A Reinforcement Learning Framework for Natural Question Generation using Bi-discriminators

Zhihao Fan[1], **Zhongyu Wei**[1]*, **Siyuan Wang**[1], **Yang Liu**[2], **Xuanjing Huang**[3]
[1]School of Data Science, Fudan University, China
[2]Liulishuo Company
[3]School of Computer Science, Fudan University, China
{14300180043,zywei,14302010062}@fudan.edu.cn, yang.liu@liulishuo.com, xjhuang@fudan.edu.cn

Abstract

Visual Question Generation (VQG) aims to ask natural questions about an image automatically. Existing research focuses on training model to fit the annotated data set that makes it indifferent from other language generation tasks. We argue that natural questions need to have two specific textual attributes from the perspectives of content and linguistic respectively, namely, *natural* and *human-written*. Inspired by the setting of discriminator in adversarial learning, we propose two discriminators, one for each attribute, to enhance the training. We then use the reinforcement learning framework to incorporate scores from the two discriminators as the reward to guide the training of the question generator. Experimental results on a benchmark VQG dataset show the effectiveness and robustness of our model compared to some state-of-the-art models in terms of both automatic and human evaluation metrics.

1 Introduction

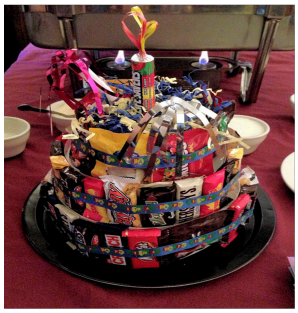
Recent years see the popularity of multi-modal research on vision and language. Visual caption generation (VCG) (Xu et al., 2015; Vinyals et al., 2015) and visual question answering (VQA) (Antol et al., 2015) attract increasing attention from research communities. VCG aims to generate descriptions for a given image with the goal of scene understanding, while VQA asks visual questions and requires an answer to it. Research for these two tasks are fueled by several manually generated corpora (Lin et al., 2014; Zhu et al., 2016).

Different from generating a statement (descriptions or answers) about an image, visual question generation (VQG) (Mostafazadeh et al., 2016) is tasked with generating a natural question which can potentially engage a human in starting a conversation when shown an image. Under this guidance, Mostafazadeh et al. (2016) collect natural questions for images via a crowd-sourcing platform and construct the first dataset for VQG. They also explore some neural network-based models for natural question generation. Those models are trained to better fit the VQG dataset that makes them indifferent from other language generation models and hard to identify the progress in naturalness for generated ones. Some human generated questions for VQG and questions for VQA are shown in Figure 1, we name questions for VQA descriptive questions and those for VQG natural ones. As we can see, VQA questions are much simpler and can be easily answered using information from the source image directly. In contrast, VQG questions are more complex and answers are not trivial. We therefore argue that the specialty in terms of content needs to be considered for natural question generation.

In this paper, We formulate the task of visual natural question generation as language generation task with specific attributes in terms of content and linguistics, i.e. *natural* and *human-written*. Recently, adversarial learning approaches (Goodfellow et al., 2014) have been applied to various tasks and show advantage of learning boundary for target data distribution. Inspired by the setting of discriminator, we propose to use two discriminators to better learn these two textual attributes. For the attribute of *human written*, we use a generative adversarial network (GAN) to learn a dynamic discriminator to

*Corresponding author

Proceedings of the 27th International Conference on Computational Linguistics, pages 1763–1774
Santa Fe, New Mexico, USA, August 20-26, 2018.



VQA:
1. How many plates are seen ?
2. Is this some sort of craft idea ?
3. How many different candies are visible ?

VQG:
1. Is this serving as a birthday cake ?
2. Where can I buy that birthday cake ?
3. How many different candies did you use to make this cake ?
4. Whose idea was it for a candy cake ?
5. Who made that candy cake ?

Figure 1: Annotated questions for tasks of VQG and VQA to an image from the dataset of MSCOCO (Antol et al., 2015).

distinguish human generated questions and machine generated questions. For the attribute of *natural*, we use questions from VQA as negative samples and questions from VQG as positive samples to train a static discriminator.

It is difficult to come up some differentiable objective function for our target. Recently, reinforcement learning has been introduced to optimize model in terms of non-differentiable metrics (Ranzato et al., 2015; Rennie et al., 2016; Zhang et al., 2017; Feng et al., 2018). Therefore, we propose a reinforcement learning framework (Williams, 1992) to incorporate scores from the two discriminators as the reward to guide the optimization of the question generator. Experiment results on a benchmark dataset show the effectiveness of our proposed framework in terms of both automatic and human evaluation.

We will introduce our framework in detail in section 2. In section 3, we present our experiment results. In section 4, we list some related works. In section 5, we conclude our work and point out some future directions.

2 Framework

Given an image I, we aim to train a generative model G_θ with parameter θ that is able to produce natural questions. The generator is designed following the fashion of Seq2Seq (Cho et al., 2014) that takes the representation of the image as input and generates a question word by word. We take two attributes of natural questions into consideration while training the generator. In particular, two discriminators are proposed to distinguish samples from two pairs of counter-question-distributions, namely *human written* vs *machine generated* and *natural* vs *descriptive*. A reinforcement learning framework is then used to combine results from the two discriminators as the reward to train the generator. The overall framework can be see in Figure 2.

2.1 Bi-discriminator configuration

We first introduce our setup of bi-discriminators in this sub-section starting with the design of a hierarchical structure for the distribution of questions.

Hierarchical structure for question distribution Suppose we have an overall domain $\mathcal{D}$ for all the questions, it can be split into two antithetic domains $\mathcal{D}_g$ (*machine generated*) and $\mathcal{D}_h$ (*human written*) according to linguistic attribute. The human written domain can be further split into $\mathcal{D}_{n+}$ (*natural*) and $\mathcal{D}_{n-}$ (*descriptive*) according to the content attribute *natural*. The hierarchical structure is described in Equation 1.

$$\begin{aligned} \mathcal{D} = \mathcal{D}_g \cup \mathcal{D}_h, \;\; \mathcal{D}_h = \mathcal{D}_{n+} \cup \mathcal{D}_{n-} \\ \mathcal{D}_{n+} \subset \mathcal{D}_h \subset \mathcal{D} \end{aligned} \tag{1}$$

To distinguish questions from the two pairs for counter-question-domains ($\mathcal{D}_g$ vs $\mathcal{D}_h$ and $\mathcal{D}_{n+}$ vs $\mathcal{D}_{n-}$), we propose two discriminators, D_1 and D_2. Discriminator D_1 is trained to discriminate whether a question comes from $\mathcal{D}_h$ or $\mathcal{D}_g$, and D_2 is trained to discriminate whether a question belongs to $\mathcal{D}_{n+}$

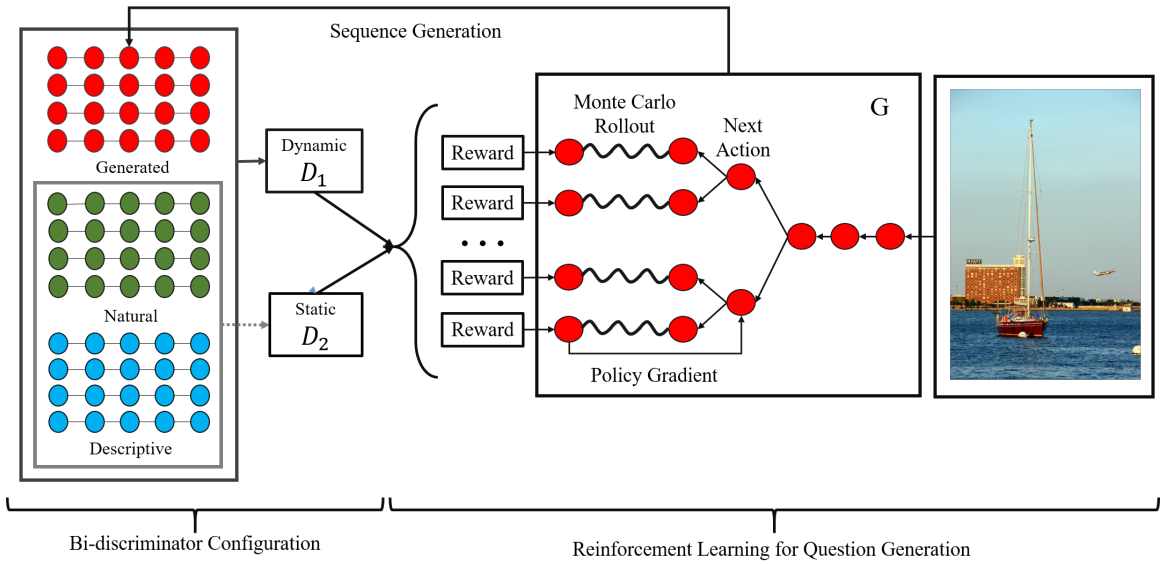


Figure 2: The overall framework of our proposed model.

or $\mathcal{D}_{n-}$. Final layer's activation of two discriminators is sigmoid, which represents the possibility that question belong to the positive domain $\mathcal{D}_*$ as formula 2 shown.

$$P_1\big(G_\theta(I) \in \mathcal{D}_h | I\big), \quad P_2\big(G_\theta(I) \in \mathcal{D}_{n+} | I\big) \tag{2}$$

where $G_\theta(I)$ stands for questions generated by the generator given image I, P_1 stands for the likelihood that a generated question is from domain $\mathcal{D}_h$ and P_2 stands for the likelihood that a generated question is from $\mathcal{D}_{n+}$.

Scores of the two discriminators are served as the reward in our reinforcement learning framework to guide G_θ to generate questions closer to questions in target domain $\mathcal{D}_*$. Under such setting, it is easy to observe that question similar to $\mathcal{D}_{n+}$ would be encouraged to generate by both D_1 and D_2, which means that our bi-discriminator environment configuration is able to encourage natural question generation in theory. In conclusion, maximizing score of $G_\theta(I)$ assigned by two discriminators is equal to encourage generate question with attributes of $\mathcal{D}_h$ (human-written) and $\mathcal{D}_{n+}$ (natural).

Discriminator D_1 for question domains $\mathcal{D}_g$ and $\mathcal{D}_h$ Discriminator D_1 is proposed to distinguish human written questions and machine generated questions. It is used to guide the generator to produce questions closer to samples from the domain of $\mathcal{D}_h$. We propose to use questions from a human generated dataset as positive samples while questions from our generator as negative samples. Considering the generator is updating during the training process, discriminator D_1 needs to be re-newed accordingly. We introduce generative adversarial network (GAN) for the training of D_1 to learn the border between $\mathcal{D}_h$ and $\mathcal{D}_g$ in pace with the updating of G_θ. The target of GAN is shown in Equation 3.

$$\min_G \max_D V(D,G) = \mathbb{E}_{p_{data}(\mathbf{x})}\big[log\, D(\mathbf{x})\big] + \mathbb{E}_{p_z(\mathbf{z})}\big[log\,\big(1 - D(G(\mathbf{z}))\big)\big] \tag{3}$$

The discriminator would converge to $D^*(\mathbf{x}) = p_{data}(x)/\big(p_{data}(\mathbf{x}) + p_g(\mathbf{x})\big)$ during the training. The optimal generator G in GAN aims to mock D to be unable to recognize generated samples. In our case, the optimal condition is that questions generated can be completely mixed with human-written questions, and the dynamic discriminator D_1 is incapable to distinguish these two kinds of questions and thus assign equal probabilities to both categories.

During training, once we have a batch of generated questions, we re-train the discriminator to minimize the loss in Equation 4,

$$L_{D_1} = -\mathbb{E}_{Q\sim D_g}\Big[log\,\big(1 - D_1(Q)\big)\Big] - \mathbb{E}_{Q\sim D_h}\Big[log\, D_1(Q)\Big] \tag{4}$$

where positive samples are from human-written domain $\mathcal{D}_h$, and negative samples are generated by current generator G_θ.

Discriminator D_2 for question domains $\mathcal{D}_{n+}$ and $\mathcal{D}_{n-}$ Discriminator D_2 is proposed to distinguish natural questions and descriptive questions. It is used to guide the generator to produce questions with information beyond the image to meet the attribute of *natural*. With human generated samples from both domains $\mathcal{D}_{n+}$ and $\mathcal{D}_{n-}$, discriminator can be trained before-hand and stay static during the training of the generator. Cross-entropy loss is used for the training of D_2. Considering that labeled samples for *natural* questions are much less than *descriptive* ones in reality, we need to consider the problem of class imbalance. As proved in the application of object detection (Lin et al., 2017b), focal loss can reduce the problem of imbalance. It slows the updating speed of a certain class when that class has been trained well. Similarly, we train the discriminator D_2 using focal loss following Equation 5.

$$
\begin{aligned}
p_t(Q,I) &= \begin{cases} P_2(Q \in \mathcal{D}_{n+}|I) & Q \in \mathcal{D}_{n+}|I \\ P_2(Q \in \mathcal{D}_{n-}|I) & Q \in \mathcal{D}_{n-}|I \end{cases} \\
L_{D_2} &= -\big(1 - p_t(Q,I)\big)^\gamma \, log \;\; p_t(Q,I)
\end{aligned} \tag{5}
$$

2.2 Reinforcement Learning for Question Generation

The reinforcement learning algorithm mainly consists of the generative model G_θ and the reward function R.

Generative model Our generator G_θ follows the design of Seq2Seq model. The only difference is that it takes image features as input instead of a sequence of words. We use *fc7* feature extracted from VGGNet to represent a given image. The decoder is a recurrent neural network that generates a question word by word. In the framework of reinforcement learning, the generation process can be described as a sequence of states and actions (known as trajectory as well), $\big(s_0, a_0, s_1, a_1, \cdots, s_T, a_T, s_{T+1}\big)$, where a_0 is the input image feature. $\mathcal{A}_t (t \geq 1)$ denotes the action space at time t. For text generation, action is the generation of a word, the action space is thus the whole vocabulary. In this task, when action is determined, the following state is also determined. Therefore, we can denote a trajectory as $\big(a_0, a_1, \cdots, a_T\big)$ for simplicity.

Reinforce learning and policy gradient training Based on the reward function R and the generative model G_θ, our goal is to maximize the expectation of reward $R(Q,I)$, where Q is a set of questions produced by G_θ and subject to $p_\theta(Q|I)$. By means of REINFORCE algorithm (Williams, 1992), the objective function is shown in Equation 6.

$$
\mathcal{J}(\theta) = \mathbb{E}_{Q \sim p_\theta(Q|I)}\big[R(Q,I)\big] \tag{6}
$$

Assuming that $p_\theta(Q|I)$ is continuously differentiable with respect to θ, the gradient of the equation 6 with respect to θ can be solved by policy gradient method shown in Equation 7 (Aleksandrov et al., 1968; Glynn, 1990; Williams, 1992).

$$
\frac{\partial \mathcal{J}(\theta)}{\partial \theta} = \mathbb{E}_{Q \sim p_\theta(Q|I)}\left[\left(\sum_{t=1}^{T} \frac{\partial}{\partial \theta} log\, p_\theta(a_t|a_{0:t-1})\right) R(Q,I)\right] \tag{7}
$$

During the training, after generating questions, we compute rewards based on reward function R and update G_θ using gradient ascent algorithm.

Monte Carlo Rollout The disadvantage of REINFORCE algorithm in language generation is that the reward can only be assigned to a complete sentence. Therefore, every single action taken for generating a complete sentence share the same reward. This hurts the effectiveness of training the generator.

To deal with such problem, we employ the strategy of Monte Carlo Rollout to assign a specific reward to each action. Suppose that we have a partial trajectory $a_{0:t-1}$ at t time, then we sample word a_t from

action space $\mathcal{A}_t$ according to policy G_θ. Next, we use Monte Carlo Rollout to generate the remaining $T-t$ tokens. The m_{th} Monte Carlo Rollout consequence is represented in equation 8:

$$\mathrm{MC}^m_{G_\theta}(a_{1:t-1}, a_t) = (a_{1:t-1}, a_t, a^m_{t+1:T}) \tag{8}$$

The reward for action a_t is computed as the mean reward of the sampling sentences, which is the following equation 9:

$$R\big((a_{0:t-1}, a_t), I\big) = \begin{cases} \frac{1}{M}\sum_{m=1}^{M} R\big(\mathrm{MC}^m_{G_\theta}(a_{1:t-1}, a_t), I\big) & t < T \\ R\big((a_{0:t-1}, a_t), I\big) & t = T \end{cases} \tag{9}$$

With Monte Carlo Rollout, the policy gradient is reconstructed as Equation 10.

$$\frac{\partial \mathcal{J}(\theta)}{\partial \theta} = \mathbb{E}_{p_\theta(Q|I)}\left[\left(\sum_{t=1}^{T} \frac{\partial}{\partial \theta} log\, p_\theta(a_t|a_{0:t-1})\right) R\big((a_{0:t-1}, a_t), I\big)\right] \tag{10}$$

Teacher Force As shown in Equation 10, ground-truth samples are not directly used to optimize the generator in the training process. In practice, this is usually in-efficient and is difficult to train a good question generator. Once the performance of generator G_θ is poor and the discriminator is able to do a good job identifying the origin of questions, it is non-trivial for the generator to find a way for improvement without guidance of ground-truth samples. To alleviate the issue, it is necessary for the generator to directly access the ground-truth questions, through which, generator G_θ learns the knowledge of target questions so that it is able to improve the performance. We follow Sutskever et al. (2014) to use the strategy of teacher force that trains the generator via MLE loss together with rewards from discriminators.

Training The overall training process of our proposed model is shown in Table 1.

Algorithm 1 Sketch of proposed model

Q stands for sample questions instances from human-generated dataset
$\hat{Q}$ stands for questions generated by the generator G(I)
Pre-train generator on VQA and VQG dataset
Pre-train D_1 and D_2
For number of training iterations **do**
  for i=1,D1-steps do
    Sample (Q, I) from real data
    Sample $\hat{Q} \sim G(I)$
    Update D_1 using (Q, I) as positive examples and $(\hat{Q}, I)$ as negative examples.

  for i=1,G1-steps do
    Sample (Q, I) from real data
    Sample $\hat{Q} \sim G(I)$
    Compute Reward r_1 for $(\hat{Q}, I)$ using D_1
    Update G on $(\hat{Q}, I)$ using reward r_1

  for i=1,G2-steps do
    Sample (Q, I) from real data
    Sample $\hat{Q} \sim G(I)$
    Compute Reward r_2 for $(\hat{Q}, I)$ using D_2
    Update G on $(\hat{Q}, I)$ using reward r_2

  for i=1,G3-steps do
    Sample (Q, I) from real data
    Teacher Force: Update G using (Q, I)
  End

3 Experiments

3.1 Dataset

We evaluate our models using MSCOCO part of Visual Question Generation (VQG) dataset[1] (Mostafazadeh et al., 2016). It contains 2500, 1250 and 1250 images for training, validation and testing respectively. Each image is accompanied with 5 natural questions produced by human annotators.

Another dataset involved is VQA. For each image in VQA, three questions are collected. The intention of building such dataset is to teach model to response to the questions with respect to concepts in some certain images. Therefore, questions in VQA are much simpler than those ones in VQG. VQG dataset contains about 80000, 40000, 80000 images for training, validation and testing respectively. VQA is used to pre-train our question generator and questions are also served as negative samples to train the discriminator D_2.

3.2 Models for Comparison

We compare our models with some baselines and some state-of-the-art methods.

- **KNN** (Mostafazadeh et al., 2016): using the question from the most similar image as the question for a target image. Cosine similarity based on *fc7* features is utilized to search for similar images.
- **Img2Seq**: it generates a question from image features following Seq2Seq fashion (Cho et al., 2014). The model is trained using the word-level loss maximum likelihood estimation (MLE).
- **Img2Seq**$_{pre-train}$: different from *Img2Seq*, this model is pre-trained on VQA.
- **MIXER-BLEU-4** (Ranzato et al., 2015): it follows the framework of reinforcement learning that uses BLEU-4 between the generated question and the human-written question as the reward to guide the parameter update of the generator with policy gradient. Since the model is trained by optimizing BLEU-4 directly, it is able to generate higher BLEU-4 score in general.
- **Reinforce**$_{\mathbf{D_1}}$: it uses the score of D_1 as the reward to guide the training of the generator. The setting is quite similar to *SeqGAN* (Yu et al., 2017). This model is a upgrade version of **Img2Seq**. It introduces adversarial learning network to better train the generator under the reinforcement learning framework.
- **Reinforce**$_{\mathbf{D_2}}$: it uses the score of D_2 as the reward to guide the training of the generator. This model is comparable to *MIXER-BLEU-4* because both models utilize a static way to produce reward (BLEU score with ground truth questions in *MIXER-BLEU-4* and classification confidence in Reinforce$_{\mathbf{D_2}}$)
- **Reinforce**$_{\mathbf{D_1+D_2}}$: this is our proposed model.

3.3 Training Details

For the generator, we use GRU cell and the number of cells is 512; the dimension of word embedding is 300 and is pre-trained using GloVe (Pennington et al., 2014). The image feature, *fc7* is the output of the 7th fully-connected layer in *VGGNet*. The original dimension of *fc7* is 4096, and we compress it to a 300-dimension vector using 2-layer fully-connected layer. The upper settings are the same for all neural network models. We set batch size, rollout size, D1-step, G1-step, G2-step and G3-step as 64, 16, 5, 1, 2 and 1, respectively. γ for the training of D_2 is set to 2.0 empirically.

3.4 Automatic Evaluation

There is no direct evaluation metric to determine whether a question is natural or not. We thus use several relevance scores for the automatic evaluation following the setting of existing researches, including Corpus BLEU-4, BLEU-4 (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005), ROUGE (Lin and Hovy, 2003) and CIDEr (Vedantam et al., 2015). The overall experiment results in terms of five relevance scores are shown in Table 1. We have some findings as follows:

[1]The original dataset contains three parts, namely, MS-COCO, Flickr and Bing, *VQG-MSCOCO*. However, images from Flickr and Bing are quite different from those in Visual Question Answering (VQA) dataset (Antol et al., 2015), and this makes it difficult to use questions from VQA as negative samples to train D_2.

Model	BLEU-4	corpus BLEU-4	METEOR	ROUGE	CIDEr
KNN	37.062	19.799	22.413	52.324	50.199
Img2Seq	36.744	21.028	23.125	54.089	51.171
Img2Seq$_{pre-train}$	37.522	22.106	23.877	55.310	54.076
MIXER-BLEU-4	**41.674**	24.808	24.382	**57.777**	60.527
Reinforce$_{\mathbf{D_1}}$	38.945	24.420	24.665	56.196	59.513
Reinforce$_{\mathbf{D_2}}$	40.063	25.237	25.492	57.503	61.745
Reinforce$_{\mathbf{D_1+D_2}}$	41.098	**26.265**	**25.634**	57.679	**63.388**

Table 1: Overall experiment results of different models in terms of relevance scores. **bold**: the best performance in that column.

- Although *KNN* model retrieves questions from the original dataset, the result is not good enough. Further analysis reveals that questions generated are not relevant to the target image. Therefore, the strategy of reusing questions from similar images is not sufficient for generating question with high quality.
- Performance of *Img2Seq*$_{pre-train}$ is better than that of *Img2Seq* in terms of all the five metrics. This indicates that samples from the domain of $\mathcal{D}_{n-}$ are also helpful for generating high quality natural questions. This is reasonable because samples from VQA also locate in the domain of $\mathcal{D}_h$ therefore pre-training is able to guide the generator to better learn the attribute of *human written*.
- By incorporating the discriminator D_2, both models of *Reinforce*$_{\mathbf{D_2}}$ and *Reinforce*$_{\mathbf{D_1+D_2}}$ is able to improve the performance in terms of all the five relevance scores compared with their counter-part models *Img2Seq*$_{pre-train}$ and *Reinforce*$_{\mathbf{D_1}}$ respectively. This confirms the effectiveness of discriminator D_2. The existence of $\mathcal{D}_{n-}$ helps the generator G_θ learns more about the unique features for questions in $\mathcal{D}_{n+}$ and better demarcates the boundary of itself.
- By incorporating the discriminator D_1, both *Reinforce*$_{\mathbf{D_1}}$ and *Reinforce*$_{\mathbf{D_1+D_2}}$ produce better performance than their counter-part approaches *Img2Seq* and *Reinforce*$_{\mathbf{D_2}}$ respectively. The disadvantage of training using MLE is that only human-generated training samples are exposed to generator (known as exposure bias) and loss is computed in word-level without considering sentence-level performance. By adding D_1, the generator is able to observe generated question during training. Therefore the issue of exposure bias can be partially addressed. Besides, the sentence level performance is also computed and used as reward to guide the training process.
- *MIXER-BLEU-4* produces the highest score of BLEU-4 and ROUGE but it performs much worse in terms of other three metrics. This indicates that simply using one criteria value as the reward is not universal.
- Our proposed model *Reinforce*$_{\mathbf{D_1+D_2}}$ produces three best values out of the five evaluation metrics. This confirms the effectiveness of our proposed framework for natural question generation. Although our model is not optimizing BLUE-4 directly, it performs comparable to *MIXER-BLEU-4* in terms of BLUE-4, indicating the robustness of our model.

3.5 Human Evaluation

We also perform human evaluation on generated questions from different models to evaluate their naturalness (Mostafazadeh et al., 2016). For a given image, we first list questions generated by different models and randomly sample one question from ground-truth to form the question pool. Then we present the image and the corresponding question pool to the annotator. S/he is asked to assign a score to each single question in term of naturalness. A guidance from (Mostafazadeh et al., 2016) is presented during the annotation. We set the score range in $(1\ 2\ 3)$ following (Mostafazadeh et al., 2016). The better the question is, the higher score it would get.

KNN What type of bird are these ?
MIXER-BLEU-4 Are those bananas ?
Img2Seq What kind of bananas are those ?
$\boldsymbol{Reinforce_{D_1}}$**:** Are these bananas fresh ?
$\boldsymbol{Reinforce_{D_1+D_2}}$**:** How much do you think those bananas cost ?

KNN How old is the man riding the horse ?
MIXER-BLEU-4 Is that a horse ?
Img2Seq Is this horse racing today ?
$\boldsymbol{Reinforce_{D_1}}$ What is the name of the horse ?
$\boldsymbol{Reinforce_{D_1+D_2}}$ Why is this horse fitted with a harness ?

Figure 3: Generated question by different models.

Model	1	2	3	Avg
KNN	214	120	66	1.63
Img2Seq	182	147	71	1.72
MIXER-BLEU-4	153	172	75	1.81
$Reinforce_{\mathbf{D_1}}$	167	153	80	1.78
$Reinforce_{\mathbf{D_1+D_2}}$	149	160	91	1.86
Ground-truth	50	79	271	2.55

Table 2: Results of human evaluation for different models.

Two annotators are invited to label the questions independently. We collect the number of various ratings for different models and add up the value of the two annotators. We randomly choose 200 images for human evaluation. Result is shown in Table 2. Figure 3 shows two sample images and corresponding questions generated by different models. Based on the result of human evaluation and the sample questions generated by the system, we have some findings as follows:

- Relatively poor performance of retrieval model indicates that questions retrieved by KNN are not relevant to the image as we can see in Figure 3.

- The performance of $Reinforce_{\mathbf{D_1+D_2}}$ is the best among all automatic question generators in terms of the number of rating 3 it obtains. This reconfirms the effectiveness of our framework. Besides, the sample questions in Figure 2. also reveals that our system is able to ask more complex questions.

- The performance of *MIXER-BLEU-4* is middle-level. This indicates that optimizing a single evaluation metric is not sufficient enough for generating high quality natural questions.

- The gap between ground-truth questions and machine generated questions is still large. This indicates that there is still a large room for question generation system to improve.

4 Related Works

This paper locates in the research filed of question generation and reinforcement learning for sequence generation. We will focus on related works from these two domains.

Question Generation Question generation has been researched for years from textual input (Rus et al., 2010; Heilman, 2011). Researchers start from rule-based method that extracts key aspects from the input text and then insert these aspects into human generated templates for interrogative sentence generation (Heilman, 2011). Recently, sequence-to-sequence model is utilized for question generation in description-question pairs (Du et al., 2017; Tang et al., 2017; Serban et al., 2016). Although these models generate better performance, the characteristics of question is still ignored. On the other hand, research about visual question generation is much less (Ren et al., 2015; Vijayakumar et al., 2018;

Mostafazadeh et al., 2016; Shijie et al., 2017). Diversity as another important characteristic of question also draws much attention. Li et al. (2016) proposed to use Maximum Mutual Information (MMI) as the objective function for result diversification. Vijayakumar et al. (2018) proposed a diverse beam search for generated multiple questions. Fan et al. (2018) utilized question type driven framework to diversify question generation.

Natural question generation cares more about a specific attribute of the generated text in terms of content. Although some attempts have been explored for this task, researchers ignore the attribute of *natural* in general. In our work, we treat this task as language generation with an additional attribute and propose a reinforcement learning framework for it. Our framework can also be used to other language generation tasks like dialogue generation with emotion information.

Reinforcement Learning For Sequence Generation *MIXER* (Ranzato et al., 2015) model uses REINFORCE method with a baseline reward estimator, directly optimizes BLEU-4 score of generated sequence. In order to compensate same reward for all action in generation and lack of ground-truth sequence knowledge, curriculum learning is utilized. Although performance in corresponding metric gets better, but model still lacks robustness and distinguishing reward for every generate step. Actor-Critic algorithm (Konda and Tsitsiklis, 2000) in sequence generation (Bahdanau et al., 2017) utilizes value network to estimate reward for every generate step. Rennie et al. (2016) utilizes self-critic $\arg\max$ to estimate generator's reward more accurately and does not need additional value network. Different reward function also gets attention for better evaluation and robustness. Liu et al. (2017) uses a combination of different metrics as the reward. In the task of visual caption generation, visual-semantic embedding (Frome et al., 2013; Kiros et al., 2015) is used to be reward (Ren et al., 2017) for better matching image and caption. SeqGAN (Yu et al., 2017), RankGAN (Lin et al., 2017a), LeakGAN (Guo et al., 2018) are also based on Reinforcement Learning, they aim to generate sequence more similar to human-written ones, thus the similarity is assigned as reward to generator.

In our setting, we borrow the idea of GAN to train one of our discriminators. Adversarial training is effective in improving generator's capability in general tasks, and it helps our generator get better guidance of natural attribute. Most of existing research under reinforcement learning focuses on using a single reward to guide the training of the generator. In this paper, we propose a setting of bi-discriminator to consider two attributes of target text and it is easy to be generalized to multiple discriminators incorporating other information.

5 Conclusion and Future Work

In this paper, we propose a reinforcement learning framework for natural question generation. It incorporates two discriminators to take two specific attributes of natural question into consideration. Experimental results on a benchmark VQG dataset show the effectiveness and robustness of our proposed model. The proposed framework can be applied to other language generation tasks with additional attribute, such as dialogue generation with emotion information. Besides, the setting of bi-discriminators can be extended to multi-discriminators to incorporate more information.

The future research can be carried out in several directions. First, during our experiment, we find that the time and space consumption of Monte Carlo Rollout is expensive. More effective and powerful methods of assigning reward for every generate step deserve research in the future. Second, We will explore more structural scoring system and better collaborative method of multiple discriminators. Third, another future direction is to incorporate some automatic evaluation metrics into our reinforcement learning framework to improve the performance further.

Acknowledgements

The work is partially supported by National Natural Science Foundation of China (Grant No.61702106), Shanghai Science and Technology Commission (Grant No.17JC1420200, Grant No.17YF1427600 and Grant No.16JC1420401).

References

VM Aleksandrov, VI Sysoyev, and SHEMENEV. VV. 1968. Stochastic optimization. *Engineering Cybernetics*, (5):11–+.

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2015. Vqa: Visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2425–2433.

Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. An actor-critic algorithm for sequence prediction. In *Proceedings of ICLR*.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Kyunghyun Cho, Bart van Merriënboer, Çalar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October. Association for Computational Linguistics.

Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to ask: Neural question generation for reading comprehension. In *Association for Computational Linguistics (ACL)*.

Jun Feng, Minlie Huang, Li Zhao, Yang Yang, and Xiaoyan Zhu. 2018. Reinforcement learning for relation classification from noisy data. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*.

Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Tomas Mikolov, et al. 2013. Devise: A deep visual-semantic embedding model. In *Advances in neural information processing systems*, pages 2121–2129.

Peter W Glynn. 1990. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM*, 33(10):75–84.

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

Jiaxian Guo, Sidi Lu, Han Cai, Weinan Zhang, Yong Yu, and Jun Wang. 2018. Long text generation via adversarial training with leaked information. In *AAAI*, pages 2852–2858.

Michael Heilman. 2011. *Automatic factual question generation from text*. Ph.D. thesis, Carnegie Mellon University.

Chris Brockett Jianfeng Gao Jiwei Li, Michel Galley. 2016. A diversity-promoting objective function for neural conversation models. In *Proceedings of NAACL-HLT*.

Ryan Kiros, Ruslan Salakhutdinov, and Richard S Zemel. 2015. Unifying visual-semantic embeddings with multimodal neural language models. In *Transactions of the Association for Computational Linguistics (TACL)*.

Vijay R Konda and John N Tsitsiklis. 2000. Actor-critic algorithms. In *Advances in neural information processing systems*, pages 1008–1014.

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 71–78. Association for Computational Linguistics.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

Kevin Lin, Dianqi Li, Xiaodong He, Ming-ting Sun, and Zhengyou Zhang. 2017a. Adversarial ranking for language generation. In *Advances in Neural Information Processing Systems*, pages 3158–3168.

Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. 2017b. Feature pyramid networks for object detection. In *CVPR*, pages 936–944. IEEE Computer Society.

Siqi Liu, Zhenhai Zhu, Ning Ye, Sergio Guadarrama, and Kevin Murphy. 2017. Improved image captioning via policy gradient optimization of spider. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct.

Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. 2016. Generating natural questions about an image. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1802–1813, Berlin, Germany, August. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *CoRR*, abs/1511.06732.

Mengye Ren, Ryan Kiros, and Richard Zemel. 2015. Exploring models and data for image question answering. In *Advances in neural information processing systems*, pages 2953–2961.

Zhou Ren, Xiaoyu Wang, Ning Zhang, Xutao Lv, and Li-Jia Li. 2017. Deep reinforcement learning-based image captioning with embedding reward. In *Proceeding of IEEE conference on Computer Vision and Pattern Recognition (CVPR)*.

Steven J. Rennie, Etienne Marcheret, Youssef Mroueh, Jarret Ross, and Vaibhava Goel. 2016. Self-critical sequence training for image captioning. *CoRR*, abs/1612.00563.

Vasile Rus, Brendan Wyse, Paul Piwek, Mihai Lintean, Svetlana Stoyanchev, and Cristian Moldovan. 2010. The first question generation shared task evaluation challenge. In *Proceedings of the 6th International Natural Language Generation Conference*, pages 251–257. Association for Computational Linguistics.

Iulian Vlad Serban, Alberto García-Durán, Caglar Gulcehre, Sungjin Ahn, Sarath Chandar, Aaron Courville, and Yoshua Bengio. 2016. Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 588–598, Berlin, Germany, August. Association for Computational Linguistics.

Zhang Shijie, Qu Lizhen, You Shaodi, Yang Zhenglu, and Zhang Jiawan. 2017. Automatic generation of grounded visual questions. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4235–4243.

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

Duyu Tang, Nan Duan, Tao Qin, and Ming Zhou. 2017. Question answering and question generation as dual tasks. *CoRR*, abs/1706.02027.

Ramakrishna Vedantam, C Lawrence Zitnick, and Devi Parikh. 2015. Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4566–4575.

Ashwin K Vijayakumar, Michael Cogswell, Ramprasath R Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. 2018. Diverse beam search: Decoding diverse solutions from neural sequence models. In *Proceedings of AAAI*.

Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. In *Reinforcement Learning*, pages 5–32. Springer.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*, pages 2048–2057.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, pages 2852–2858.

Li Zhang, Flood Sung, Feng Liu, Tao Xiang, Shaogang Gong, Yongxin Yang, and Timothy M. Hospedales. 2017. Actor-critic sequence training for image captioning. *CoRR*, abs/1706.09601.

Fan Zhihao, Wei Zhongyu, Li Piji, Lan Yanyan, and Huang Xuanjing. 2018. A question type driven framework to diversify visual question generation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-18*.

Yuke Zhu, Oliver Groth, Michael Bernstein, and Li Fei-Fei. 2016. Visual7w: Grounded question answering in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4995–5004.