# Matching-based Generative Adversarial Network for Neural Dialogue Generation

**Hengyi Cai[1,2,4], He Bai[1,3,4], Songfang Huang[1]**
[1]IBM Research
[2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[3]National Laboratory of Pattern Recognition, Institute of Automation, CAS, Beijing, China
[4]University of Chinese Academy of Sciences, Beijing, China
caihengyi@ict.ac.cn, he.bai@nlpr.ia.ac.cn, huangsf@cn.ibm.com

## Abstract

Sequence-to-sequence(seq2seq) approach has obtained great attention in the field of open-domain dialogue generation. However, the standard seq2seq model is prone to generate common responses lacking meaningfulness and coherence, making the human-machine conversations clearly distinguishable from the human-human conversations. This coherence includes both the internal consistency of the long response and the external relevance between the post and its response. In this work, we propose a novel dialogue generation framework called Matching-based Generative Adversarial Network(MatchGAN) to improve the external relevance. Instead of imitating the ground truth by maximum likelihood estimation with supervised learning, this model can generate post-relevant responses through the generative-adversarial learning with a seq2seq-based generator and a matching-based discriminator, allowing the generated human-machine conversation much more like a human-human conversation by discriminating whether a response is matching with the post. Furthermore, we introduce the target-side attention mechanism to maintain the internal consistency of the generated responses. This new framework is able to generate coherent responses with high quality. Experiments show that our model achieves a substantial improvement on both metric-based and human evaluations among various baselines.

## Introduction

Building a human-machine dialogue system that can generate conversations indistinguishable from the human-human dialogues is a long-cherished goal. Recently, variants of sequence-to-sequence(seq2seq) architecture have been successfully applied for building such chat-bots(Serban et al. 2016; Shang, Lu, and Li 2015; Li et al. 2016b). However, these methods for dialogue generation are still far from desirable. In particular, these models tend to produce short, boring responses (e.g., “I don’t know”, “Yes, yes”) or long, inconsistent sentences (e.g., “I love to cook, I’m a vegan so I don’t like cooking”) (Sordoni et al. 2015; Li et al. 2016a). Although these responses are fluent and grammatical, the chat-bots look still very unattractive as they always seem to talk to themselves no matter what humans talk to them.



These issues are partly due to the fact that even to the same post utterance, there might be multiple appropriate responses with different semantics and speaking styles in the conversational task, namely the 1-to-N relation between the post and its response(Zhou et al. 2017). This is different from the translation task whose source input and target output are semantically equivalent, and results in a higher conditional entropy of the target distribution $P(Y|X)$ in the conversational setting. Additionally, the seq2seq model is prone to degenerating into language model that pay little attention to the post utterance(Wang et al. 2017) (similar observations that have also been noted in other tasks, such as story generation(Lewis, Dauphin, and Fan 2018)). These findings indicate that the seq2seq model together with the maximum likelihood objective function is not adequate to capture the relation between the post and its response, which is critical for a chat-bot to create a conversation like the human-human dialogue.

Very recently, Zhang et al. (2018) attempt to reinforce the relevance between the post and response by promoting their semantic similarity. However, the post and its corresponding response are not necessarily to be similar, and modeling the complex and under-specified relation between the post and its response with the semantic similarity is inadequate. There is still no competent method for modeling such implicit relation in dialogue generation. Besides, previous works pay little attention to the internal consistency of the generated sentences, which is critical for avoiding the generation of self-contradictory or segment-duplicated responses.

In this paper, we concentrate on improving both the **internal consistency** of a response and the **external relevance** between the response and its post during conversational learning. More specifically, for the external relevance, we propose a novel dialogue generation framework named Matching-based Generative Adversarial Network(MatchGAN). Unlike some existing GAN-based response generation models(Li et al. 2017; Xu et al. 2017) in which the discriminator is designed as a binary classifier to distinguish the synthetic from the real response, we design a matching-based network to learn the representation of the post-response’s relation and discriminate whether a response is matched with the given post. Here we assume the post-response from the human-human conversations is a matching pair while the ones from human-machine conversations are not. Given a post and the generated response, our discriminator can output the matching degree of these two sentences and then, the matching degree can be used to guide the generator to generate a more context-aware and meaningful response given such post. Moreover, our discriminator has the ability to explicitly leverage *1-to-N* dialogue corpus for training, which is the potential benefit of the proposed framework (detailed in the experimental part). As for the internal consistency, we introduce the target-side attention mechanism into the generator used in this framework to better maintain the internal consistency of utterances during responses generation. The proposed matching-based discriminator is used to reinforce the generator for context-aware (external relevance) responses modeling and the target-side attention mechanism is introduced for generating more semantic-consistent (internal consistency) responses.

In experiments, we evaluate our models on three large-scale human-human conversation datasets, i.e., Baidu-Tieba, StackOverflow and PERSONA-CHAT. Empirical results show that our methods outperform current popular models with respect to both metric-based evaluations and human judgments. Our main contributions are three folds: (1) We propose a novel approach called MatchGAN for dialogue generation, which is able to learn the implicit relation between the post and its corresponding reply for context-aware response generation. (2) We introduce the target-side attention mechanism into the dialogue generator to better maintain the internal consistency of the generated response utterances. (3) We perform the extensive empirical study on three dialogue datasets with different scales and characteristics. The proposed methods outperform five representative existing approaches in both metric-based and human evaluations.

## Related Work

Response generation for open-domain dialogue can be viewed as a source-to-target transduction problem. Ritter, Cherry, and Dolan (2011) first model the generation of conversations as a statistical machine translation(SMT) problem in an end-to-end and purely data-driven fashion. Recently, the promising results of applying deep learning in machine translation and language modeling further spur the enthusiasm of researches in neural generative dialogue systems (Shang, Lu, and Li 2015; Vinyals and Le 2015; Serban et al. 2016; Chen et al. 2018).

However, Zhou et al. (2017) point out that the task of conversation modeling with large degree of response diversity is much more challenging than machine translation which estimates the probability of a target language sentence conditioned on the source language sentence with the same meaning. To tackle the issue of dull response, Li et al. (2016a) use Maximum Mutual Information (MMI) as the objective function in neural models to reduce the proportion of generic responses. Li et al. (2016c) detect entities from contextual information and search for relevant entities in a large knowledge base to provide information clues for content introducing. Yao et al. (2017) explore an implicit content introducing method by incorporating an additional cue word using the cue word gated recurrent unit(GRU) and the fusion unit for generative short-text conversation system. For coherent dialogue generation, Zhang et al. (2018) use the semantic similarity between the post and response as reward to do policy gradient on seq2seq model. Our work generalizes this conventional notion of similarity and aims to model the matching relations between the post and its response. Mei, Bansal, and Walter (2017) propose to use the dynamic attention model to promote coherence of the generated responses. Bosselut et al. (2018) propose to learn neural rewards to model cross-sentence ordering as a means to approximate desired discourse structure for dialogue generation. Another category of related work is dialogue generation with generative adversarial networks (Li et al. 2017; Xu et al. 2017). While the implicit interactions between the post and response are not considered in these works, our work designs a novel discriminator to measure these relations in an adversarial fashion. As for the target-side attention mechanism, Xia et al. (2017) introduce target-target attention model to enhance the semantic adequacy in neural machine translation and image caption. However, they do not fully investigate it in the dialogue generation setting. Shao et al. (2017) incorporate target-side attention into neural conversation models by data manipulating. In this work, we explore a more natural way to incorporate target-side attention into the generation process without any additional data manipulation.

## Methodology

### Overview

MatchGAN learns the model from the matching patterns between generated responses and post utterances in a generative adversarial framework. The sketch of MatchGAN is illustrated in Figure 1, which consists of a generator for generating responses and a matcher for discriminating the matching degree of the generated text and post utterance. Although there are many possible instantiations of the generator and matcher for our framework, in this work, we purposefully choose a simplistic architecture, leaving more sophisticated modeling for future work. Specifically, the generator is a seq2seq model with both source-side and target-side attention, and the matcher is responsible for measuring matching degrees of posts and generated responses. Parameters of the matcher are learned jointly with the generator instead of separate training.

### Generator with Target-side Attention

The backbone of the generator $G$ is a seq2seq model, which takes the input sentence $X = \{x_1, x_2, \cdots, x_{T_X}\}$ of $T_X$ words from vocabulary $\Gamma$, and generates the output sequence $Y = \{y_1, y_2, \cdots, y_{T_Y}\}$ of $T_Y$ words from vocabulary $\Lambda$. Formally, the encoder network sequentially reads the word in X and encodes it as a context vector $c$ through a recurrent neural network(RNN). The decoder network sequentially generates a reply $Y$ with context vector $c$ as input. Typically, LSTM-based(Hochreiter and Schmidhuber 1997) seq2seq framework models the conditional probability of the

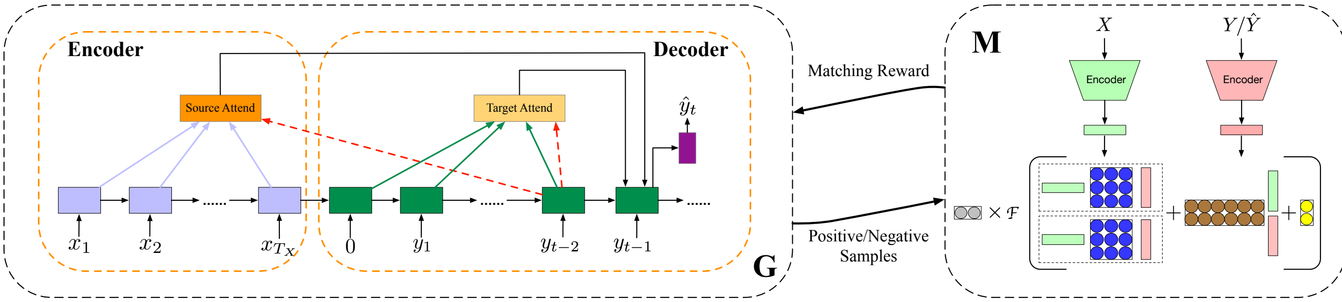


Figure 1: An illustration of the proposed MatchGAN. Given a post sequence $X$=$\{x_1, x_2, \cdots, x_{T_X}\}$, the generative network(G) samples each word $\hat{y}_i$ of the generated response $\hat{Y}$ using both source-side and target-side attention mechanisms. The matching-based discriminator($M$) then takes the input $\langle X, Y\rangle$ and $\langle X, \hat{Y}\rangle$ as the positive and negative samples respectively. The matching reward given by $M$ is used to guide the generator learning. The overall framework is trained in the adversarial paradigm using reinforcement learning.

target sequence as follows:

$$\begin{aligned} P(Y|X) &= \prod_{t=1}^{T_Y} P(y_t|y_{[0:t-1]}; X) \\ &= \prod_{t=1}^{T_Y} P(y_t|y_{t-1}, h'_{t-1}, c), \end{aligned} \tag{1}$$

where $h'_{t-1}$ is the hidden state of $y_{t-1}$ output by LSTM.

To better model the conditional language model $P(Y|X)$, the attention mechanism(Bahdanau, Cho, and Bengio 2014) is used to enable the decoder to "attend" to different parts of the source sentence at each step of the output generation. To put it formally, LSTM-based seq2seq models with attention parameterize the per-symbol conditional probability as:

$$\begin{aligned} P(y_t|y_{[0:t-1]}; X) = \mathrm{LSTM}(&y_{t-1}, h'_{t-1}, \\ &\mathrm{Attention}(h'_{t-1}, \mathbf{x})), \end{aligned} \tag{2}$$

where Attention() is a function which uses $h'_{t-1}$ to address the most relevant parts in $X$ and then outputs the summary of encoder symbols $\mathbf{x}$. Typically, $\mathrm{Attention}(h'_{t-1}, \mathbf{x}) = \sum_{i=1}^{T_X} a_{t,i} h_i$, where $a_{t,i}$ is the attention weight over $h_i$ at time $t$. $a_{t,i}$ is usually defined as:

$$e_{t,i} = g(h'_{t-1}, h_i); \quad a_t = softmax(e_t), \tag{3}$$

where $h_i$ is the hidden state of $x_i$. $g$ is a function that calculates the similarity between $h_i$ and $h'_{t-1}$. In this work we use bilinear function as $g(h'_{t-1}, h_i) = v^T tanh(W_h h_i + W_{h'} h'_{t-1})$, where $v$, $W_h$ and $W_{h'}$ are parameter matrices.

The aforementioned source-side attention has shown the promising performance when applied to tasks that can be cast as sequence-to-sequence learning, including machine translation. However, responding to conversations is much more different with translation. In machine translation, the semantics of the target sequence $Y$ is almost entirely determined by the given source sequence $X$, while the reply generated by the agent in a conversational setting should not only be consistent with the given post utterances, but also be consistent with what has been responded for semantic consistency and content readability. Therefore, the target-side attention is proposed to augment the attention mechanism by taking into account the generated target sequence at each step of the decoding process. As illustrate in the left part of Figure 1, at time step $i-1$, the decoder in generator receives an input token embedding $y_{i-1}$, previous hidden state $h'_{i-1}$, $\mathrm{Attention}(h'_{i-1}, \mathbf{x})$ and $\mathrm{Attention}(h'_{i-1}, \mathbf{y}_{<j-1})$, i.e., the output distribution at time step $i-1$ follows:

$$\begin{aligned} P(y_i|y_{[0:i-1]}; X) = \mathrm{LSTM}(&y_{i-1}, h'_{i-1}, \\ &\mathrm{Attention}'(h'_{i-1}, \mathbf{x}, \mathbf{y}_{<i-1})), \end{aligned} \tag{4}$$

where $\mathbf{y}_{<i-1}$ represents the summary of the decoder symbols before time step $i-1$. In this work, we implement the $\mathrm{Attention}'()$ as the weighted sum of hidden states generated before time step $i-1$ similar to $\mathrm{Attention}()$.

## Matching-based Discriminator

We utilize a siamese network consisting of two RNNs with a neural tensor layer to model the interactions between the post and response. As shown in the right part of Figure 1, the first RNN (Post-RNN) is used to encode the given post, and the second RNN (Response-RNN) is used to encode the response(generated response or target response). The final hidden state of each RNN represents a summary of the input sentence and is used as the input for the neural tensor layer.

In this work, we model the matching degree of the post and response with a non-linear tensor layer, which has been successfully applied to explicitly model multi-dimensional interactions of data (Qiu and Huang 2015; Socher et al. 2013). Given a post $X$ and its corresponding response $Y$, we use two RNNs mentioned above to model them into the fixed vector $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\mathbf{y} \in \mathbb{R}^{n_y}$ respectively. Following (Socher et al. 2013), the neural tensor layer utilizes a bilinear tensor layer to directly relates $\mathbf{x}$ and $\mathbf{y}$ across multiple dimensions. To put it formally, the tensor layer computes a matching score by the following function:

$$m(\mathbf{x}, \mathbf{y}) = \mathbf{u} f(\mathbf{x}^T \mathbf{W}^{[1:d]} \mathbf{y} + \mathbf{V} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \mathbf{b}), \tag{5}$$

where $f$ is a nonlinearity applied element-wise, $\mathbf{W}^{[1:d]} \in \mathbb{R}^{n_x \times n_y \times d}$ is a tensor and the bilinear tensor product

$\mathbf{x}^T \mathbf{W}^{[1:d]} \mathbf{y}$ results in a vector $h \in \mathbb{R}^d$, where each entry is computed by one slice $i = 1, \cdots, d$ of the tensor: $h_i = \mathbf{x}^T \mathbf{W}^i \mathbf{y}$; The other parameters are the standard form of a neural network: $\mathbf{V} \in \mathbb{R}^{d \times (n_x + n_y)}$, $\mathbf{b} \in \mathbb{R}^d$ and $\mathbf{u} \in \mathbb{R}^d$. The right part of Figure 1 shows a visualization of the proposed matcher $M$.

The main advantage of the proposed matcher is that it is able to model the sentence representations and interactions jointly. The representations of post and response are modeled by the Post-RNN and Response-RNN respectively, and the interactions between the post and its response are modeled by the neural tensor layer. The final output of the matcher $M$ is the matching score of post $X$ and response $Y$. Other more sophisticated matching models including convolutional neural network architectures(Hu et al. 2014) and self-attentive matching network(Wu et al. 2018) are also able to be applied in this framework directly for learning the interaction between post and response. This must be left for future work.

To encourage the generator to generate context-aware responses, the matcher $M$ is required to assign higher reward to the real response with the given post and lower reward to the generated ones. More specifically, we use the following loss function to train the matcher $M$:

$$\begin{aligned} \mathcal{L} = -\mathbb{E}_{X \sim \mathcal{P}_h, Y \sim \mathcal{P}_h}[M_\phi(\{X, Y\})] + \\ \mathbb{E}_{X \sim \mathcal{P}_h, Y \sim G_\theta}[M_\phi(\{X, Y\})], \end{aligned} \tag{6}$$

where $\theta$ and $\phi$ are the variable parameters in $G$ and $M$, respectively. $X$ is the post utterance and $Y$ is the corresponding response. The $\mathbb{E}$ is the expectation operator, and $\mathcal{P}_h$ is the real data from human-written utterances. $Y \sim \mathcal{P}_h$ and $Y \sim G_\theta$ denote that $Y$ is from human-written responses and machine-generated responses, respectively.

## Policy Gradient Training

We use policy gradient method(Williams 1992) to train the generator $G$ for maximizing the expected reward of generated responses:

$$J(\theta) = \mathbb{E}_{Y \sim G_\theta}(M_\phi(\{X, Y\})|\theta). \tag{7}$$

Given the input post $X$, the agent generates a response $Y$ by sampling from the policy $G_\theta$. The post $X$ and the generated response $Y$ are then fed to the matcher $M_\phi$. The gradient of Eq.(7) is approximated as follows:

$$\begin{aligned} \nabla_\theta J(\theta) \approxeq [M_\phi(\{X, Y\}) - b(\{X, Y\})] \nabla \log G_\theta(Y|X) \\ = [M_\phi(\{X, Y\}) - b(\{X, Y\})] \nabla \sum_{t=1}^{T_Y} \log P(y_t | y_{[0:t-1]}; X), \end{aligned} \tag{8}$$

where $G_\theta(Y|X)$ denotes the probability of the generated response $Y$ given post utterance $X$. $b(\{X, Y\})$ denotes the baseline value to reduce the variance of the estimate while keeping it unbiased[1]. The matcher $M$, which takes target response paired with post $X$ as the positive sample and machine-generated response paired with post $X$ as the negative sample, is simultaneously updated with the generator $G$ in an adversarial fashion.

One drawback of the policy gradient method described above is that the generator only gets the reward when the complete response has been generated, yielding the problem of sparse training signals. The reward mechanism is expected to give separate rewards for each token during response generation in the training phase for better credit assignment. We use Monte Carlo Search(MCS) to achieve such a goal. More specifically, for a sequence of length $T$, to evaluate the action-value for a word at the time step $t$, MCS with a rollout policy $G_\theta$ is applied to sample the unknown last $T - t$ tokens for $N$ times. These $N$ complete sequences shared with a common post $X$(i.e., $N$ pairs of $\{$post, response$\}$) are fed to the matcher $M_\phi$, then the average match score of these samples is used as a reward for the token at time step $t$. The downside of MCS is that the repeated sampling process is significantly time-consuming which becomes worse with the longer sequence. One way to mitigate this problem is to dynamically adjust $N$ with $t$: $N = \lceil N_0 \times (1 - \frac{t-1}{T}) \rceil$. $N_0$ is set to 10 in this work.

During the adversarial training of MatchGAN, we found that gradient vanishing occurs when the matcher $M$ is much stronger then the generator $G$, i.e., the value of reward is too small to update the generator $G$. Inspired by ranking idea from RankGAN(Lin et al. 2017) and the rescaled activation in (Guo et al. 2018), we adopt the similar technique to rescale the reward before it is being fed into $G$. For a mini-batch with $B$ responses generated by $G$, the reward matrix is denoted as $R_{B \times T}$. For each time step $t$, we rescale the $t$-th column vector $R^t$ as follows:

$$R_i^t = \sigma(\delta \cdot (0.5 - \frac{rank(i)}{B})), \tag{9}$$

where rank($i$) is the $i$-th element's high-to-low ranking in current column vector. $\delta$ is a hyper-parameter that controls the smoothness of the rescale function. $\sigma(\cdot)$ is an activation function. We take $\delta = 12.0$ and the sigmoid function as $\sigma(\cdot)$ in this work.

# Experiments

## Datasets

The proposed methods are evaluated on three human-human dialogue datasets: one multi-turn Chinese dialogue dataset, one single-turn 1-to-N English conversation dataset and one multi-turn English chit-chat dataset, i.e., Baidu-Tieba, StackOverflow and PERSONA-CHAT.

**Baidu-Tieba** This massive Chinese dataset of human-human multi-turn conversations is crawled from the Baidu Tieba[2] forum. We use 1,000,000 ⟨post-response⟩ pairs for training, 5,000 for validation and 10,000 for testing. We kept

[1]Similar to (Li et al. 2017), we train another neural network model(i.e., the critic) to estimate the value(or future reward) of current state(hidden state of the last time step) under the current policy $G_\theta$. The critic network takes the current state as input and maps it to a scalar. The network is optimized based on the smooth L1 loss between the estimated reward and the observed reward.

[2]http://tieba.baidu.com

the length of each utterance in [6, 80] for better training. The vocabulary size of this dataset is 5,564.

**StackOverflow**[3] As mentioned earlier, the 1-to-N relation between a post and its response usually exists in human conversations. We provide a new StackOverflow conversation corpus which consists of almost 0.5 million 1-to-N ⟨post-responses⟩ extracted from public available Stack Overflow database[4] for evaluating the performance of proposed approaches under the 1-to-N conversational setting. We use 500,000 1-to-N ⟨post-responses⟩ pairs for training, 3,000 1-to-1 pairs for validation and 5,000 1-to-1 pairs for testing. We kept the length of each utterance in [10, 100] and the total number of distinct tokens is 24,012. There are 2.2 responses on the average of each post. When training on this dataset, the matching-based discriminator takes these $n$ pairs ⟨post-response⟩ of each 1-to-N conversation as positive samples and the 1-to-1 generated ⟨post-response⟩ as the negative sample. The intuition is that the higher the accuracy of the matching-based discriminator, the more likely it emits sound feedback(rewards) to the generator for context-aware response generation. To make the comparison fairer, all positive samples are used for baseline methods training.

**PERSONA-CHAT** This multi-turn dialogue dataset is introduced in (Kiela et al. 2018) which consists of 139,239 utterances between crowd-workers who were randomly paired and each asked to act the part of a given provided persona. In this work, we aim at response generation and thus do not consider the persona information by simply removing the persona text. We recommend readers refer to (Kiela et al. 2018) for detail. We kept 131,438 ⟨post-response⟩ utterances for training, 7,801 for validation and testing. The vocabulary size of this dataset is 9,293.

## Baselines

We compare our proposed model with the following five representative baselines:

**Seq2Seq-attn** A seq2seq model with attention proposed in (Bahdanau, Cho, and Bengio 2014). This model is trained with the traditional negative log-likelihood(NLL) loss.

**Seq2Seq-tgt-attn** A Seq2Seq-attn model which also attends on the target-side sequence. We build this baseline to investigate whether the proposed target-side attention mechanism is able to improve the quality of generated responses compared with Seq2Seq-attn.

**Adver-REGS** An adversarial strategy proposed in (Li et al. 2017)[5], which consists of a seq2seq model as the generator and an RNN-based binary classifier as the discriminator to evaluate the output sequence. The overall method is trained with a reinforcement learning framework.

**PG-BLEU** The PG-BLEU(Bahdanau et al. 2016) computes the BLEU(Papineni et al. 2002) score to measure the similarity between the generated response and the human-written response, then takes the BLEU score as the reward to update the generator with policy gradient. The advantage is that this model can directly optimize the task-specific score, i.e., BLEU. It is worthy noting that while PG-BLEU grasps the similarities depend on the n-grams matching from the token-level among sentences, MatchGAN explores the implicit interactions inside the embedded features of the post sentence and response sentence.

**MMI** Li et al. (2016a) proposed using Maximum Mutual Information(MMI) as the objective function to train the seq2seq model for diverse response generation. We use the backward seq2seq to model the $P(X|Y)$ used in MMI.

To better evaluate the performance of the proposed approaches, we introduce two variants of our model, denote as MatchGAN and MatchGAN-tgt, respectively. The generator used in MatchGAN-tgt is augmented with the target-side attention mechanism.

## Implementation Details

In experiments, we implement baseline methods and our models using the ParlAI(Miller et al. 2017) framework. We first give the details about input embeddings. For Baidu-Tieba, we use character-level embeddings rather than word-level embeddings, due to the word sparsity, segmentation mistakes and unknown Chinese words which lead to inferior performance than character-level(Hu, Chen, and Zhu 2015). For StackOverflow and PERSONA-CHAT, we use word embeddings trained by word2vec on a large Wikipedia corpus[6]. We then give the training details. We implemente the generator of MatchGAN using 2 LSTM layers on both the encoder and the decoder. Post-RNN and Response-RNN used in the matcher are also implemented using 2 LSTM layers. We first pre-train the generator by predicting target sequence given the post sequence, and then fine-tune overall model(generator and discriminator) until the specific metric does not get improvements on the validation datasets. For a fair comparison among all the baseline methods and our methods, the sizes of hidden state are all set to 256, and the embedding sizes of Baidu-Tieba, StackOverflow and PERSONA-CHAT are all set to 300. We use Adam optimization method(Kingma and Ba 2014) with initial learning rate of 0.001. Dropout rate of 0.2 is applied to prevent the model from over-fitting.

## Experimental Results

**Automatic Evaluation** Evaluating the performance of dialogue system is not a trivial task. For automatic evaluation of the proposed methods, we use metrics that can reflect the diversity, language fluency and word overlap with the ground truth responses, i.e., the number of distinct ngrams(*dist*), perplexity(PPL) and BLEU. Besides, we also use the metric of $R_C@k$, which is often used in language

[3]This dataset is not a strictly open-domain chit-chat corpus. Here we use this corpus to investigate whether the MatchGAN is able to benefit from the explicit 1-to-N relations embodied in the dataset.

[4]http://download.brentozar.com/StackOverflow2010.7z

[5]https://github.com/jiweil/Neural-Dialogue-Generation

[6]http://www.psych.ualberta.ca/∼westburylab/downloads/westburylab.wikicorp.download.html

tasks. Here the model is asked to select the $k$ most likely responses from a candidate set containing $C$ responses, and it is correct if the true response is among these k candidates. We hypothesize that if a model is perform better on the generation task, it will eventually lead to improvements for the ranking task. We use $R_{20}$@1 and $R_{20}$@5 in the experiments.

| Baidu-Tieba | PPL | $R_{20}$@1 | $R_{20}$@5 | BLEU | *dist-1* |
|---|---|---|---|---|---|
| Seq2Seq-attn | 44.26 | 7.63 | 33.4 | 11.03 | 0.3633 |
| Seq2Seq-tgt-attn | 43.77 | 7.63 | 33.7 | 11.62 | 0.3679 |
| Adver-REGS | 43.65 | 7.87 | 34.6 | 11.55 | 0.2343 |
| PG-BLEU | 44.11 | 7.82 | 34.3 | 11.52 | 0.2575 |
| MMI | 44.23 | 7.53 | 33.8 | 12.21 | 0.4561 |
| MatchGAN | 43.62 | 7.74 | 33.9 | 13.19 | 0.7766 |
| MatchGAN-tgt | **42.64** | **8.33** | **35.7** | **13.77** | **0.8157** |
| **StackOverflow** | **PPL** | **$R_{20}$@1** | **$R_{20}$@5** | **BLEU** | ***dist-1*** |
| Seq2Seq-attn | 52.54 | 8.48 | 37.3 | 11.21 | 0.4319 |
| Seq2Seq-tgt-attn | 50.97 | 9.1 | 39.9 | 12.57 | 1.176 |
| Adver-REGS | 49.1 | 9.04 | 42.4 | 11.07 | 0.6269 |
| PG-BLEU | 49.75 | 9.29 | 40.8 | 11.49 | 0.4392 |
| MMI | 53.35 | 7.73 | 35.7 | 11.68 | 1.535 |
| MatchGAN | 48.3 | **9.73** | 42.1 | 12.62 | **1.678** |
| MatchGAN-tgt | **48.03** | 9.35 | **42.6** | **13.32** | 1.307 |
| **PERSONA-CHAT** | **PPL** | **$R_{20}$@1** | **$R_{20}$@5** | **BLEU** | ***dist-1*** |
| Seq2Seq-attn | 37.28 | 9.27 | 33.9 | 16.41 | 0.4943 |
| Seq2Seq-tgt-attn | 32.55 | 9.60 | 34.2 | 16.99 | 0.7179 |
| Adver-REGS | 33.79 | 9.51 | 34.4 | 16.94 | 0.5999 |
| PG-BLEU | **32.28** | 9.11 | 34.4 | 17.36 | 0.5561 |
| MMI | 35.13 | 8.42 | 32.1 | 17.44 | 0.6231 |
| MatchGAN | 32.74 | 9.96 | 34.9 | 17.53 | **1.028** |
| MatchGAN-tgt | 32.75 | **10.1** | **35.2** | **17.62** | 0.8303 |

Table 1: Performance of MatchGAN, MatchGAN-tgt and five baselines on dialogue generation tasks(%). Higher is better except for PPL. *dist-1* is the number of distinct unigrams divided by total number of generated words.

The automatic evaluation results in Table 1 demonstrate the characteristics of each approach. One can see that Seq2Seq-tgt-attn outperforms Seq2Seq-attn baseline in terms of five metrics on all the three corpora, which demonstrates that the proposed target-side attention mechanism consistently improves the performance of dialogue generation tasks. We see that of our baseline models, MMI has higher *dist-1* on three response generation tasks, which is not surprising given that the model is trained using the diversity-promoting objective function. However, MMI has lower $R_{20}$@1 and $R_{20}$@5, indicating its inability to figure out the ground-truth responses from candidate set. As for *dist-1*, our models generate remarkably more distinct tokens than the baselines, indicating that our models is able to generate more diverse and informative responses compared to the baselines. Among the two proposed MatchGAN variants, MatchGAN-tgt performs best, because of its capability to learn the post-response interactions and the generator augmented with target-side attention jointly for coherent response modeling. It is noteworthy that MatchGAN yields a significant performance boost compared with Seq2Seq-attn on StackOverflow, with an $R_{20}$@5 score increase of up to 12.87% and a more than 200% jump in unigram diversity. The interpretation of this huge performance improvement is that the proposed matcher is able to benefit from the 1-to-N relations contained in the dataset and thus helps the generator to learn better. Our models obtain higher $R_{20}$@1, $R_{20}$@5, BLEU, *dist-1* and lower PPL than baseline models. These results demonstrate that the post-response matching patterns measured by our proposed discriminator as well as the target-side attention mechanism can benefit dialogue generation tasks.

**Human Evaluation** Besides the automatic evaluation, we also use human evaluation to better assess the quality of generated responses. Human evaluation rates the model in three aspects: the **matching degree with context**, **internal consistency** and **overall quality** of the generated responses. All these metrics range from 0 to 3, where 0 represents complete error, 1 for partially correct, 2 for almost correct and 3 for absolutely correct. Given 200 randomly sampled post utterances and their generated responses, four annotators are required to give judgments independently.

The human evaluation results are shown in Table 2. Among the various models, MatchGAN-tgt performs significantly better than the baselines when generating context-aware responses, which indicates that our model is capable of capturing the post-response matching patterns to generate more semantic-relevant sentences. We can also see that MatchGAN-tgt obtains particularly high scores in internal consistency. This demonstrates that the proposed approach is better to generate more semantic-consistent responses, and is able to match the post sentence at the same time. The pairwise Cohen's Kappa agreement scores are 0.49 on matching degree, 0.51 on internal consistency, and 0.58 on overall quality, which indicate a strong annotator agreement.

## Analysis

In this part, we use the PERSONA-CHAT corpus for analyzing how matching-based discriminator works in MatchGAN from quantity analysis. We then perform case study to evaluate the response quality of the baselines and the proposed approaches.

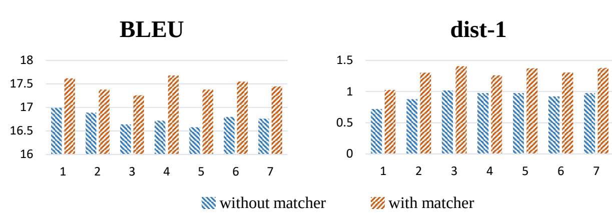


Figure 2: MatchGAN's performance comparison(%) on PERSONA-CHAT corpus across different contexts.

**Quantity Analysis** We study how MatchGAN performs in different utterance number of context. Figure 2 shows the changes of BLEU and *dist-1* on PERSONA-CHAT across contexts with different number of utterance. As demonstrated, the proposed matcher consistently improves the performance on both BLEU and *dist-1* compared with the counterpart for contexts having different utterance text length. In

| Model | Baidu-Tieba | | | StackOverflow | | | PERSONA-CHAT | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mat. | Int. | Ove. | Mat. | Int. | Ove. | Mat. | Int. | Ove. |
| Seq2Seq-attn | 1.11 | 1.18 | 1.11 | 1.14 | 1.04 | 1.1 | 1.85 | 1.77 | 1.81 |
| Seq2Seq-tgt-attn | 1.42 | 1.56 | 1.3 | 1.42 | 1.5 | 1.2 | 2.09 | 2.13 | 1.97 |
| Adver-REGS | 1.26 | 1.37 | 1.22 | 1.32 | 1.14 | 1.26 | 2.01 | 1.86 | 1.87 |
| PG-BLEU | 1.39 | 1.33 | 1.33 | 1.38 | 1.24 | 1.3 | 1.93 | 1.77 | 1.87 |
| MMI | 1.22 | 0.6 | 0.88 | 0.88 | 0.32 | 0.66 | 1.2 | 1.1 | 1.13 |
| MatchGAN | **2.03** | 1.99 | 1.84 | 1.44 | 1.44 | 1.38 | 2.14 | 2.22 | 2.15 |
| MatchGAN-tgt | 1.85 | **2.07** | **1.87** | **1.52** | **1.9** | **1.54** | **2.32** | **2.47** | **2.38** |

Table 2: Manual evaluation results on Baidu-Tieba, StackOverflow and PERSONA-CHAT for different judgment aspects. Mat., Int. and Ove. stand for matching degree with context, internal consistency and overall quality, respectively.

addition, model with the matcher outperforms its counterpart in terms of BLEU by a large margin for contexts having longer utterance text length(>3), implying the advantage of using matching-based discriminator in our framework.

| | |
|---|---|
| post | 个人觉得，上半场德佩是继叔叔之后第二大进攻威胁。*(Personally speaking, Depay is the second major threat after uncle in the first half.)* |
| response | 我也觉得孟菲斯除了那两脚毒了点，其他的还是可圈可点的。*(I also think that Memphis is still commendable except for those two vicious feet.)* |
| Seq2Seq-attn | 德佩的传球是什么？ *(What is Depe's pass?)* |
| Seq2Seq-tgt-attn | 这个赛季的表现不错, 而且他的位置感不错。*(He has performed well this season, besides, he has a good sense of position.)* |
| Adver-REGS | 我觉得他是个好球员，但是他不是一个人，而是他的位置。*(I think he is a good player, but he is not a person, but his position.)* |
| PG-BLEU | 德佩的表现不错。*(Depay did a good job.)* |
| MMI | 不好好看球啊。我看球没用啊。我说过他们不用他踢中卫。*(Why don't watch the game seriously? There is no need for me to watch the game. I said they didn't need him to play the central defender.)* |
| MatchGAN | 我觉得林佳德踢得挺好，毕竟上赛季踢得不错，毕竟他是铁腰，虽然很容易受伤。*(I think Lingard played very well. After all, he played very well last season, and he has an iron waist, although he was very vulnerable to injury.)* |
| MatchGAN-tgt | 我觉得马塔比他更适合打边后卫。*(I think Mata is a better full-back than he is.)* |
| post | *hank williams.... I grew up on country music.* |
| response | *I like country too! I am kinda tired tonight from a long day at the shipyard.* |
| Seq2Seq-attn | *I love it, I love it.* |
| Seq2Seq-tgt-attn | *I love the country, but I'm not a fan of them.* |
| Adver-REGS | *I've never heard of that. I've never heard of that.* |
| PG-BLEU | *I'm a UNK. I'm going to see the sunset.* |
| MMI | *Well what about that band. . lol ? ? !* |
| MatchGAN | *I'm a country fan too. I'm a country boy!* |
| MatchGAN-tgt | *I love to sing. I love the beatles.* |

Table 3: Different responses generated by the baselines and our models on Baidu-Tieba and PERSONA-CHAT.

**Case Study** Table 3 shows some generated responses for demonstration. We can see that (1) Compared with Seq2Seq-attn, responses generated by Seq2Seq-tgt-attn are more informative and internal-consistent. Adverbs and conjunctions(e.g., *besides*, *but*) used in the sentences are more appropriate and there is no contradiction between clauses. These examples show that the target-side attention can capture the fine-grained patterns in the response-side and improve the internal-consistency of responses. (2) MatchGAN and MatchGAN-tgt is able to generate meaningful and informative responses compared with baseline models. It is worthy noting that contents generated by our models usually contain cue entities which are well matched with the corresponding posts. For example, *Lingard* and *Mata* correspond to *Depay* in Baidu-Tieba(they are all professional footballers) and *the beatles* corresponds to *hank williams* in PERSONA-CHAT(they are all influential singers), indicating typical cases of **topic continuity** in our tasks, which means that our models are more likely to elicit further interactions. This also reflects our expectation: the proposed MatchGAN framework generalizes the conventional notion of similarity and aims to model the matching relations between the post and its response. (3) Our models are able to avoid inconsistence even facing with long generated responses compared with baseline methods. Adver-REGS and MMI generate longer responses compared with Seq2Seq-attn, however, their replies are prone to be self-contradictory and less-correlated with the given posts. Similar observations have also been obtained for many other posts, but we have to omit them for space limitations.

## Conclusion

This paper describes a novel framework for dialogue generation which aims at generating coherent responses in terms of both internal consistency and external relevance based on the paradigm of generative adversarial learning. We first introduce the target-side attention into the encoder-decoder generator to maintain the internal consistency of the generated responses. We then propose a matching-based discriminator to model the post-response interactions and use reinforcement learning to encourage the generator for meaningful and context-aware responses generation. Extensive experiments conducted on three large-scale datasets have verified the effectiveness of our proposed approaches by showing significant improvements over multiple baselines in terms of metric-based evaluations and human judgments, Note that, the implications of this work extend well to other tasks like question answering, image caption and potentially any task where mutual correspondences must be modeled. We would like to make further studies in our future work.

## References


Bahdanau, D.; Brakel, P.; Xu, K.; Goyal, A.; Lowe, R.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2016. An actor-critic algorithm for sequence prediction. *CoRR* abs/1607.07086.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. *CoRR* abs/1409.0473.

Bosselut, A.; Çelikyilmaz, A.; He, X.; Gao, J.; Huang, P.; and Choi, Y. 2018. Discourse-aware neural rewards for coherent text generation. In *NAACL-HLT*, 173–184.

Chen, H.; Ren, Z.; Tang, J.; Zhao, Y. E.; and Yin, D. 2018. Hierarchical variational memory network for dialogue generation. In *WWW*, 1653–1662.

Guo, J.; Lu, S.; Cai, H.; Zhang, W.; Yu, Y.; and Wang, J. 2018. Long text generation via adversarial training with leaked information. In *AAAI*.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780.

Hu, B.; Lu, Z.; Li, H.; and Chen, Q. 2014. Convolutional neural network architectures for matching natural language sentences. In *NIPS*, 2042–2050.

Hu, B.; Chen, Q.; and Zhu, F. 2015. LCSTS: A large scale chinese short text summarization dataset. In *EMNLP*, 1967–1972.

Kiela, D.; Weston, J.; Zhang, S.; Dinan, E.; Urbanek, J.; and Szlam, A. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *ACL*, 2204–2213.

Kingma, D. P., and Ba, J. L. 2014. Adam: Amethod for stochastic optimization. In *ICLR*.

Lewis, M.; Dauphin, Y.; and Fan, A. 2018. Hierarchical neural story generation. In *ACL*, 889–898.

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

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, X.; Mou, L.; Yan, R.; and Zhang, M. 2016c. Stalematebreaker: A proactive content-introducing approach to automatic human-computer conversation. In *IJCAI*, 2845–2851.

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

Lin, K.; Li, D.; He, X.; Sun, M.; and Zhang, Z. 2017. Adversarial ranking for language generation. In *NIPS*, 3158–3168.

Mei, H.; Bansal, M.; and Walter, M. R. 2017. Coherent dialogue with attention-based language models. In *AAAI*, 3252–3258.

Miller, A. H.; Feng, W.; Batra, D.; Bordes, A.; Fisch, A.; Lu, J.; Parikh, D.; and Weston, J. 2017. Parlai: A dialog research software platform. In *EMNLP*, 79–84.

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

Qiu, X., and Huang, X. 2015. Convolutional neural tensor network architecture for community-based question answering. In *IJCAI*, 1305–1311.

Ritter, A.; Cherry, C.; and Dolan, W. B. 2011. Data-driven response generation in social media. In *ACL*, 583–593.

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.

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

Shao, Y.; Gouws, S.; Britz, D.; Goldie, A.; Strope, B.; and Kurzweil, R. 2017. Generating high-quality and informative conversation responses with sequence-to-sequence models. In *EMNLP*, 2210–2219.

Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. Y. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*, 926–934.

Sordoni, A.; Galley, M.; Auli, M.; Brockett, C.; Ji, Y.; Mitchell, M.; Nie, J.; Gao, J.; and Dolan, B. 2015. A neural network approach to context-sensitive generation of conversational responses. In *NAACL*, 196–205.

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

Wang, D.; Jojic, N.; Brockett, C.; and Nyberg, E. 2017. Steering output style and topic in neural response generation. In *EMNLP*, 2140–2150.

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

Wu, H.; Liu, Y.; Chen, Y.; Zhao, W. X.; Dong, D.; Yu, D.; Zhou, X.; and Li, L. 2018. Multi-turn response selection for chatbots with deep attention matching network. In *ACL*, 1118–1127.

Xia, Y.; Tian, F.; Qin, T.; Yu, N.; and Liu, T. 2017. Sequence generation with target attention. In *ECML, PKDD*, 816–831.

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 *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, 617–626.

Yao, L.; Zhang, Y.; Feng, Y.; Zhao, D.; and Yan, R. 2017. Towards implicit content-introducing for generative short-text conversation systems. In *EMNLP*, 2190–2199.

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

Zhou, G.; Luo, P.; Cao, R.; Lin, F.; Chen, B.; and He, Q. 2017. Mechanism-aware neural machine for dialogue response generation. In *AAAI*, 3400–3407.