Posterior-GAN: Towards Informative and Coherent Response Generation with Posterior Generative Adversarial Network

Shaoxiong Feng,† Hongshen Chen,‡ Kan Li,† Dawei Yin‡
† School of Computer Science & Technology, Beijing Institute of Technology
‡ Data Science Lab, JD.com
{shaoxiongfeng,likan}@bit.edu.cn, ac@chenhongshen.com, yindawei@acm.org

Abstract
Neural conversational models learn to generate responses by taking into account the dialog history. These models are typically optimized over the query-response pairs with a maximum likelihood estimation objective. However, the query-response tuples are naturally loosely coupled, and there exist multiple responses that can respond to a given query, which leads the conversational model learning burdensome. Besides, the general dull response problem is even worsened when the model is confronted with meaningless response training instances. Intuitively, a high-quality response not only responds to the given query but also links up to the future conversations. To facilitate the modeling of these triples, we further propose a novel encoder-decoder based generative adversarial learning framework, Posterior Generative Adversarial Network (Posterior-GAN), which consists of a forward and a backward generative discriminator to cooperatively encourage the generated response to be informative and coherent by two complementary assessment perspectives. Experimental results demonstrate that our method effectively boosts the informativeness and coherence of the generated response on both automatic and human evaluation, which verifies the advantages of considering two assessment perspectives.

Introduction
Generative conversational models are drawing an increasing amount of interests (Shang, Lu, and Li 2015; Vinyals and Le 2015; Serban et al. 2016; Serban et al. 2017b; Serban et al. 2017a; Li et al. 2016a; Li et al. 2016b; Li et al. 2017a; Mou et al. 2016; Zhao, Zhao, and Eskenazi 2017; Xing et al. 2017; Xu et al. 2017; Xu et al. 2018a; Zhang et al. 2018a; Zhang et al. 2018b). Most existing generative conversational models are based on a Seq2Seq architecture (Sutskever, Vinyals, and Le 2014). These models consider conversation history to learn to generate responses and are optimized over the query-response pairs.

However, the query-response tuples are naturally loosely coupled, there exist multiple responses that can respond to a given query, so call one-to-many phenomenon, which leads the conversational model learning burdensome. In DailyDialog (Li et al. 2017b) corpus, at least 13% utterances contain more than one response (Csaky, Purgai, and Recski 2019). What’s more, the notorious general dull response problem is even worsened when the model is confronted with meaningless response training instances. In another public available corpus OpenSubtitles (OSDb), 113K sentences contain the sequence “I don’t know” in the training set (Li et al. 2016a). Not to mention other similar meaningless responses like “haha”, “what are you talking about?”, etc. The one-to-many phenomenon and non-negligible proportion of generic responses in the training corpus cause the neural response generation model prone to generate short, bland, or even irrelevant responses. In Figure 1 for the given query talking about the traveling experience, the first response is much shorter and uninformative compared with the other two responses, while the second one seems to be informative enough but meanders away from the conversational subject, especially in terms of the following conversations. The third response is not only informative but also coherent with both the query and the next utterance.

It is often the case that a high-quality response not only responds to the given query but also links up to the future conversations, in this paper, we propose to utilize query-response-future turn triples instead of the query-response pairs to train the response generation model. Convention-
ally, neural dialogue generation model is optimized with a maximum likelihood estimation (MLE) objective given the query-response tuples. However, such an objective is obviously inadequate for triple learning, where the future conversation is introduced during training. Moreover, the MLE objective encourages the model to repetitively overproduce high-frequency words in the ambiguous and noisy training corpus (Zhang et al. 2018a) and tend to deterministically output some “average” of diverse real-world responses (Csaky, Purgai, and Recski 2019). To extend the neural dialogue learning from tuples to triples, and induce the generated response to be not only informative but also coherent regarding both the input context and the future conversations, we further propose a novel encoder-decoder based generative adversarial learning framework, Posterior Generative Adversarial Network (Posterior-GAN), to handle the query-response-future turn triple modeling. The framework leverages a forward and a backward generative discriminator to guide the generated response: the forward discriminator that extracts as much sentence-level semantic information as possible from the response to predict the real-world future conversations outputs high rewards if the generated response is informative enough with respect to the subsequent future conversations, and the backward discriminator that assesses the response based on the full information of real-world future conversations instead of the input context encourages the generated response to be more coherent in terms of the following conversations and guides the conversation smoothly linking up to the future turn.

We highlight our contributions as follows:

- We identify an unexplored type of metadata, query-response-future turn triples, for response generation. Compared to general query-response tuples, the triples help the model use bidirectional information to learn the response generation in training.
- We propose a novel encoder-decoder based generative adversarial learning framework, Posterior-GAN, to facilitate the query-response-future turn modeling, which induces the generated response to be informative and coherent by constructing two generative discriminators, a forward one and a backward one respectively.
- We perform detailed experiments to demonstrate the effectiveness of the proposed framework and verifies the ability of bidirectional generative discriminators on assessing the quality of response.

**Method**

**Overview**

In this paper, we extend the conventional query-response tuple \((x, y)\) neural dialogue learning into query-response-future turn triple \((x, y, z)\) to encourage the generated response to be informative and coherent with respect to both the given query and the future conversations.

Here, a novel posterior generative adversarial network (Posterior-GAN) is proposed to undertake the triple learning and mitigate the overproduction of repetitive responses problem under the MLE objective. Posterior-GAN contains a generator that is responsible for generating response, and two discriminators cooperatively discriminating whether the generated response is coherent and informative in a forward and a backward manner by taking both the preceding context and the future conversations into account. The generator \(G\) is constructed upon the Seq2Seq structure. Given the input query \(x = (x_1, \ldots, x_t, \ldots, x_T)\) of \(T\) words from the vocabulary \(\Gamma\), the model generates response \(y = (y_1, \ldots, y_m, \ldots, y_{L})\) of \(L\) words. For the discriminator, instead of a traditional classification-based discriminator, we utilize two symmetric generative discriminators with cross-entropy based rewards: a forward generative discriminator \(D_{φ1}\) and a backward generative discriminator \(D_{φ2}\). The general architecture is illustrated in Figure 2.

**Generator**

In our implementation, the generator consists of a two-layer bidirectional LSTM encoder and a four-layer LSTM decoder. The word embedding is sequentially fed to two-layer bidirectional LSTM resulting with a hidden state representing the past and future information simultaneously. To better handle the long-range dependencies in multi-turn conversations, we also apply attention mechanism (Bahdanau, Cho, and Bengio 2015) in the decoding phase.

**Discriminator**

Traditional discriminators in generative adversarial networks are classification-based approach, such as a binary classifier, which takes in the query-response pair \((x, y)\) and recognizes the true probability of pair \((x, y)\) being true as a reward. Essentially, it models the joint probability \(p(x, y)\). However, as Xu et al. (2018a) illustrated, when a query-generated response pair fits the distribution of real-world pairs, the classifier-based discriminators may result in saturated similar indistinguishable rewards for both the synthesized response and the ground truth response.

As shown in Figure 3 in this paper, instead of modeling the joint probability \(p(x, y)\), we introduce the future conversations \(z\) and utilize the conditional probability \(p(z|y)\) and \(p(y|z)\) as rewards. The forward discriminator \(p(z|y)\) outputs high rewards if \(y\) is informative enough to perceive the subsequent future conversations, and a backward discriminator \(p(y|z)\) encourages the generated response to be more coherent in terms of the following conversations and generates high rewards if the generated response bridge the gap between the query \(x\) and the future conversation \(z\). Note that, in order to induce the generated response to be informative

![Figure 2: Illustration of Posterior-GAN. Brown for the query, yellow for the current response, and green for the future turn. \(G_{θ}\) represent generator. \(D_{φ1}\) and \(D_{φ2}\) represent forward and backward generative discriminator respectively.](image-url)
and coherent in terms of both the given query and the future conversations, and stabilize the adversarial training process (Li et al. 2017a; Wu et al. 2018), we also optimize the generator by teacher forcing periodically.

Forward Generative Discriminator Intuitively, in multi-turn conversations, a high-quality response not only responds to the query but also is informative enough to perceive future conversations. The forward generative discriminator takes in the response $y$ (the predicted response $y_{Gθ}$ or the real-world response $y_{true}$), and generates the future turn $z$, a sequence of $K$ words. It discriminates whether $y$ is informative and appropriate enough to induce the future turn.

In detail, for a response $y$ of $L$ words, the reward of generating the real-world future turn $z$ is defined as the averaged negative cross entropy of each word of $z$:

$$ R_1(y) = \frac{1}{K} \sum_{k=1}^{K} \log D_{φ1}(z_k | y, z_{<k}) . \quad (1) $$

We maximize the reward for real-world response $y_{true}$ for generating the future response $z$ and minimize the reward for the generated response $y_{Gθ}$ of predicting $z$. We expect the general, meaningless generated responses are of lower rewards while the informative responses are of higher rewards. The loss function of the forward generative discriminators is formulated as follows:

$$ J(φ1) = - \left( E_{y_{true} \sim p_{data}}[R_1(y_{true})] - E_{y_{Gθ} \sim Gθ}[R_1(y_{Gθ})] \right) . \quad (2) $$

In contrast, the reward of the existing classifier-based discriminators are calculated as follows:

$$ R(y) = D_θ \left( true \mid x, y \right) , $$

where $D_θ$ is a binary classifier judging how likely $(x, y)$ is from the real-world data. One major problem of the classifier-based discriminator is that the reward is easy to saturate, where for a given context, different generated responses usually achieve similar rewards from the saturated region of the non-linear classification function like sigmoid (Xu et al. 2018a). As a result, the discriminator fails to distinguish detailed fine-grained differences among the generated responses in such a situation. In forward generative discriminator, the response $y$ is differentiated by the ability of seeing the future few turns of conversations $z$. Such a cross-entropy based reward not only does not saturate but also discriminates the response in terms of $z$.

Backward Generative Discriminator To further induce the generated response to be more coherent with both the preceding and the following conversations, we propose a backward generative discriminator $D_{φ2}$.

Given the real-world future conversation $z$, the reward for $m^{th}$ word in response $y$ of $L$ word (the real-world response $y_{true}$, and the generated response $y_{Gθ}$) is calculated at the word level:

$$ R_2(y_m) = - \log D_{φ2}(y_m | z, y_{<m}) . \quad (3) $$

We maximize the reward $R_2$ for the real-world response $y_{true}$ and minimize the reward $R_2$ for the response $y_{Gθ}$ produced by generator $Gθ$. We formulate the loss function of the backward generative discriminator as follows:

$$ J(φ2) = - \left( E_{y_{true} \sim p_{data}}[R_2(y_{true})] - E_{y_{Gθ} \sim Gθ}[R_2(y_{Gθ})] \right) . \quad (4) $$

If a response $y$ matches well with the given context $x$, but is irrelevant with the following conversations $z$, in previous discriminators, it may be endowed with a high reward. Whereas the backward discriminator models the generative probability $p(y | z)$ given the future conversation $z$, it induces the generated response to be more coherent by bridging the gap between the preceding and the subsequent conversations. We expect the responses which are coherent with both the preceding and the following conversations gain higher rewards while the responses that are irrelevant with the subsequent turns achieve lower rewards.

Optimization

In this work, the policy gradient method (Sutton et al. 1999; Williams 1992) is employed for optimization. The generator (policy) is trained to maximize the cumulative total reward of generated response:

$$ J(θ) = E_{y_{Gθ} \sim Gθ} \left( Q_{θ_{Dθ1},θ_{Dθ2}}^G(x, y_{Gθ}, z) \right) , \quad (5) $$

where $Q_{θ_{Dθ1},θ_{Dθ2}}^G(x, y_{Gθ}, z)$ is the cumulative total reward for a generated response $y_{Gθ}$ starting from initial state $x$, taking action $a$ according to the policy $Gθ$. The gradient of Eq. (5) is approximated using the likelihood ratio trick (Williams 1992):

$$ \nabla_θ J(θ) \simeq \frac{\nabla_θ \log Gθ(y_{n|m} | x, y_{<m})}{\sum_{n=1}^{N} \sum_{m=1}^{L} R^n_m} , \quad (6) $$
where $N$ is the number of sampling via the policy $G_{θ}$, and $R_{n}^{m}$ is the final reward of $n^{th}$ word in response $y_{G_{θ}}^{m}$ by combining the reward $R_1$ and $R_2$ as $\sum_{i=m}^{L} X_i (R_1 (y_n) - MIN(R_1)) R_2(y_n)$. The term $γ$ is the discount rate. The $MIN(R_1)$ is defined as the minimum response reward of each batch in training samples.

Only using policy gradient methods to optimize the generator directly will lead to a very fragile training process (Li et al. 2017a), because the generator never has access to the real-world response throughout the training process. Thus, we adopt the following three strategies to promote and stabilize the training process.

**Curriculum Learning Strategy.** For an utterance, the first $T$ words are optimized by MLE and the rest uses the policy gradient to calculate the loss. Then policy gradient is gradually adopted at every word (Li et al. 2016b).

**Baseline Strategy.** It facilitates the training process to be more steady by encouraging the model generates responses that achieve higher rewards than the baseline and suppressing the response generation with lower rewards compared with the baseline. In practice, we calculate the average word rewards of each batch in training samples as the baseline. When we only use the forward generative discriminator to judge the generated response, the baseline is set to the average response rewards of every batch in training samples.

**1-value reward Strategy.** Following previous work (Li et al. 2016b; Li et al. 2017a; Xu et al. 2018a), we also utilize teacher forcing to train the generator periodically. In this work, teacher forcing forces the generator to keep the given query in mind. We use the maximum likelihood estimation (MLE) objective in the teacher forcing phase, which can be viewed as setting the reward of the real-word response to 1 when using the policy gradient.

### Experiment

#### Datasets

**DailyDialog:** This dataset consists of high-quality multi-turn dialog, which is provided by (Li et al. 2017b). We construct the query-response-future turn triples by treating each round in the dataset as response, three previous rounds as query, and three latter rounds as future turn. The length of response is limited to (5,40) by discarding the triples whose response is shorter than 5 words and truncating response over the maximum length to 40 words. The size of query and future turn is limited to less than 80 words. We randomly sample 28K, 3K, and 1.5K triples for training, validation, and testing sets, respectively.

**OpenSubtitles (OSDb):** OSDb[^1] is a very large and noisy open-domain dataset containing roughly 60M-70M scripted lines. We first preprocess the triples as we do with DailyDialog, then select one subset in our experiment and split it into 1500K, 50K, and 25K triples for training, validation, and testing set, respectively.

[^1]: [http://opus.lingfil.uu.se/OpenSubtitles.php](http://opus.lingfil.uu.se/OpenSubtitles.php)

### Comparison Models

We compare the proposed Posterior-GAN with the following state-of-the-art models:

**Seq2Seq-att:** The generator is a sequence-to-sequence model (Sutskever, Vinyals, and Le 2014) with attention mechanism (Bahdanau, Cho, and Bengio 2015). A maximum likelihood estimation (MLE) objective is used to train the model.

**Adver-REGS:** Adver-REGS (Li et al. 2017a) uses a sequence-to-sequence model to generate response. A binary classifier based discriminator calculates reward to train generator with policy gradient.

**DP-GAN:** DP-GAN (Xu et al. 2018a) also consists of a generator and a discriminator. Different from Adver-REGS, this discriminator is a cross-entropy based language model which alleviates the reward saturation problem.

### Training Details

Based on the loss and the metrics on the validation set, we train the comparison models and our model with the following hyperparameters: The word embedding size is 256. The hidden size is set to 256. To conduct a fair comparison among all the models, We set the encoder layer to 2 and the decoder layer to 4. The encoder is a bidirectional LSTM. The vocabulary for DailyDialog and OpenSubtitles is of size 20,000 and 50,000, respectively. The batch size is set to 256 for pre-training and adversarial training. All the parameters is initialized using a normal distribution $N(0, 0.0001)$. All the models are trained end-to-end using Adam (Kingma and Ba 2015) with a learning rate of 0.0001 and a global norm clipping at 2.0. For Adver-REGS, DP-GAN, and our model, before adversarial learning, we pre-train the generator for 10 epochs. In adversarial training, we alternatively train the generator every 1000 steps and optimize the discriminator every 5000 steps.

### Evaluation Metrics

We evaluate the model in terms of following automatic evaluation metrics:

- **BLEU** (Papineni et al. 2002), a word-overlapping based metric, which calculates word overlapping degree between the generated response and the real-world response. Recently plenty of work adopts its to reflect the lexical similarity of response (Li et al. 2016a; Zhao, Zhao, and Eskenazi 2017; Zhang et al. 2018a).

- **Embedding-based Metrics.** Embedding Average (Average), Embedding Greedy (Greedy) and Embedding Extreme (Extrema) (Liu et al. 2016) are used in the experiments. The three embedding-based metrics first calculate semantic embedding based on the vectors of all individual tokens in responses and then calculate the similarity between the generated response and the real-world response by cosine distance. They are widely used to evaluate the semantic similarity of response (Serban et al. 2017; Zhang et al. 2018b; Csaky, Purgai, and Recski 2019).

- **Distinct.** Dist-\{1,2,3\} are employed to reflect the degree of diversity of the generated responses, which are widely
### Experimental Results

#### Overall Performance

| Models   | Dist-1 | Dist-2 | Dist-3 | BLEU  | Greedy | Average | Extrema |
|----------|--------|--------|--------|-------|--------|---------|---------|
| Seq2Seq-att | 0.0277 | 0.1625 | 0.3868 | 0.1878 | 0.4825 | 0.5993  | 0.3080  |
| Adver-REGS  | 0.0541 | 0.2877 | 0.5542 | 0.2116 | 0.4857 | 0.6215  | 0.3542  |
| DP-GAN     | 0.0656 | 0.3088 | 0.5630 | 0.1992 | 0.4749 | 0.6144  | 0.3409  |
| Posterior-GAN(F) | 0.0659 | 0.2995 | 0.5578 | 0.2067 | 0.4734 | 0.6218  | 0.3343  |
| Posterior-GAN(B) | 0.0578 | 0.2950 | 0.5545 | 0.2130 | 0.4818 | 0.6220  | 0.3442  |
| Posterior-GAN(A) | 0.0678 | 0.3549 | 0.6006 | 0.2183 | 0.4916 | 0.6260  | 0.3544  |

Table 1: The automatic metrics evaluation results. Higher is better. “(F)”, “(B)” and “(A)” represent Posterior-GAN with a forward generative discriminator, a backward generative discriminator and both two discriminators, respectively.

| Models   | Coherence | Informativeness |
|----------|-----------|----------------|
| Seq2Seq-att | 3.8550    | 3.8933          |
| Adver-REGS  | 3.4717    | 3.3683          |
| DP-GAN     | 3.4400    | 3.2350          |
| Posterior-GAN | 3.2883  | 3.2250          |

Table 2: The human evaluation results. We calculate each score by averaging the rank of each model in corresponding metrics. Lower is better.

#### Ablation Test

Comparing the forward and backward generative discriminator in Posterior-GAN, we observe that forward generative discriminator achieves better performance on the diversity metrics, whereas backward generative discriminator performs better on lexical and semantic similarities. The difference lies in that backward generative discriminator directly calculates the reward for all individual tokens of the generated response in terms of the future conversation, in the supplement of the generation perspective based on the query, and forward generative discriminator measures whether the generated response is informative enough to predict the subsequent real-world turns.

#### Qualitative Evaluation

Due to the known fact that quantitative metrics and human perception have a certain degree of deviation (Stent, Marge, and Singhai 2005), e.g., the conceptual difference of informativeness and diversity (Zhang et al. 2018b), we use human evaluation as a qualitative way to further evaluate our model and comparison models. We randomly select 200 samples from the test sets in the two dialogue datasets separately. Each sample consists of query, future turn and responses generated by different models. We invite three annotators to rank the generated responses with respect to two aspects: coherence and informativeness. Ties are allowed. Coherence indicates how likely the generated response is relevant to both query and future turn. Informativeness specifies how much the information related to the context is contained in the generated response. The annotators are all well educated and are ignorant of the models by which the responses are generated. To ensure a stable comparison, the evaluated models consist of 2 Seq2Sea-att models with different initialization, Adver-REGS, DP-GAN, and 3 Posterior-GAN outputs. The results of the human evaluation are shown in Table 2. We also report the inter-annotator agreement to demonstrate the consistency of three annotators. The spearman’s rank correlation coefficient for coherence and informativeness is 0.3948 and 0.3406, with \( p < 0.0001 \). Augmented with adversarial
Case Study

and coherence of the response. In this section, we further analyze why future turn and two symmetric generative discriminators have a positive effect on the performance of model and increase informativeness and coherence of the response.

**Query:** I need some flowers for my girlfriend. // No problem. 
Would you like some artificial carnations? // Oh, no. Carnations are not very elegant. Artificial flowers have no passion.

**Seq2Seq-att:** How how much how much? I'd like.
**Adver-REGS:** How did you want to spend? How much did.
**DP-GAN:** How long did you want to spend? They was.

**Posterior-GAN:** How much did you want to spend? It’s very expensive.

**Future turn:** Money is no object. // Our most elegant flower is golden lily. // I will take ten.

**Query:** How may I help you today? // I need to open a second account. // What kind of account would you like to open?

**Seq2Seq-att:** I want to need to deposit to deposit.
**Adver-REGS:** I need to open the savings account.
**DP-GAN:** I want to open cash at cash. I need

**Posterior-GAN:** I need to transfer a savings account account.

**Future turn:** Do you have another account with us? // I sure do. // Would you like to transfer money from that account into your new one?

**Query:** What’s your case? // I was pulled over for running a red light, but I never did. // Do you believe that the officer lied?

**Seq2Seq-att:** Well, you was so in that.
**Adver-REGS:** We never like to lie.
**DP-GAN:** Because I’ve seen them.

**Posterior-GAN:** Of course, you can’t know that.

**Future turn:** Your license plate was caught on camera? // A picture of my license plate was never taken. // Since there is no picture of your license plate on record, I’m going to let you go.

Table 3: The responses generated by the proposed models and comparison models on DailyDialog.

| Models            | Averaged Greedy Matching | Frequency-based Similarity |
|-------------------|--------------------------|---------------------------|
| Seq2Seq-att       | 0.5942                   | 0.5478                    |
| Adver-REGS        | 0.6403                   | 0.7165                    |
| DP-GAN            | 0.6514                   | 0.6739                    |
| Posterior-GAN     | 0.7276                   | 0.7460                    |

Table 4: The results of Embedding-based Averaged Greedy Matching for response $y$ with the given query $x$ and future conversations $z$, which reflects the coherence of the response, and Frequency-based Similarity, which illustrates the informativeness of the response.

other examples, but we do not show them for limited space.

**Automatic Analysis** To further demonstrate the above observations, we design two metrics to verify the superiority of the generated responses. The informativeness of the responses is reflected by comparing the word frequency similarity between the generated response and the ground-truth response, where the responses are represented as a vector, and each element in the vector is denoted as the frequency of a word. Here, we use 2350 most frequent words from the training set of DailyDialog corpus without stop words and meaningless words. To validate the coherence of the generated response, we calculate the average matching degree of the generated response with the given query and the subsequent real-world conversations by utilizing the embedding-based greedy matching metric, which prefers response with keywords that have high semantic similarity with those in the real-world context (Liu et al. 2016). The results are shown in Table 4. Regarding both the frequency similarity and matching degree, our model consistently outperforms the comparison models, which indicates that our model generates more informative and coherent responses.

**Visualization** We also visualize the reward distribution of two symmetric discriminators to get some insight into the behavior of the model on DailyDialog. Figure illustrates the reward distribution and representative conversations. The reward distribution is roughly divided into three regions. We observe that some responses are of higher rewards on $R_1$ (subtract the minimum value) but low rewards on $R_2$ (the average $R_2(y_m)$ of response) as in the yellow region C of Figure The response from regions C provides specific information to answer the preceding context. The responses from red region B of Figure are of higher rewards on $R_2$, whereas the rewards on $R_1$ is much lower. Although the response ‘no’ is lack of informativeness, it is coherent to the given query and the future turn. By simultaneously integrating both $R_1$ and $R_2$, the generator pays more attention to responses that are not only informative but also coherent and achieves high rewards on both generative discriminators.

**Related Work**

A tremendous amount of effort has been paid to increase the informativeness and diversity of neural dialogue generation model. Li et al. (2016a) adopted Maximum Mutual Infor-
Figure 4: The distribution of sample rewards calculated by the forward generative discriminator \( R_1 \) and the backward generative discriminator \( R_2 \) on DailyDialog. \( R \) is the combination of \( R_1 \) and \( R_2 \). We use three regions A, B, and C to represent three types of samples. Samples in region A gain high rewards in both discriminators. Samples in region B achieve higher reward in the backward generative discriminator than in the forward one, while Samples in region C obtain higher reward in the forward generative discriminator than in the backward one.

In this paper, we propose the query-response-future turn triples instead of the conventional query-response pairs for neural dialog response generation. To facilitate the triple modeling and alleviate the overproducing of generic and repetitive responses problem, Posterior-GAN that consists of a forward and a backward encoder-decoder based generative discriminators is further introduced. Augmented with future conversations and Posterior-GAN in training, detailed experiments and analysis demonstrate that the model effectively generates more informative and coherent responses.

Acknowledgements

This research is supported by National Key R&D Program of China (No.2016YFB0801100), Beijing Natural Science Foundation (No.4172054, L181010), and National Basic Research Program of China (No.2013CB329605). Kan Li is the corresponding author.
References

[Bahdanau, Cho, and Bengio 2015] Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural machine translation by jointly learning to align and translate. In ICLR.

[Clark and Cao 2017] Clark, S., and Cao, K. 2017. Latent variable dialogue models and their diversity. In EACL (2), 182–187.

[Csaky, Purgai, and Recski 2019] Csaky, R.; Purgai, P.; and Recski, G. 2019. Improving neural conversational models with entropy-based data filtering. In ACL (1), 5650–5669.

[Fedus, Goodfellow, and Dai 2018] Fedus, W.; Goodfellow, I. J.; and Dai, A. M. 2018. Maskgan: Better text generation via filling in the ______. In ICLR (Poster).

[Goodfellow et al. 2014] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. C.; and Bengio, Y. 2014. Generative adversarial nets. In NIPS, 2672–2680.

[Gu et al. 2019] Gu, X.; Cho, K.; Ha, J.; and Kim, S. 2019. Dialogwae: Multimodal response generation with conditional wasserstein auto-encoder. In ICLR (Poster).

[Kingma and Ba 2015] Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In ICLR (Poster).

[Li et al. 2016a] Li, J.; Galley, M.; Brockett, C.; Gao, J.; and Dolan, B. 2016a. A diversity-promoting objective function for neural conversation models. In HLT-NAACL, 110–119.

[Li et al. 2016b] Li, J.; Monroe, W.; Ritter, A.; Jurafsky, D.; Galley, M.; and Gao, J. 2016b. Deep reinforcement learning for dialogue generation. In EMNLP, 1192–1202.

[Li et al. 2017a] Li, J.; Monroe, W.; Shi, T.; Jean, S.; Ritter, A.; and Jurafsky, D. 2017a. Adversarial learning for neural dialogue generation. In EMNLP, 2157–2169.

[Li et al. 2017b] Li, Y.; Su, H.; Shen, X.; Li, W.; Cao, Z.; and Niu, S. 2017b. Dailydialog: A manually labelled multi-turn dialogue dataset. In IJCNLP(1), 986–995.

[Liu et al. 2016] Liu, C.; Lowe, R.; Serban, I.; Noseworthy, M.; Charlin, L.; and Pineau, J. 2016. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In EMNLP, 2122–2132.

[Mou et al. 2016] Mou, L.; Song, Y.; Yan, R.; Li, G.; Zhang, L.; and Jin, Z. 2016. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. In COLING, 3349–3358.

[Papineni et al. 2002] Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W. 2002. Bleu: a method for automatic evaluation of machine translation. In ACL, 311–318.

[Serban et al. 2016] Serban, I. V.; Sordoni, A.; Bengio, Y.; Courville, A. C.; and Pineau, J. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In AAAI, 3776–3784.

[Serban et al. 2017a] Serban, I. V.; Klinger, T.; Tesauro, G.; Talamadupula, K.; Zhou, B.; Bengio, Y.; and Courville, A. C. 2017a. Multiresolution recurrent neural networks: An application to dialogue response generation. In AAAI, 3288–3294. AAAI Press.

[Serban et al. 2017b] Serban, I. V.; Sordoni, A.; Lowe, R.; Charlin, L.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2017b. A hierarchical latent variable encoder-decoder model for generating dialogues. In AAAI, 3295–3301.

[Shang, Lu, and Li 2015] Shang, L.; Lu, Z.; and Li, H. 2015. Neural responding machine for short-text conversation. In ACL (1), 1577–1586.

[Shen et al. 2018] Shen, X.; Su, H.; Niu, S.; and Demberg, V. 2018. Improving variational encoder-decoders in dialogue generation. In AAAI, 5456–5463.

[Stent, Marge, and Singhai 2005] Stent, A.; Marge, M.; and Singhai, M. 2005. Evaluating evaluation methods for generation in the presence of variation. In CICLing, volume 3406 of Lecture Notes in Computer Science, 341–351.

[Sutskever, Vinyals, and Le 2014] Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In NIPS, 3104–3112.

[Sutton et al. 1999] Sutton, R. S.; McAllester, D. A.; Singh, S. P.; and Mansour, Y. 1999. Policy gradient methods for reinforcement learning with function approximation. In NIPS, 1057–1063.

[Vinyals and Le 2015] Vinyals, O., and Le, Q. V. 2015. A neural conversational model. CoRR abs/1506.05869.

[Williams 1992] Williams, R. J. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning 8:229–256.

[Wu et al. 2018] Wu, L.; Tian, F.; Qin, T.; Lai, J.; and Liu, T. 2018. A study of reinforcement learning for neural machine translation. In EMNLP, 3612–3621.

[Xing et al. 2017] Xing, C.; Wu, W.; Wu, Y.; Liu, J.; Huang, Y.; Zhou, M.; and Ma, W. 2017. Topic aware neural response generation. In AAAI, 3351–3357.

[Xu et al. 2017] Xu, Z.; Liu, B.; Wang, B.; Sun, C.; Wang, X.; Wang, Z.; and Qi, C. 2017. Neural response generation via GAN with an approximate embedding layer. In EMNLP, 617–626.

[Xu et al. 2018a] Xu, J.; Ren, X.; Lin, J.; and Sun, X. 2018a. Diversity-promoting GAN: A cross-entropy based generative adversarial network for diversified text generation. In EMNLP, 3940–3949.

[Xu et al. 2018b] Xu, X.; Dusek, O.; Konstas, I.; and Rieser, V. 2018b. Better conversations by modeling, filtering, and optimizing for coherence and diversity. In EMNLP, 3981–3991.

[Yu et al. 2017] Yu, L.; Zhang, W.; Wang, J.; and Yu, Y. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In AAAI, 2852–2858.

[Zhang et al. 2018a] Zhang, H.; Lan, Y.; Guo, J.; Xu, J.; and Cheng, X. 2018a. Reinforcing coherence for sequence to sequence model in dialogue generation. In IJCAI, 4567–4573.

[Zhang et al. 2018b] Zhang, Y.; Galley, M.; Gao, J.; Gan, Z.; Li, X.; Brockett, C.; and Dolan, B. 2018b. Generating infor-
mative and diverse conversational responses via adversarial information maximization. In *NeurIPS*, 1815–1825.

[Zhao, Zhao, and Eskénazi 2017] Zhao, T.; Zhao, R.; and Eskénazi, M. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *ACL (1)*, 654–664.