree. After 100 Bert-Scores were calculated, the average was calculated.The result are reported below:

Method	BERT-score
MLE	89.18%
ADV	**89.67%**

Table 8: Random sampling of 100 samples of BERT-score results

In order to prevent uneven distribution, the test data were divided into 10 groups and the BERT-score of the mean MLE model and ADV model are calculated respectively. The main results are shown below:

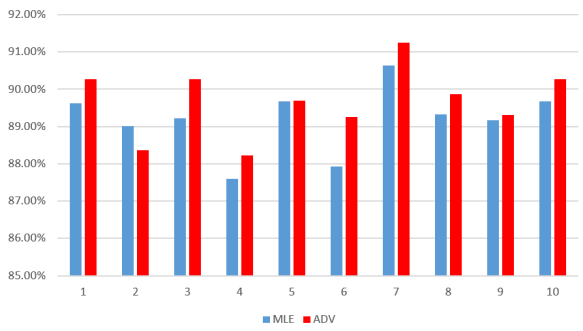


Figure 5: An overview of bert-score.

According to the Figure 5, the average Bert-score of group 2 is higher than that of ADV, and the similarity of the texts generated by ADV model is higher than that generated by MLE model in the other 9 groups, which proves the effectiveness of the algorithm.

B Human evaluation

Table 9 and Table 10 represent the experimental details statement fluency, and the conditions of slot value, contained for the human evaluation of the GRU-based model, respectively. Table 11 and Table 12 represent the same experimental details for the T5-based model, respectively.

Statement Fluency						
	USER		MLE		ADV	
	average value	≥3(%)	average value	≥3(%)	average value	≥3(%)
expert1	4.81	96%	4.12	81%	4.49	87%
expert2	4.38	97%	3.79	86%	3.92	94%
expert3	4.57	96%	4.09	96%	4.06	95%
average	4.59	96.30%	4	87.70%	**4.16**	**92%**

Table 9: Statement fluency experiment with GRU model.A total of 3 experts participated in the evaluation. Sentences are scored on a scale of 1 to 5. The average value represents the average score, and ≥3(%) represents the proportion of the sentence evaluation score greater than or equal to three points in all sentences

The conditions of slot value contained						
	USER		MLE		ADV	
	Contain /Part /No	Acc	Contain /Part /No	Acc	Contain /Part /No	Acc
expert1	79/12/9	79%	53/28/19	53%	64/19/17	64%
expert2	82/11/7	82%	53/26/21	53%	66/20/14	66%
expert3	81/13/6	81%	54/31/15	54%	62/23/15	62%
average	-	80.70%	-	53.30%	-	**64%**

Table 10: The conditions of slot value contained with GRU-based model.Contain, Part, and No respectively represent whether the answer of the dialogue is fully contained, partially contained, or not containing the dialogue state. Accuracy is only calculated for cases where the dialogue state is completely contained.

Statement Fluency						
	USER		MLE		ADV	
	average value	≥3(%)	average value	≥3(%)	average value	≥3(%)
expert1	4.84	100%	4.49	98%	4.71	98%
expert2	4.8	100%	4.71	98%	4.75	98%
expert3	4.71	100%	4.9	100%	4.98	100%
average	4.78	100.00%	4.70	98.67%	**4.81**	**98.67%**

Table 11: Statement fluency experiment with T5-based model. A total of 3 experts participated in the evaluation. Sentences are scored on a scale of 1 to 5. The average value represents the average score, and ≥3(%) represents the proportion of the sentence evaluation score greater than or equal to three points in all sentences.

The conditions of slot value contained						
	USER		MLE		ADV	
	Contain /Part /No	Acc	Contain /Part /No	Acc	Contain /Part /No	Acc
expert1	36/12/3	70.6%	39/8/4	76.5%	40/7/4	78.4%
expert2	37/10/4	72.5%	39/8/4	76.5%	41/6/4	80.4%
expert3	36/12/3	70.6%	39/9/3	76.5%	41/6/4	80.4%
average	–	71.23%	–	76.5%	–	**79.73%**

Table 12: The conditions of slot value contained with T5-based model.Contain, Part, and No respectively represent whether the answer of the dialogue is fully contained, partially contained, or not containing the dialogue state. Accuracy is only calculated for cases where the dialogue state is completely contained.