DAL: Dual Adversarial Learning for Dialogue Generation

Shaobo Cui [1], Rongzhong Lian [2], Di Jiang [2], Yuanfeng Song [2], Siqi Bao [2], Yong Jiang [1]
[1] Tsinghua University, China
[2] Baidu Inc., China
cuishaobo16@mails.tsinghua.edu.cn
{lianrongzhong,jiangdi,songyuanfeng,baosiqi}@baidu.com
jiangy@sz.tsinghua.edu.cn

Abstract

In open-domain dialogue systems, generative approaches have attracted much attention for response[1] generation. However, existing methods are heavily plagued by generating safe responses and unnatural responses. To alleviate these two problems, we propose a novel framework named Dual Adversarial Learning (DAL) for high-quality response generation. DAL innovatively utilizes the duality between query generation and response generation to avoid safe responses and increase the diversity of the generated responses. Additionally, DAL uses adversarial learning to mimic human judges and guides the system to generate natural responses. Experimental results demonstrate that DAL effectively improves both diversity and overall quality of the generated responses. DAL outperforms state-of-the-art methods regarding automatic metrics and human evaluations.

1 Introduction

In recent years, open-domain dialogue systems are gaining much attention owing to their great potential in applications such as educational robots, emotional companion, and chitchat. The existing approaches for open-domain dialogue systems can be divided into two categories: retrieval-based approaches (Hu et al., 2014; Ji et al., 2014) and generative approaches (Ritter et al., 2011; Shang et al., 2015). The retrieval-based approaches are based on conventional information retrieval techniques and strongly rely on the underlying corpus (Wang et al., 2013; Lu and Li, 2013). Since the capability of retrieval-based approaches is strongly limited by corpus, generative approaches are attracting more attention in the field of open-domain dialogue research. The *de facto* backbone of generative approaches is the Seq2Seq model (Bahdanau et al., 2014) , which is essentially an encoder-decoder neural network architecture. Despite their success, Seq2Seq model and its variants (Sordoni et al., 2015; Vinyals and Le, 2015) are heavily plagued by **safe responses** (generic and dull responses such as "I don't know" or "Me too") and **unnatural responses** (such as "I want to go, but I don't want to go").

In this paper, we propose a novel framework named Dual Adversarial Learning (DAL) to alleviate the aforementioned two problems. DAL consists of two generative adversarial networks (GANs): one for query generation and the other for response generation. The response generation model is used to transfer from the query domain $\mathcal{Q}$ to the response domain $\mathcal{R}$, while the query generation model is for transformation from $\mathcal{R}$ to $\mathcal{Q}$. Here we consider the response generation task and the query generation task as **dual** tasks. The generators of these two GANs are connected through the duality constraint. As such, in DAL, there are two kinds of signals that jointly instruct the optimization of generators: (1) the dual signal from the duality constraint between these two generators; (2) the adversarial signal from the discriminators. The dual signal is utilized to model the mutual relation between query generation and response generation. We use an instance to better illustrate this mutual relation: for a given query "Where to have dinner?", compared with a safe response "I dont know", a more diverse and specific response "The Indian cuisine around the corner is great" usually has a higher probability of being transformed back to the given query. DAL takes full advantage of this intuition via dual learning, which avoids generating safe responses and improves the diversity of the generated responses. Additionally, in order to make the generated responses as natural as possible, the adversarial signal in DAL mimics human judges to alle-

[1] We use *query* and *response* to denote the first and second utterances in a single-turn dialogue.

Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation (NeuralGen), pages 11–20
Minneapolis, Minnesota, USA, June 6, 2019.

viate unnatural responses. We compare DAL with state-of-the-art methods through extensive experiments, and DAL demonstrates superior performance regarding automatic metrics, human evaluations, and efficiency.

There are **crucial differences** between our dual approach and Maximum Mutual Information (MMI) (Li et al., 2016) though both utilize the reverse dependency to improve the diversity of the generated responses. Due to the challenging mutual information objective, the distribution $p(r|q)$ is same as that in vanilla Seq2Seq in MMI. More specifically, $p(r|q)$ in MMI is trained only by maximum likelihood estimation (MLE) objective at training time (we use $p(r|q)$ to denote the probability distribution of predicting the response r given the query q). The mutual information in MMI is utilized only at inference time, and the inference process is not only time-consuming but also inaccurate in MMI. However, $p(r|q)$ in our dual approach is trained by not only the maximum likelihood estimation objective but also the diversity objective (duality constraint) at training time. Since the dual approach directly incorporates the reverse dependency information at the training time, it can avoid the time-consuming inference plaguing MMI. Additionally, the dual approach does not need to maintain a large size optional response set for the time-consuming reranking strategy in MMI-bidi (one variant of MMI). The dual approach shows its efficiency superiority over MMI in real-life applications, which is shown in our efficiency experiment.

Our dual approach is quite different from the reinforcement learning based structure having two Seq2Seq models in (Zhang et al., 2018) [2] . In (Zhang et al., 2018), G_1, which generates a response $\hat{r}$ given a query q, uses the conditional probability $P_2(q|\hat{r})$ calculated by G_2 as the coherence measure to guide G_1 in the reinforcement learning process. Similarly, G_2, which generates a query $\hat{q}$ given a response r, uses the conditional probability $P_1(r|\hat{q})$ calculated by G_1 as the coherence measure to guide G_2 in the reinforcing learning process. However, in our work, we utilize the joint probability $p(q, r)$ to connect these two Seq2Seq models and thus avoid unstable and time-consuming reinforcement learning in the dual approach. Besides, our DAL framework is strongly different from previous structures that are composed of two GANs, such as CycleGAN (Zhu et al., 2017), DiscoGAN (Kim et al., 2017) and DualGAN (Yi et al., 2017). Those works can only be utilized on the image translation task and two generators are connected by *cycle consistency*, i.e., for each image x in domain $\mathcal{X}$, the image translation cycle is supposed to bring x to the original image: $x \rightarrow G_1(x) \rightarrow G_2(G_1(x)) \approx x$. However, *cycle consistency* is difficult to be applied into the text generation task. In our paper, we use the *joint distribution* of query-response pairs rather than *cycle consistency* to enforce the duality between these two dual generators.

The contributions of this paper are as follows:

• To the best of our knowledge, this is the first work that adopts the duality to avoid safe responses for dialogue generation. It sheds light on the utility of query generation in improving the performance of response generation.

• DAL is a novel framework that integrates dual learning and adversarial learning, which complementary and jointly contributes to generating both diverse and natural responses.

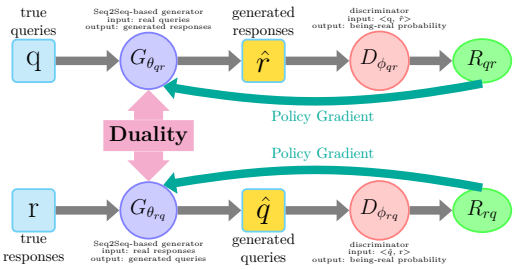
The rest of this paper is organized as follows. The related work is firstly reviewed. The DAL framework is introduced in Section 3 and the training of DAL is described in Section 4. Experimental results are shown in Section 5, followed by the conclusion of this paper in Section 6.

2 Related Work

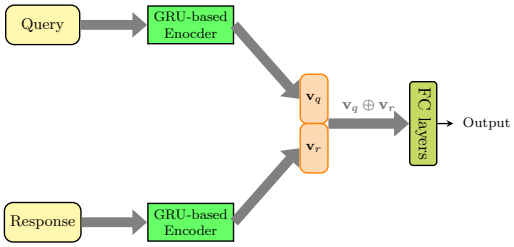
Dual Learning Many machine learning tasks have emerged in dual forms, such as dual neural machine translation (dual-NMT) (He et al., 2016), image classification and conditional image generation (van den Oord et al., 2016). Dual learning (He et al., 2016) is proposed on the assumption that the dual correlation could be used to improve both the primal task and its dual task: the primal task aims to map from input space $\mathcal{X}$ to output space $\mathcal{Y}$, whereas the dual task takes samples from space $\mathcal{Y}$ and maps to space $\mathcal{X}$. Tang et al. (2017) implemented a dual framework for the question answering system. Their model regards the answer selection (given a question and its several candidate answers, select the most satisfying answer to answer the question) and the question generation as dual tasks, which increases the performance of both.

Adversarial Learning Adversarial learn-

[2] Our dual approach is finished independently with this work in addition to the crucial difference. We did not notice this paper until our work is done.



(a) The architecture of DAL.



(b) The architecture of the discriminator.

Figure 1: Dual Adversarial Learning.

ing (Goodfellow et al., 2014), or Generative Adversarial Networks (GAN), has proven to be a promising approach for generation task. A GAN usually contains two neural networks: a generator G and a discriminator D. G generates samples while D is trained to distinguish generated samples from true samples. By regarding the sequence generation as an action-taking problem in reinforcement learning, Li et al. (2017) proposed to apply GAN to dialogue generation, in which the output of the discriminator is used as the reward for the generator's optimization.

Work on the Safe Response Problem There is some existing work on the safe response problem. The first kind of approach is to introduce specific keywords (Mou et al., 2016) or topic information (Xing et al., 2017) into the generated responses. These methods help to increase the dialogue coherence (Peng et al., 2019) by keywords introduction. However, these methods shift the difficulty from diverse response generation to keyword or topic prediction, which are also challenging tasks. The second kind of approach takes the reverse dependency (the query generation task given the responses) into consideration. Li et al. (2016) considered the reverse dependency and proposed Maximum Mutual Information (MMI) method, which is empirically plagued by ungrammatical responses (MMI-antiLM) and huge decoding space (MMI-bidi).

3 DAL Framework

In this section, we firstly given an overview of DAL framework and then elaborate the discriminators and the generations. We also present the reason why duality promotes diversity.

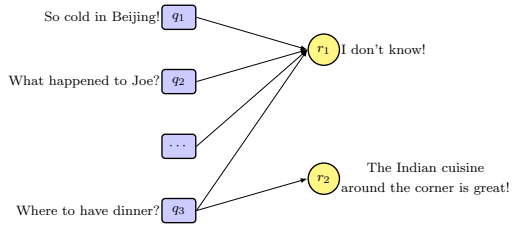
3.1 Overview

The architecture of DAL is presented in Figure 1(a). The real query and response are denoted by q and r, whereas the generated query and response are denoted as $\hat{q}$ and $\hat{r}$. DAL consists of two GANs (one for query generation and the other for response generation). Generators are denoted by $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$ and the corresponding discriminators are denoted as $D_{\phi_{qr}}$ and $D_{\phi_{rq}}$. The input of $G_{\theta_{qr}}$ is a real query q and the output is the generated response $\hat{r}$. Similarly, for $G_{\theta_{rq}}$, the input is a real response r and the output is the generated query $\hat{q}$. For $D_{\phi_{qr}}$, the input is the *ficto-facto* query-response pair $\langle q, \hat{r} \rangle$, and the output R_{qr} is estimated probability of the query-response pair being human-generated, which is estimated by $D_{\phi_{qr}}$. Analogously, the input of $D_{\phi_{rq}}$ is the *ficto-facto* pair $\langle \hat{q}, r \rangle$, and the output R_{rq} is the estimated probability of the input pair being human-generated. $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$ are connected by the duality constraint derived from the joint probability $P(q, r)$. The adversarial signal from discriminators, R_{qr}, R_{rq}, are passed to the corresponding generators as the reward through policy gradient.

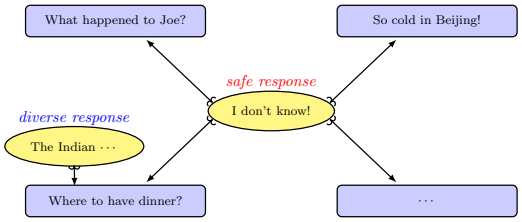
3.2 Discriminator

The discriminator mimics a human judge and guides the generator to generate natural utterances. The architecture of the discriminator is shown in Figure 1(b). Gated Recurrent Unit (GRU) based (Bahdanau et al., 2014) neural networks are used to obtain the query embedding $\mathbf{v}_q$ and the response embedding $\mathbf{v}_r$. The concatenation vector $\mathbf{v}_q \oplus \mathbf{v}_r$ is used as the abstract representation of the query-response pair. $\mathbf{v}_q \oplus \mathbf{v}_r$ is further passed through two fully-connected layers. The output of the last fully-connected layer is the estimated probability of the query-response pair being human-generated. The objective of the discriminator is formalized as follows:

$$\begin{aligned} \min_{\phi} & - \mathbb{E}_{\langle q,r \rangle \sim p_{data}} \left[\log \left(D_{\phi}(\langle q, r \rangle) \right) \right] \\ & - \mathbb{E}_{\langle q,r \rangle \sim G_{\theta}} \left[\log \left(1 - D_{\phi}(\langle q, r \rangle) \right) \right] \end{aligned} \tag{1}$$



(a) An example corpus.



(b) Queries and responses with duality constraint.

Figure 2: An example to illustrate why duality promotes diversity.

where p_{data} denotes the real-world query-response distribution. For the response generation task, D_ϕ is $D_{\phi_{qr}}$ and G_θ is $G_{\theta_{qr}}$, while for the query generation task, D_ϕ is $D_{\phi_{rq}}$ and G_θ is $G_{\theta_{rq}}$.

3.3 Dual Generators

Both generators adopt the Seq2Seq structure, in which GRU is used as the basic unit. The constraint between the dual tasks (query generation and response generation) can be represented with the joint probability $P(q, r)$:

$$P(q, r) = P_q(q)P(r|q; \theta_{qr}) = P_r(r)P(q|r; \theta_{rq}) \quad (2)$$

where $P_q(q)$ and $P_r(r)$ are language models pre-trained on the query corpus and the response corpus. In this paper, we use smoothed bi-gram language models for both $P_q(q)$ and $P_r(r)$. $P(r|q; \theta_{qr})$ and $P(q|r; \theta_{rq})$ are the dual generators. Both $P(r|q; \theta_{qr})$ and $P(q|r; \theta_{rq})$ can be obtained through the markov chain rule:

$$\begin{cases} P(r|q; \theta_{qr}) = \prod_{t=1}^{|r|} P(r^t|r^{0:t-1}, q; \theta_{qr}) \\ P(q|r; \theta_{rq}) = \prod_{t=1}^{|q|} P(q^t|q^{0:t-1}, r; \theta_{rq}) \end{cases}$$

where $P(r^t|r^{0:t-1}, q; \theta_{qr})$ and $P(q^t|q^{0:t-1}, r; \theta_{rq})$ are formulations of decoders in Seq2Seq models.

3.4 Duality Promotes Diversity

To better illustrate why duality increases the diversity of the generated responses, we show some query-response pair examples in Figure 2(a). In Figure 2(a), each directional arrow starts from a query while ends at its corresponding response. It can be observed that: (1) Safe response r_1 : "I don't know" connects to many queries, i.e., $\{q_1, q_2, q_3, \cdots\}$. (2) More diverse and specific response r_2 : "The Indian cuisine around the corner is great", nevertheless, exactly corresponds to only one query q_3 : "Where to have dinner?". [3]

In the training process of $G_{\theta_{rq}}$, the increase of $\log P(q_3|r_2; \theta_{rq})$, denoted by $\Delta \log P(q_3|r_2; \theta_{rq})$ [4], is much bigger than the increase of $\log P(q_3|r_1; \theta_{rq})$, denoted by $\Delta \log P(q_3|r_1; \theta_{rq})$. Formally,

$$\Delta \log P(q_3|r_2; \theta_{rq}) \gg \Delta \log P(q_3|r_1; \theta_{rq})$$

The reason behind this phenomenon is as follows. The safe response r_1 relates with queries $\{q_1, q_2, q_3, \cdots\}$. When $G_{\theta_{rq}}$ is provided with $\langle q_1, r_1 \rangle$ or $\langle q_2, r_1 \rangle$, $G_{\theta_{rq}}$ is optimized to increase the log conditional probability $\log P(q_1|r_1; \theta_{rq})$ or $\log P(q_2|r_1; \theta_{rq})$, it is inevitable that $\log P(q_3|r_1, \theta_{rq})$ will decrease to a certain extent, since these log conditional probabilities share the same parameters θ_{rq}. The same principle applies to $\log P(q_2|r_1, \theta_{rq})$ when $G_{\theta_{rq}}$ is provided with $\langle q_1, r_1 \rangle$ or $\langle q_3, r_1 \rangle$. However, the diverse response r_2 is uniquely connected to the query q_3, in that case, $G_{\theta_{rq}}$ takes all efforts to increase $\log P(q_3|r_2, \theta_{rq})$.

With the duality constraint in Eq. 2, we obtain:

$$\frac{P(q|r; \theta_{rq})}{P(r|q; \theta_{qr})} = \frac{P_q(q)}{P_r(r)} = k(q, r). \quad (3)$$

Since both $P_q(q)$ and $P_r(r)$ are obtained from the pre-trained language models, both of them are constant for any query-response pair $\langle q, r \rangle$. $k(q, r) = \frac{P_q(q)}{P_r(r)}$ is also constant for any $\langle q, r \rangle$. Take the log formulation of Eq. 3, we can obtain:

$$\log P(q|r; \theta_{rq}) - \log P(r|q; \theta_{qr}) = \log k(q, r).$$

From above equation, we observe that the increase of $\log P(q|r; \theta_{rq})$, denoted as $\Delta \log P(q|r; \theta_{rq})$,

[3] There may exist several other queries that can be replied using "The Indian cuisine around the corner is great". But this number is much smaller than those that can be replied using "I don't know". For simplicity, we only show only one query here for the response "The Indian cuisine around the corner is great". This would not affect the following analysis.

[4] The reason why the probability is in log formulation is that the probability which the maximum likelihood objective optimize is in log formulation rather than origin formulation

and the increase of $\log P(r|q;\theta_{qr})$, denoted by $\Delta \log P(r|q;\theta_{qr})$, is supposed to be equal for any query-response pair $\langle q, r\rangle$, since $\log k(q, r)$ is constant during the training process. Therefore,

$$\Delta \log P(q_3|r_2;\theta_{rq}) \gg \Delta \log P(q_3|r_1;\theta_{rq})$$

in turn makes

$$\Delta \log P(r_2|q_3;\theta_{qr}) \gg \Delta \log P(r_1|q_3;\theta_{qr}).$$

When $G_{\theta_{qr}}$ finishes its training process, we obtain $P(r_2|q_3;\theta_{qr}) \gg P(r_1|q_3;\theta_{qr})$. This indicates that it is more likely for $G_{\theta_{qr}}$ to assign higher probability to the diverse response given the query.

We use Figure 2(b) to visually explain this intuition. We suppose that both queries and responses "possess" their own spatial space. The coordinates of the ellipse and the rectangle represent the locations of the query q and the response r in the spatial space. The distance between q and r represents the probability of transforming between q and r, namely $P(q|r)$ and $P(r|q)$. The shorter the distance, the larger the probability. When $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$ are provided with a query-response pair $\langle q, r\rangle$, the training objectives of $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$ are to increase the probability $P(r|q)$ and $P(q|r)$, i.e., to shorten the distance between q and r. Since the safe response r_1 corresponds to $\{q_1, q_2, q_3, \cdots\}$, the position of this safe response is determined by all involved queries. Because each of these involved queries attempts to "drag" r_1 close to itself, the safe response r_1 "chooses" to keep a distance with each of them to balance the involved queries. However, the diverse response r_2 corresponds to exactly one query q_3. r_2 "selects" to stay as close to q_3 as possible. As it can be seen from the figure, the distance between q_3 and r_2 is much shorter than the distance between q_3 and r_1, i.e., $P(r_2|q_3)$ is much larger than $P(r_1|q_3)$. In other words, with the duality constraint, $G_{\theta_{qr}}$ tends to generate diverse responses rather than safe responses.

4 Training of DAL

Duality Constraint for Diversity Direct enforcement of the constraint in Eq. 2 is intractable. The duality constraint in Eq. 2 can be *relaxed* into a regularization term (Tang et al., 2017):

$$\begin{aligned}\Upsilon = [&\log P_r(r) + \log P(q|r;\theta_{rq}) \\ &- \log P_q(q) - \log P(r|q;\theta_{qr})]^2.\end{aligned} \quad (4)$$

We minimize Υ to enforce the duality constraint in order to generate more diverse responses.

Adversarial Signal for Naturalness The decoding phase in the Seq2Seq model involves sampling discrete words. This discrete sampling makes the optimization of the generator based upon the discriminator's guidance non-differentiable. To circumvent the non-differentiable obstacle, we optimize each generator through reinforcement learning. The policy gradient is applied to pass the discriminator's adversarial signal to the generator. The discriminator D_ϕ gives a score $J(\theta)$ based on its judgment of how likely the generated $\langle q, r\rangle$ is human-generated:

$$J(\theta) = \mathbb{E}_{\langle x,y\rangle \in G_\theta}[D_\phi(\langle x, y\rangle)].$$

For response generation, $J(\theta)$ is $J(\theta_{qr})$, G_θ is $G_{\theta_{qr}}$, D_ϕ is $D_{\phi_{qr}}$, x is the real query and y is the generated response. Analogously, in query generation, $J(\theta)$ is $J(\theta_{rq})$, G_θ is $G_{\theta_{rq}}$, D_ϕ is $D_{\phi_{rq}}$, x is the real response and y is the generated query. $J(\theta)$ is used as the reward for the optimization of G_θ. With the likelihood ration trick (Williams, 1992; Sutton et al., 2000), the gradient of $J(\theta)$ can be approximated as:

$$\nabla_\theta J(\theta) \simeq [D_\phi(\langle x, y\rangle) - b] \cdot \nabla_\theta \log(p(y|x;\theta)),$$

where b is used to reduce the variance of the estimation while keeping the estimation unbiased, and $p(y|x;\theta)$ is the probability distribution defined by the generator G_θ.

Combined Gradient In DAL, the gradient for updating each generator is the weighted combination of $\nabla_\theta J(\theta)$ (for natural responses) and $\nabla_\theta \Upsilon$ (for avoidance of safe responses):

$$\begin{cases}\nabla_{\theta_{qr}} G_{\theta_{qr}} = \nabla_{\theta_{qr}} \Upsilon - \lambda_{qr} \cdot \nabla_{\theta_{qr}} J(\theta_{qr}) \\ \nabla_{\theta_{rq}} G_{\theta_{rq}} = \nabla_{\theta_{rq}} \Upsilon - \lambda_{rq} \cdot \nabla_{\theta_{rq}} J(\theta_{rq})\end{cases}. \quad (5)$$

Teacher Forcing When the generator is trained with only the adversarial signals from the discriminator and the duality constraint, the training process of the generator easily collapses. This is because the discriminator sometimes is remarkably better than the corresponding generator in certain training batches. The discriminator can easily discriminate all the generated utterances from real ones. The generator realizes that it generates low-quality samples but cannot figure out the good standard. To stabilize the training process, after each update with the combined gradient $\nabla_{\theta_{qr}} G_{\theta_{qr}}$ or $\nabla_{\theta_{rq}} G_{\theta_{rq}}$, the generators are

provided with real query-response pairs and are strengthened with maximum likelihood training, which is also known as Teacher Forcing (Li et al., 2017; Lamb et al., 2016). The training procedure

Algorithm 1 Training of DAL.

Input: Pre-trained language models: $P_q(q)$ on query corpus and $P_r(r)$ on response corpus.
Output: $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$

```
Randomly initialize G_{θqr}, G_{θrq}, D_{φqr}, D_{φrq}.
Pre-train G_{θqr} and G_{θrq} using MLE.
Pre-train D_{φqr} and D_{φrq} by Eq. 1.
while models have not converged do
    for i = 1, ··· , d do
        Update D_{φqr} and D_{φrq} by Eq. 1.
    end for
    for j = 1, ··· , g do
        Sample ⟨q, r⟩ from real-world data.
        Update G_{θqr} by ∇_{θqr} G_{θqr} in Eq. 5.
        Teacher Forcing: update G_{θqr} with ⟨q, r⟩
        Update G_{θrq} by ∇_{θrq} G_{θrq} in Eq. 5.
        Teacher Forcing: update G_{θrq} with ⟨q, r⟩
    end for
end while
```

of DAL is presented in Algorithm 1. Firstly, we use maximum likelihood estimation to pre-train $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$. Analogously, $D_{\phi_{qr}}$ and $D_{\phi_{rq}}$ are also pre-trained according to Eq. 1. After the pre-training phase, each generator is optimized by both duality constraint and adversarial signal, followed with the regularization of Teacher Forcing. The corresponding discriminators are simultaneously optimized.

5 Experiments

5.1 Experimental Settings

Baselines In order to verify the performance of DAL, we compare the following methods: *Seq2Seq:* the standard Seq2Seq model (Sutskever et al., 2014). *MMI-anti:* the mutual information method (Li et al., 2016), which uses an anti-language model in inference. *MMI-bidi:* the mutual information method (Li et al., 2016), which first generates a N-best response set with $p(r|q)$ and then reranks this response set with $p(q|r)$ in inference. *Adver-REIN:* the adversarial method adopting REINFORCE algorithm (Li et al., 2017). *GAN-AEL:* the adversarial method with an approximate embedding layer to solve the non-differentiable problem (Xu et al., 2017). *DAL-Dual (ours):* DAL trained only with maximum likelihood (Teacher Forcing) and duality constraint ($\nabla_{\theta_{qr}}\Upsilon$ or $\nabla_{\theta_{rq}}\Upsilon$). *DAL-DuAd (ours): DAL-Dual* with adversarial learning (Algorithm 1).

Both *DAL-Dual* and *DAL-DuAd* are methods proposed by us: the former incorporates the dual signal only, while the later combines the dual signal and the adversarial signal. In *DAL-Dual*, the guidance of each generator can be formulated as

$$\nabla_\theta G_\theta = \nabla_\theta \text{MLE} + \lambda_{dual} \cdot \nabla_\theta \Upsilon,$$

where ∇_θMLE is the guidance from teacher forcing and $\nabla_\theta\Upsilon$ is the guidance from the duality constraint. In *DAL-DuAd*, the guidance of each generator can be formulated as

$$\nabla_\theta G_\theta = \nabla_\theta \text{MLE} + \lambda_{dual} \cdot \nabla_\theta \Upsilon + \lambda_{gan} \cdot \nabla_\theta J(\theta),$$

where $\nabla_\theta J(\theta)$ is the adversarial signal.

Experimental Settings A Sina Weibo dataset (Zhou et al., 2017) is employed to train the models. We treat each query-response pair as a single-turn conversation. Attention mechanism (Luong et al., 2015) is applied in all the methods to enhance the performance. All the methods are implemented based on the open source tools Pytorch(Paszke et al., 2017) and OpenNMT (Klein et al., 2017). 1,000,565 query-response pairs are employed as the training data, 3,000 pairs as the validation data. The test data is another unique 10,000 query-response pairs. The length of all the dialogue utterances in the training corpus ranges from 5 to 50. Batch size is set to 64. The vocabulary size is set to 50,000. The dimension of word embedding is set to 500. All the methods adopt a beam size of 5 in the decoding phase. The maximum length of the target sequence is set to 50. Gradient clipping strategy is adopted when the norm exceeds a threshold of 5. There are 2 fully-connected layers (1000*500, 500*1) in the discriminator structure of *DAL-DuAd*. The vanilla *Seq2Seq*, *MMI-anti* and *MMI-bidi* use SGD as the optimizer, whose initial learning rate is 1.0. *Adver-REIN*, *GAN-AEL*, *DAL-Dual*, and *DAL-DuAd* use Adam (Kingma and Ba, 2014) as the optimizer, whose initial learning rate is 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Both Adam and SGD used in all the methods adopt a decay rate of 0.5 after the 8th epoch. The dropout (Srivastava et al., 2014) probability is set to 0.5. λ_{dual} is set

to 0.025 for both $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$. λ_{dual} is set to 0.025 and λ_{gan} is set to 1 for both $G_{\theta_{qr}}$ and $G_{\theta_{rq}}$. In Algorithm 1, d is set to 1 and g is set to 5. In *MMI-bidi*, the size of the N-best list is set to 5. In *MMI-anti*, γ is set to 0.15 and λ is set to 0.3.

5.2 Experimental Results

We firstly evaluate DAL on the task of generating of diverse responses. Then we resort to human annotators to evaluate the overall quality of the generated responses. Finally, we present several cases generated by all the involved method.

Response Diversity DISTINCT is a well-recognized metric to evaluate the diversity of the generated responses (Li et al., 2016; Xing et al., 2017). In our experiment, we employ DISTINCT-1 and DISTINCT-2, which calculate distinct unigrams and bigrams in the generated responses respectively. Table 1 presents the results of the five methods.

Method	DISTINCT-1	DISTINCT-2
Seq2Seq	0.031	0.137
MMI-anti	0.033	0.141
MMI-bidi	0.034	0.143
Adver-REIN	0.036	0.145
GAN-AEL	0.038	0.149
DAL-Dual (*ours*)	**0.052**	**0.209**
DAL-DuAd (*ours*)	**0.049**	**0.201**

Table 1: Results of diversity evaluation.

From Table 1, we have the following observations: (1) Both *MMI-anti* and *MMI-bidi* slightly improve the performance as compared with *Seq2Seq*. *MMI-bidi* heavily relies on the diversity of the N-best response set generated by $p(r|q)$. When N is not large enough to include some infrequently-occurring responses into the optional set, this set may lack diversity, and thus the ultimate response obtained with the reranking strategy also lacks diversity. However, when N is large, some responses having low coherence with the given query will be included in the optional set, and such responses may be selected as the final response, which hurts the performance of *MMI-bidi*. Therefore, the selection of N is an arduous task. *MMI-anti* also heavily relies on the anti-language model to obtain diverse responses. (2) Compared with *Seq2Seq*, our *DAL-Dual* improves diversity by 67.7% measured by DISTINCT-1 and 52.6% measured by DISTINCT-2, which reveals the effectiveness of the dual approach in improving diversity. (3) As expected, compared with *Adver-Rein* and *GAN-AEL*, our *DAL-DuAd* further improves the diversity of the generated responses. This observation proves our assumption that, with the guidance of discriminators $D_{\phi_{qr}}$ and $D_{\phi_{rq}}$, the generator $G_{\theta_{rq}}$ is able to influence the generator $G_{\theta_{qr}}$ to produce more diverse responses. We do notice that *DAL-Dual* achieves slightly better performance than *DAL-DuAd* on diversity. The reason is that sometimes adversarial methods tend to generate some short but quality responses such as "Let's go!" for given queries such as "We can have dinner together tonight. " or "There is an exhibition at the National Museum.". However, this short but natural response would harm diversity.

Response Quality Since the word overlap-based metrics such as BLEU (Papineni et al., 2002) and embedding-based metrics are inappropriate for response quality evaluation due to their low correlation with human judgment (Liu et al., 2016; Mou et al., 2016), we resort to human annotators to evaluate the overall quality of the generated responses. We employ 3 annotators to evaluate the quality of 200 responses generated from each of the aforementioned methods. **2**: the response is natural, relevant and informative. **1**: the response is appropriate for the given query but may not be very informative. **0**: the response is completely irrelevant, incoherent or contains syntactic errors. The final score for each response is the average of the scores from all the annotators. The human evaluation results are listed in Table 2.

Method	Human rating	Kappa
Seq2Seq	0.470	0.56
MMI-anti	0.568	0.46
MMI-bidi	0.523	0.60
Adver-REIN	0.767	0.49
GAN-AEL	0.758	0.52
DAL-Dual (*ours*)	0.730	0.47
DAL-DuAd (*ours*)	**0.778**	0.50

Table 2: Results of human elevation: response quality.

The agreement among annotators is calculated with Fleiss' kappa (Fleiss, 1971). The agreement ratio is in a range from 0.4 to 0.6, showing moderate agreement. Based on the results, we have the following observations: (1) *DAL-DuAd* achieves the highest quality score, indicating that our *DAL-DuAd* has the ability to produce coherent and informative responses. (2) *Adver-REIN* and *GAN-AEL* also obtain fairly good pointwise scores. This is because the adversarial learning mechanism effectively guides the generated responses to be close to the human-generated responses. (3) Com-

		Chinese	English (translated)
First case	Query	我无聊的把鼠标垫的内容一字不落的看完了	I feel so bored that I read all the words printed on the mouse pad.
	Seq2Seq	不要这样嘛！	Come on!
	MMI-anti	哈哈哈哈哈哈！	Haha Haha Haha!
	MMI-bidi	不错不错！	Good, good!
	Adver-REIN	没事没事！	It's nothing!
	GAN-AEL	哈哈。。。	Haha ...
	DAL-Dual	你太可爱了！	You are so cute!
	DAL-DuAd	我也是这么想的.	I also catch such an idea.
Second case	Query	昨天刚看了《等风来》，想去下尼泊尔，这个神奇的国度.	I watched the movie *Up in the Wind* yesterday. It inspires me to visit Nepal, such an amazing country.
	Seq2Seq	我也想去，可是没去过.	I want to go, but I haven't.
	MMI-anti	不错啊！	Good!
	MMI-bidi	真的假的???	Seriously???
	Adver-REIN	我也想去，可是没去过.	I want to go, but I haven't.
	GAN-AEL	也我就怕语言问题.	Also I am concern about the language.
	DAL-Dual	真的很神奇！	It's really amazing!
	DAL-DuAd	好神奇的国度!	What an amazing country!

Figure 3: Case study.

pared with *Seq2Seq*, *MMI-anti* and *MMI-bidi*, our *DAL-Dual* obtains relatively satisfactory performance on overall quality. It shows that the dual signal can also improve the overall quality.

Case Study We present several cases in Figure 3. For the first case involving *the content on the mouse pad*, most of the baselines generate generic responses such as"Come on!", "Haha!" or "It's nothing!". On the contrary, our *DAL-Dual* and *DAL-DuAd* method produce much more diverse and informative responses, such as "You are so cute!" and "I also catch such an idea.". These two entertaining responses are also topically coherent and logically consistent with the given query. In the second cases, our methods are also capable of capturing the topic *amazing country* shown in the query, and well generate the diverse and coherent responses following the topic of the query, such as "What an amazing country!" or "It is really amazing!". In contrast, the baselines still tend to provide safe responses lacking diversity to different queries.

5.3 Comparison of Efficiency

Efficiency is a crucial factor for real-life applications such as online chatbots. We conduct an experiment to evaluate the efficiency of all the methods under study. The efficiency experiment is conducted ten times on one Tesla K40m GPU whose memory is 11471M. The average time consumed by each method to generate the responses for 1000 queries is reported in Figure 4. *MMI-bidi-5*, *MMI-bidi-10* and *MMI-bidi-20* denote the *MMI-bidi* method with the N-best size of 5, 10 and 20 respectively. We can see that *MMI-anti* and *GAN-AEL* are the most time-consuming in all the baselines. Besides, we note that *MMI-bidi* method with the reranking strategy, even with a relatively small N-best size of 5, consumes much longer time than our methods, which severely limits its *MMI-bidi*'s application in practice. However, *Seq2Seq*, *Adver-REIN*, *DAL-Dual* and *DAL-DuAd* have very similar efficiency performance. Compared with *Seq2Seq* and *Adver-REIN*, *DAL-Dual* and *DAL-DuAd* achieve much better performance on diversity and overall quality. Therefore, DAL is more suitable for real-life applications.

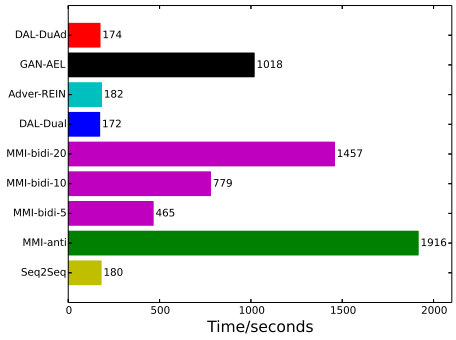


Figure 4: Time consumed by different methods.

6 Conclusion

We propose a novel framework named DAL to alleviate two prominent problems (safe responses and unnatural responses) plaguing dialogue generation. The dual learning proposed in this paper is the first effort to utilize the reverse dependency between queries and responses to reduce the probability of safe response generation and improve the diversity of the generated responses. Adversarial learning makes the generated responses as natural to human-generated ones as possible. DAL seamlessly integrates dual learning and adversarial learning, which are complementary to each other. Experimental results show that DAL achieves better performance than the state-of-the-art methods in terms of diversity, overall quality and efficiency.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NIPS*, pages 2672–2680.

Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tieyan Liu, and Wei-Ying Ma. 2016. Dual learning for machine translation. In *NIPS*, pages 820–828.

Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional neural network architectures for matching natural language sentences. In *NIPS*, pages 2042–2050.

Zongcheng Ji, Zhengdong Lu, and Hang Li. 2014. An information retrieval approach to short text conversation. *arXiv preprint arXiv:1408.6988*.

Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. 2017. Learning to discover cross-domain relations with generative adversarial networks. In *ICML*, pages 1857–1865.

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

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush. 2017. Opennmt: Open-source toolkit for neural machine translation. *Proceedings of ACL 2017, System Demonstrations*, pages 67–72.

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. In *NIPS*, pages 4601–4609.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *NAACL-HLT*, pages 110–119.

Jiwei Li, Will Monroe, Tianlin Shi, Sèbastien Jean, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. In *EMNLP*, pages 2157–2169.

Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *EMNLP*, pages 2122–2132.

Zhengdong Lu and Hang Li. 2013. A deep architecture for matching short texts. In *NIPS*, pages 1367–1375.

Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. In *EMNLP*, pages 1412–1421.

Lili Mou, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. 2016. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. In *COLING*, pages 3349–3358.

Aaron van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. 2016. Conditional image generation with pixelcnn decoders. In *NIPS*, pages 4790–4798.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL*, pages 311–318.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *NIPS-W*.

Jinhua Peng, Zongyang Ma, Di Jiang, and Hua Wu. 2019. Integrating bayesian and neural networks for discourse coherence. In *Companion Proceedings of the 2019 World Wide Web Conference*. International World Wide Web Conferences Steering Committee.

Alan Ritter, Colin Cherry, and William B Dolan. 2011. Data-driven response generation in social media. In *EMNLP*, pages 583–593.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *ACL*, volume 1, pages 1577–1586.

Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In *NAACL-HLT*, pages 196–205.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *NIPS*, pages 3104–3112.

Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*, pages 1057–1063.

Duyu Tang, Nan Duan, Tao Qin, and Ming Zhou. 2017. Question answering and question generation as dual tasks. *arXiv preprint arXiv:1706.02027*.

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

Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. 2013. A dataset for research on short-text conversations. In *EMNLP*, pages 935–945.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. 2017. Topic aware neural response generation. In *AAAI*, pages 3351–3357.

Zhen Xu, Bingquan Liu, Baoxun Wang, SUN Chengjie, Xiaolong Wang, Zhuoran Wang, and Chao Qi. 2017. Neural response generation via gan with an approximate embedding layer. In *EMNLP*, pages 617–626.

Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. 2017. Dualgan: Unsupervised dual learning for image-to-image translation. In *ICCV*, pages 2868–2876. IEEE.

Hainan Zhang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng. 2018. Reinforcing coherence for sequence to sequence model in dialogue generation. In *IJCAI*, pages 4567–4573.

Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2017. Emotional chatting machine: Emotional conversation generation with internal and external memory. *arXiv preprint arXiv:1704.01074*.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, pages 2242–2251. IEEE.