

# Meta-CoTGAN: A Meta Cooperative Training Paradigm for Improving Adversarial Text Generation

Haiyan Yin, Dingcheng Li, Xu Li, and Ping Li

Cognitive Computing Lab

Baidu Research

No.10 Xibeiwang East Road, Beijing, 10085, China

10900 NE 8th St. Bellevue, WA 98004, USA

haiyanyin18@gmail.com, {lidingcheng, lixu13}@baidu.com, pingli98@gmail.com

## Abstract

Training generative models that can generate high-quality text with sufficient diversity is an important open problem for Natural Language Generation (NLG) community. Recently, generative adversarial models have been applied extensively on text generation tasks, where the adversarially trained generators alleviate the *exposure bias* experienced by conventional maximum likelihood approaches and result in promising generation quality. However, due to the notorious defect of *mode collapse* for adversarial training, the adversarially trained generators face a quality-diversity trade-off, i.e., the generator models tend to sacrifice generation diversity severely for increasing generation quality. In this paper, we propose a novel approach which aims to improve the performance of adversarial text generation via efficiently decelerating *mode collapse* of the adversarial training. To this end, we introduce a cooperative training paradigm, where a language model is cooperatively trained with the generator and we utilize the language model to efficiently shape the data distribution of the generator against *mode collapse*. Moreover, instead of engaging the cooperative update for the generator in a principled way, we formulate a meta learning mechanism, where the cooperative update to the generator serves as a high level meta task, with an intuition of ensuring the parameters of the generator *after* the adversarial update would stay resistant against *mode collapse*. In the experiment, we demonstrate our proposed approach can efficiently slow down the pace of *mode collapse* for the adversarial text generators. Overall, our proposed method is able to outperform the baseline approaches with significant margins in terms of both generation quality and diversity in the testified domains.

## Introduction

Generative models are trained to learn the true data distribution from the training set and are capable of generating new data points when the training is completed. In recent years, they have been successfully applied to a wide range of applications, including image generation (Arjovsky, Chintala, and Bottou 2017), stylization (Ulyanov, Vedaldi, and Lempitsky 2016), semi-supervised classification (Radford, Metz, and Chintala 2015), and natural language generation (Bahdanau, Cho, and Bengio 2015; Li, Sun, and Li 2019; Sun,

Li, and Li 2018), etc. In this paper, we tackle the emerging task of text generation, which is typically modeled as a sequential discrete data generation process (Lamb et al. 2016). Such tasks play a pivot role in many real world applications, such as machine translation (Sutskever, Vinyals, and Le 2014), text summarization (Li et al. 2017; Liu, Zhong, and Li 2012), and dialogue systems (Wen et al. 2015; Li et al. 2016).

The training of sequential text generation models has been greatly relying on applying *teacher forcing* over autoregressive models, i.e., optimizing with maximum likelihood estimation (MLE) (Che et al. 2017). However, training the generative models with *teacher forcing* would suffer from *exposure bias* (Rajeswar et al. 2017), i.e., the models are fed to their predicted data rather than the ground-truth data at inference time and thus result in generating poor samples due to the accumulated error. To address the *exposure bias* issue, a major on-going research for text generation centers on utilizing adversarial training techniques to derive better text generation models. Generally, such attempts could be classified into the following two strands (Chen et al. 2018): the first line of approaches combine generative adversarial network (GAN) (Goodfellow et al. 2014) with reinforcement learning (RL), denoted as RL-based; the second line of approaches solely play the two-player adversarial game without using RL, denoted as RL-free.

Both RL-based and RL-free text generation approaches suffer from *mode collapse*, a notoriously known challenge for training GAN-based models (Arjovsky, Chintala, and Bottou 2017). That is, as the adversarial training progresses, the generated distribution tends to contrast towards generating subset of modes for the data. As a result, the generator outputs repeated sentences and thus no longer expressively represents the data generating distribution. Such effect has been quantitatively evaluated in a recent study, which shows that the entropy of the generator’s output distribution would experience a clear drop when moving from MLE training to adversarial training phase (Caccia et al. 2018). To derive better text generation models with GAN-based techniques, one critical thing is to achieve a better quality-diversity trade-off by efficiently slowing down the *mode collapse* of the adversarial generator, i.e., to let the generator get abundant gradi-

ent information from adversarial update for making its output more real (i.e., improve quality) while bearing with small *mode collapse* effect (i.e., decrease diversity). However, limited number of existing RL-based or RL-free approaches explicitly consider dealing with *mode collapse* of GAN training. In this work, we propose a cooperative training mechanism which explicitly tackles the challenge of *mode collapse* for adversarial training, resulting in an improved text generation model.

Overall, the contributions of this paper are three-fold. Firstly, we propose a novel cooperative training approach where we utilize a language model to efficiently shape the output distribution of the adversarial text generator. Our proposed approach could efficiently slow down the *mode collapse* of the adversarial text generator and thus lead the text generation towards a better quality-diversity trade-off. Secondly, to optimize the cooperative training loss for the generator, we propose a novel meta-learning mechanism. In our setting, the cooperative training task serves as a meta task and the adversarial training serves as a base task. Thus, our proposed approach ensures that the generator parameters *after* the adversarial update would be resistant for *mode collapse*. Thirdly, we conduct extensive experiments on synthetic and real-world datasets to demonstrate that our proposed approach is able to produce better text generation models in terms of both the quality and the diversity.

## Related Work

Besides the conventional approaches of training language models with *teacher forcing*, today’s approaches for text generation could be generally classified as RL-based or RL-free approaches. Most RL-based approaches formulate text generation as a Markov Decision Process (MDP). Often, the generator is updated by policy gradient algorithm (Sutton et al. 2000) or its variants using reward signals derived from GAN’s discriminator. Prominent examples for this type of approaches include SeqGAN (Yu et al. 2017), RankGAN (Lin et al. 2017), LeakGAN (Guo et al. 2018) and MaskGAN (Fedus, Goodfellow, and Dai 2018). The noisy reward signals derived from the discriminator model makes such RL-based models suffer from high-variance gradients to update the generator’s parameters. Besides high-variance of gradient, the RL-based approaches also face the difficulties brought by partial sequence evaluation, slow learning, and sensitive hyperparameters (Caccia et al. 2018). Considering such challenges for the RL-based approaches, in this work, our proposed method resides in, but not restricted to, the category of RL-free approach for text generation. Prominent examples of RL-free approaches include TextGAN (Zhang et al. 2017), FM-GAN (Chen et al. 2018), GSGAN (Kusner and Hernández-Lobato 2016), and RelGAN (Nie, Narodytska, and Patel 2019). Such approaches feed the generator with low variance gradient and often lead to more stable training.

Most of the adversarial text generation models are firstly pretrained by MLE, and then are continuously optimized by adversarial training under either RL-based or RL-free mechanism. When switched from MLE training to adversarial training phase, the generator models for both RL-

based and RL-free approaches would suffer from *mode collapse* issue. In this work, our core intuition is to utilize a cooperatively trained language model to decelerate the *mode collapse* of adversarial training. Such intuition of utilizing language model to facilitate adversarial text generation aligns with the works proposed in (Xu et al. 2018; Lu et al. 2018). In (Xu et al. 2018), the discriminator for adversarial training is modeled as a language model, which maximizes the probability for real data and minimizes that for generated data. Furthermore, the output derived from the language model is adopted as reward signal to promote generation diversity under an RL-based set-up. Our work is mostly related to the cooperative training method proposed in (Lu et al. 2019), where a language model is trained online to offer a target distribution for minimizing the Jensen-Shannon divergence between the real data distribution and the generated distribution. In our work, we adopt a similar strategy to train the language model, but the cooperative training for the generator model is different from (Lu et al. 2019). Furthermore, we propose a distinct meta learning set-up to optimize the cooperative training loss for the generator. To the best of our knowledge, our work is the first attempt that adopts meta learning on text generation GANs.

## Preliminaries

The task of text generation is typically modelled as sequential discrete data generation process. Let  $\{x_j\}_{j=1}^N$  be the  $N$  data points drawn from an underlying data generating distribution  $p_{data}$ . Each data point is represented as a sequence of discrete tokens:  $\mathbf{x} = (y_1, \dots, y_T)$ , where  $y_i$  denotes the  $i$ -th token and  $T$  denotes the length of the sequence. Let  $G_\theta$  denote the generator model parameterized by  $\theta$ . Conventional text generation approaches typically train a language model with maximum likelihood estimation (MLE) as follows:

$$\min_{\theta} \mathbb{E}_{\mathbf{x} \sim p_{data}} [-\log G_\theta(\mathbf{x})],$$

where the probability of each sequence  $\mathbf{x}$  is represented in an autoregressive manner:

$$G_\theta(\mathbf{x}) = \prod_{i=1}^T G_\theta(y_i | y_{<i}; \theta),$$

with  $y_{<i}$  denoting the sequence of previous tokens  $y_1, \dots, y_{i-1}$ .

The approaches utilizing GANs for text generation attempt to play a two-player game between the generator  $G_\theta$  and a discriminator  $D$ . Let the discriminator  $D$  be parameterized by  $\phi$ . Under the adversarial set-up, the generator  $G_\theta$  is trained to generate realistic sentences given samples from  $p_{data}$ , and the discriminator  $D_\phi$  attempts to distinguish between  $G_\theta$ ’s generating distribution  $p_\theta$  and the real data distribution  $p_{data}$ . Thus, the above mentioned process could be formulated as an adversarial training mechanism as follows:

$$\min_{\theta} \max_{\phi} \mathbb{E}_{\mathbf{x} \sim p_{data}} [\log(D_\phi(\mathbf{x}))] + \mathbb{E}_{\mathbf{x} \sim p_\theta} [\log(1 - D_\phi(G_\theta(\mathbf{x}))), \quad (1)$$

where the generator  $G_\theta$  and discriminator  $D_\phi$  attempt to minimize and maximize the function, respectively. We denote the adversarial loss in (1) in terms of the generator

model and the discriminator model as  $\mathcal{L}_{adv}(\theta)$  and  $\mathcal{L}_{adv}(\phi)$ , respectively.

With the autoregressive generation process, the  $i$ -th token  $y_i$  is generated by sampling from the generator’s output distribution, conditioned on its previous tokens  $y_{<i}$ . Performing such sampling introduces considerable difficulty for the generator to utilize the discriminator’s prediction outcome. That is, the backpropagation route for adversarial loss, i.e.,

$$\frac{\partial \mathcal{L}_{adv}}{\partial \theta} = \sum_{i=0}^{T-1} \frac{\partial \mathcal{L}_{adv}}{\partial y_{t+1}} \frac{\partial y_{t+1}}{\partial \theta},$$

becomes non differentiable w.r.t. the generator’s parameters  $\theta$ , since  $\frac{\partial y_{t+1}}{\partial \theta}$  would be zero due to the sampling. To overcome the above issue, the RL-based approaches mostly rely on the REINFORCE algorithm (Williams 1992) or its variants to derive the gradient to optimize the generator, where the discriminator’s predictions could be utilized to derive reward signals. The RL-free approaches often relax the non-differentiable sampling function by some continuous approximations, such as *soft-argmax* (Zhang et al. 2017) or *gumbel-softmax* (Jang, Gu, and Poole 2017). In this paper, our proposed approach adopts the *gumbel-softmax* relaxation which models the effect of sampling as introducing noise to the input so that the outputs become continuous and differentiable. Specifically, the noise is modeled by a Gumbel distribution, which is formed as follows:

$$g_t^{(i)} = -\log(-\log(U_t^{(i)})), \text{ with } U_t^{(i)} \sim \text{Uniform}(0, 1),$$

where  $g_t^{(i)}$  denotes the Gumbel noise to be applied to the  $i$ -th logits. With the Gumbel noise, the token for next step  $y_{t+1}$  is derived in a deterministic manner:

$$y_{t+1} = \text{one\_hot}\left(\arg \max_{1 \leq i \leq V} (\mathbf{o}_t^{(i)} + g_t^{(i)})\right),$$

where  $\mathbf{o}_t \in \mathbb{R}^V$  denotes the logits output by the generator for sampling token  $y_{t+1}$ , and  $V$  denotes vocabulary size. To make the discriminator’s loss differentiable, the *argmax* operator is replaced by a *softmax* function  $\sigma(\cdot)$ , i.e.,  $\hat{\mathbf{y}}_{t+1} = \sigma(\beta(\mathbf{o}_t + \mathbf{g}_t))$ , where  $\beta$  is a real-valued temperature hyperparameter, with  $\beta > 0$ .

## Methodology

Language generators trained with adversarial training mechanism (both RL-based and RL-free approaches) suffer from *mode collapse* when switched from *teacher forcing* to the adversarial training phase. In this section, we introduce a novel meta cooperative training algorithm to overcome such challenges. Overall, our objective is to achieve a better quality-diversity trade-off for the language generators via decelerating *mode collapse* of their adversarial training. That is, the algorithm allows the generator to get abundant gradient information from the adversarial training for increasing generation quality, while sacrificing little in terms of generation diversity. Overall, we engage a language model to decelerate the *mode collapse* of the generator’s output distribution. The language model is cooperatively trained with the generator  $G_\theta$  during adversarial training. We utilize the output of language model over samples

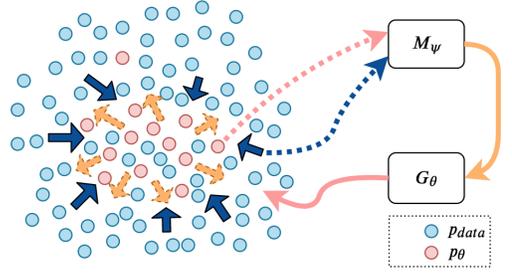


Figure 1: Depiction for our proposed cooperative training mechanism. The generator trained with adversarial training tend to suffer from *mode collapse* (dark blue arrows). We engage a language model to supervise the data distribution of  $G_\theta$  to decelerate *mode collapse* (yellow arrows). The language model is trained from a mixture distribution of samples from  $p_\theta$  and  $p_{data}$ . The supervision from language model to the language generator works on samples from  $p_{data}$ . The generator is updated by the adversarial loss and the cooperative training loss.

from real data distribution  $p_{data}$  to shape the generator’s output distribution. Furthermore, the supervision is formulated with a meta optimization setup.

## Cooperative Training Formulation

We introduce a cooperative training paradigm that engages an interleaved training procedure for an adversarial generator  $G_\theta$ , an adversarial discriminator  $D_\phi$ , and a language model  $M_\psi$ , where  $\psi$  denotes the parameters for the language model. Figure 1 depicts a high-level overview for the proposed cooperative training procedure. When the generator  $G_\theta$  is trained by the adversarial loss, its generation diversity would progressively decrease for increasing the generation quality due to *mode collapse* issue. To overcome such challenge, we cooperatively train a language model  $M_\psi$ . The language model would pose a supervision over  $G_\theta$ ’s output distribution towards preserving desirable generation probability for the real data.

During the cooperative training process, the language model is optimized consistently by MLE loss. To offer a smoothly changing target distribution for the generator, it is trained with data from a mixture distribution with balanced samples from real data and generated data, i.e.,  $\frac{1}{2}(p_{data} + p_\theta)$ . Formally, the cooperative training loss  $\mathcal{L}_{cot}(\psi)$  for updating the language model with MLE is defined in (2). It could be interpreted as minimizing the direct KL divergence between  $M_\psi$  and an optimal mixture density model  $M^*$  which has a distribution of  $\frac{1}{2}(p_{data} + p_\theta)$ .

$$\begin{aligned} & \nabla_\psi \mathcal{L}_{cot}(\psi) \\ &= -\frac{1}{2} \nabla_\psi \left( \mathbb{E}_{\mathbf{x} \sim p_\theta} \log(M_\psi(\mathbf{x})) + \mathbb{E}_{\mathbf{x} \sim p_{data}} \log(M_\psi(\mathbf{x})) \right) \\ &= \nabla_\psi \mathbb{E}_{\mathbf{x} \sim M^*} \left( \log \frac{M^*(\mathbf{x})}{M_\psi(\mathbf{x})} \right) = \nabla_\psi \text{KL}(M^* || M_\psi), \end{aligned} \quad (2)$$

Consistently updating the language model  $M_\psi$  with samples from real data and using the *teacher forcing* loss makes

it experience mild *mode collapse* effect. Thus, its output predictions could offer an effective supervision over the generator  $G_\theta$ 's output distribution for decelerating *mode collapse*. Moreover, updating  $M_\psi$  with the mixture distribution, compared to only using the real data distribution, would offer a target distribution that is smoothly changing towards the generator's update, which turns out to be more beneficial. Formally, the cooperative training loss for the generator model is proposed as follows,

$$\begin{aligned} \mathcal{L}_{cot}(\theta) &= \text{KL}(M_\psi(\mathbf{x}) \parallel G_\theta(\mathbf{x})) \\ &= \sum_{i=1}^T M_\psi(y_i) \log \frac{M_\psi(y_i)}{G_\theta(y_i)}, \end{aligned} \quad (3)$$

where  $y_i$  is the  $i$ -th token from the sequence  $\mathbf{x}$ . Thus, the KL-loss distills the output distribution given by the language model to the generator (Hinton, Vinyals, and Dean 2015; Rusu et al. 2016; Yin and Pan 2017). When considering the *mode collapse*, we would only be interested in preserving the distribution for the real data from  $p_{data}$ , rather than those from  $p_\theta$ . Therefore, when optimizing (3), we only adopt samples from the real data distribution  $p_{data}$  to compute the KL-loss. With the above cooperative training loss, the gradient for updating the generator's parameters is derived as follows,

$$\begin{aligned} \nabla_\theta \mathcal{L}_{cot}(\theta) &= \nabla_\theta \left( \sum_{i=1}^T M_\psi(y_i) \log \frac{M_\psi(y_i)}{G_\theta(y_i)} \right) \\ &= - \sum_{i=1}^T M_\psi(y_i) \nabla_\theta \log(G_\theta(y_i)). \end{aligned}$$

As such, the effect of applying cooperative training on the generator is equivalent to increasing the density of the real data in a weighted manner.

### Meta Cooperative Optimization

In this section, we introduce a meta learning paradigm to interleave the optimization of the adversarial training loss  $\mathcal{L}_{adv}(\theta)$  and the cooperative training loss  $\mathcal{L}_{cot}(\theta)$  for the generator model parameters. Unlike the conventional meta-learning approaches that work on achieving faster learning (Finn, Abbeel, and Levine 2017), task generalization (Li et al. 2018) or deriving adaptive models (Al-Shedivat et al. 2018), our intuition is to preserve the generative distribution for the adversarial text generator model to decelerate its *mode collapse*.

To this end, optimizing the adversarial loss  $\mathcal{L}_{adv}(\theta)$  is modelled as a base task, and optimizing the cooperative training loss  $\mathcal{L}_{cot}(\theta)$  is modeled as the meta task. With such setting, the meta optimization scheme ensures that after optimizing the generator parameters  $\theta$  with the adversarial training loss  $\mathcal{L}_{adv}(\theta)$  for increasing generation quality, the resultant parameters would demonstrate considerable resistance towards *mode collapse*, i.e., increasing generation quality while preserving considerable generation diversity.

Formally, we first perform one gradient update on the generator parameters  $\theta$  by optimizing the base task loss:

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{adv}(\theta).$$

Then, we obtain new samples from the real data distribution:  $\mathbf{x} \sim p_{data}$  and inference the meta-loss  $\mathcal{L}_{cot}(\theta')$  for the real samples on the updated parameters  $\theta'$ . The meta gradient is weighted by  $\lambda > 0$  and added to the base task gradient to update the parameters  $\theta$ . Finally, the adversarial update under our proposed meta cooperative training paradigm could be formulated as below:

$$\begin{aligned} \mathcal{L}_D &= \max_{\phi} \mathbb{E}_{\substack{\mathbf{x}_r \sim p_{data}, \\ \mathbf{x}_y \sim p_\theta}} \mathcal{L}_{adv}(\phi) \\ \mathcal{L}_G &= \min_{\theta} \mathbb{E}_{\substack{\mathbf{x}_r \sim p_{data}, \\ \mathbf{x}_y \sim p_\theta}} \left( \mathcal{L}_{adv}(\theta) + \lambda \mathcal{L}_{cot}(\theta') \right) \\ \mathcal{L}_M &= \min_{\psi} \mathbb{E}_{\substack{\mathbf{x}_r \sim p_{data}, \\ \mathbf{x}_y \sim p_\theta}} \mathcal{L}_{cot}(\psi) \end{aligned}$$

The full algorithm for meta cooperative training is presented in Algorithm 1.

---

#### Algorithm 1 Meta Cooperative Training

---

**Input:**  $G_\theta, D_\phi, M_\psi$ , learning rate  $\alpha/\beta/\gamma$ , training data distribution  $p_{data}$

**Output:** Generator  $G_\theta$

---

- 1: Randomly initialize  $\theta, \phi, \psi$
  - 2: Pretrain  $G_\theta$  with samples from  $p_{data}$
  - 3: Assign the weight from  $G_\theta$  to  $M_\psi$
  - 4: **while** not done **do**
  - 5:   Sample  $\mathbf{x}_r \sim p_{data}$
  - 6:   Generate  $\mathbf{x}_f$  with  $G_\theta$
  - 7:   Compute adversarial loss  $\mathcal{L}_{adv}(\theta)$
  - 8:    $\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{adv}(\theta)$
  - 9:   Compute  $M_\psi(\mathbf{x}_r)$  with language model
  - 10:    $g_m = \nabla_\theta \lambda \mathcal{L}_{cot}(M_\psi(\mathbf{x}_r), G_{\theta'}(\mathbf{x}_r))$    ▷ Compute meta gradient
  - 11:    $\theta = \theta - \alpha (\nabla_\theta \mathcal{L}_{adv}(\theta) + \lambda g_m)$    ▷ Generator update
  - 12:    $\phi = \phi - \beta \nabla_\phi \mathcal{L}_{adv}(\phi)$    ▷ Discriminator update
  - 13:    $\psi = \psi - \gamma \nabla_\psi \mathcal{L}_{cot}(\psi)$    ▷ Language model update
  - 14: **end while**
  - 15: **return** Generator  $G_\theta$
- 

## Experiments

We denote our proposed meta cooperative training generative adversarial networks as Meta-CoTGAN. In the experiment section, first, we compare our proposed algorithm with its closest cooperative training counterpart, CoT (Lu et al. 2019) on the synthetic dataset. Then we show the comparison result between our method and several RL-based and RL-free approaches on two commonly used real-world text generation datasets: COCO Image Captions (Chen et al. 2015) and EMNLP 2017 WMT News<sup>1</sup>.

**Implementation Details** We implement our proposed algorithm on top of RelGAN (Nie, Narodytska, and Patel 2019), an RL-free adversarial text generation model that

<sup>1</sup><http://statmt.org/wmt17/translation-task.html>

is among the state-of-the-art approaches. Specifically, RelGAN adopts a relational memory to model the long-distance dependencies among the input tokens, and a *gumbel-softmax* relaxation to overcome the non-differentiable issue in the generator training. The relational memory adopts 1 memory slot, multi-head attention with 2 heads, and the attention key size is set to be 512. The language model for cooperative training adopts the identical network architecture as the generator, and the weights for the generator’s parameters are assigned to the language model after pretraining. The discriminator adopts multiple representations with size to be 64. We adopt Adam (Kingma and Ba 2015) as the optimization algorithm for updating all the model parameters. The source code of our framework is based on PaddlePaddle<sup>2</sup> platform.

**Evaluation Metrics** For comparison, we evaluate the models in terms of *sample quality* and *sample diversity* simultaneously. Following most of today’s text generation works (e.g., (Yu et al. 2017; Lu et al. 2019)), the *sample quality* is evaluated by the BLEU score metrics when testified on real datasets, and  $NLL_{oracle}$  loss when testified on the synthetic dataset. The  $NLL_{oracle}$  loss is defined as the negative log likelihood derived from the target LSTM model for the data generated by  $G_{\theta}$ . The *sample diversity* is evaluated in terms of  $NLL_{gen}$  loss, which is in the following form:

$$NLL_{gen} = -\mathbb{E}_{\mathbf{x}_{1:K}} \log P_{\theta}(\mathbf{x}_1, \dots, \mathbf{x}_K),$$

where the density of the real data is evaluated on the generator model. Thus, models with better sample diversity would have a broader coverage over the real data space and result in lower  $NLL_{gen}$  loss. Models that suffer from severe *mode collapse* would no longer represent the real data well and result in higher  $NLL_{gen}$  loss.

**Baseline Models** To evaluate the efficiency of our proposed approach, we consider MLE as well as the RL-based baselines, including SeqGAN (Yu et al. 2017), RankGAN (Lin et al. 2017) and LeakGAN (Guo et al. 2018). Also, we compare with the most related RL-free baseline RelGAN (Nie, Narodytska, and Patel 2019). During evaluation, we follow the temperature settings proposed in RelGAN and present the results for our method when evaluated with temperature values of 100 and 1000, respectively.

### Synthetic Dataset

Our first evaluation domain is the synthetic oracle dataset, which is first proposed in (Yu et al. 2017). The experiment engages a randomly initialized LSTM model as the target model to simulate real-world sequences and generate data from real data distribution. The synthetic experiments are conducted with the sequence length set to be 20. The objective for experimenting in this domain is to compare our proposed method with its closest cooperative training counterpart CoT. While these two models adopt same way to train the language model, we investigate on the efficiency

of adopting the respective cooperative training losses on the generator model as proposed in these two methods.

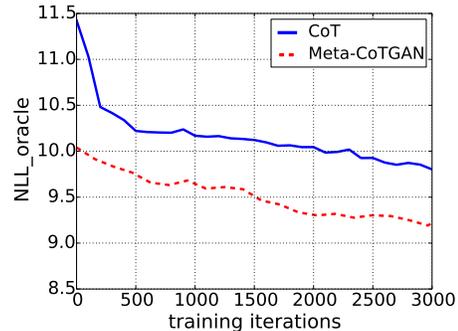


Figure 2: Evaluation result on synthetic oracle with length 20 in terms of  $NLL_{oracle}$  loss. Overall, our model could converge to significantly better standard than CoT.

We demonstrate the learning curves for  $NLL_{oracle}$  loss in Figure 2. Note that CoT takes no pretraining stage and its  $NLL_{oracle}$  loss progressively decreases. Our method takes a pretraining stage and the loss decreases in both the pre-training stage and the adversarial training stage. We could notice that upon convergence, the  $NLL_{oracle}$  loss for our method is significantly lower than CoT. This demonstrates that the cooperative training mechanism proposed by CoT is not comparable to our method in terms of *sample quality*. We also present the evaluation scores for  $NLL_{oracle}$  and  $NLL_{gen}$  in Table 1. When comparing  $NLL_{gen}$ , our method could achieve much lower loss scale than CoT. This demonstrates that our proposed algorithm convey greater efficiency in preserving the *sample diversity*. Overall, considering the inferior performance and long training time of this model, we do not consider it further in the following real-world dataset experiments.

Method	$NLL_{oracle}$	$NLL_{gen}$
CoT	8.19	7.54
Meta-CoTGAN	<b>7.69</b>	<b>6.86</b>

Table 1: Evaluation result on synthetic oracle with sequence length 20. For CoT, we present their *best* score for  $NLL_{gen}$ .

### COCO Image Captions Dataset

Our second evaluation domain is the COCO Image Captions dataset. We follow the pre-processing method proposed in (Zhu et al. 2018). The training and testing set consist of 10, 000 sentences respectively. The sentences in COCO have minimum length of 7 and maximum length of 37. The vocabulary size is 4,682.

We present the scores of BLEU-2 to BLEU-5 for measuring *sample quality*, and the  $NLL_{gen}$  score for measuring *sample diversity* in Table 2. Overall, our method demonstrates significant advantage over all the *sample quality/diversity* metrics. Notably, our method leads to  $NLL_{gen}$  loss significantly lower than the other baseline approaches.

<sup>2</sup><https://www.paddlepaddle.org.cn/>

Method	BLEU-2	BLEU-3	BLEU-4	BLEU-5	NLL <sub>gen</sub>
MLE	0.731	0.497	0.305	0.189	0.718
SeqGAN	0.745	0.498	0.294	0.180	1.082
RankGAN	0.743	0.467	0.264	0.156	1.344
LeakGAN	0.746	0.528	0.355	0.230	0.679
RelGAN (100)	0.849 ± 0.030	0.687 ± 0.047	0.502 ± 0.048	0.331 ± 0.044	0.756 ± 0.054
RelGAN (1000)	0.814 ± 0.012	0.634 ± 0.020	0.455 ± 0.023	0.303 ± 0.020	0.655 ± 0.048
Meta-CoTGAN (100)	<b>0.858 ± 0.003</b>	<b>0.692 ± 0.005</b>	<b>0.518 ± 0.007</b>	<b>0.363 ± 0.009</b>	<b>0.578 ± 0.036</b>
Meta-CoTGAN (1000)	0.842 ± 0.011	0.675 ± 0.019	0.502 ± 0.026	0.349 ± 0.024	0.583 ± 0.028

Table 2: Evaluations on COCO Image Captions dataset. For RelGAN and Meta-CoTGAN, the temperature (in parentheses) is set to be 100 and 1000, and results are averaged over 6 runs (random seeds). For NLL<sub>gen</sub> (last column), the smaller the better.

Method	BLEU-2	BLEU-3	BLEU-4	BLEU-5	NLL <sub>gen</sub>
MLE	0.768	0.473	0.240	0.126	2.382
SeqGAN	0.777	0.491	0.261	0.138	2.773
RankGAN	0.727	0.435	0.209	0.101	3.345
LeakGAN	0.826	0.645	0.437	0.272	2.356
RelGAN (100)	0.881 ± 0.013	0.705 ± 0.019	0.501 ± 0.023	0.319 ± 0.018	2.482 ± 0.031
RelGAN (1000)	0.837 ± 0.012	0.654 ± 0.010	0.435 ± 0.011	0.265 ± 0.011	2.285 ± 0.025
Meta-CoTGAN (100)	<b>0.882 ± 0.014</b>	<b>0.734 ± 0.017</b>	<b>0.542 ± 0.016</b>	<b>0.358 ± 0.015</b>	2.299 ± 0.011
Meta-CoTGAN (1000)	0.868 ± 0.015	0.703 ± 0.014	0.500 ± 0.016	0.318 ± 0.016	<b>2.205 ± 0.053</b>

Table 3: Evaluations on EMNLP2017 WMT News dataset. See the caption of Table 2 for more details.

This indicates that our method could provide an efficient control over the *mode collapse* for the adversarial training and eventually leads to superior *sample diversity*. While decelerating the *mode collapse*, the cooperative training could result in model with better *sample quality* as well.

To further validate this, we present the learning curves for the *sample diversity* metric and BLEU-5 as a representative *sample quality* metric in Figure 3. We could observe that the NLL<sub>gen</sub> for RelGAN would fast go up, which is a sign of *mode collapse*. However, that for MetaCoTGAN progresses rather slowly. It shows that our proposed method could efficiently decelerate *mode collapse* and control the NLL<sub>gen</sub> loss from explode. When investigating on the *sample quality* metric, we could observe the BLEU-5 score for RelGAN would go up faster than MetaCoTGAN. But eventually, our model could achieve a significantly higher standard than RelGAN. Also, we observe that when NLL<sub>gen</sub> for RelGAN explode (e.g., after 400 epochs), the repeat rate is rather high and therefore the generator just becomes useless. However, our method could preserve much better diversity. Also, we observe from the generated real sentences that our model could generate quite long sentences, while most of the GAN models that fall short (Caccia et al. 2018).

### EMNLP2017 WMT News Dataset

Our third evaluation domain is the EMNLP2017 WMT News dataset. The size of this dataset is much larger than Image COCO, involving a training set of 270,000 sentences. The testing set consists of 10,000 sentences. The sentences have maximum length of 51. The vocabulary size is 5,255.

The results for EMNLP dataset are presented in Table 3.

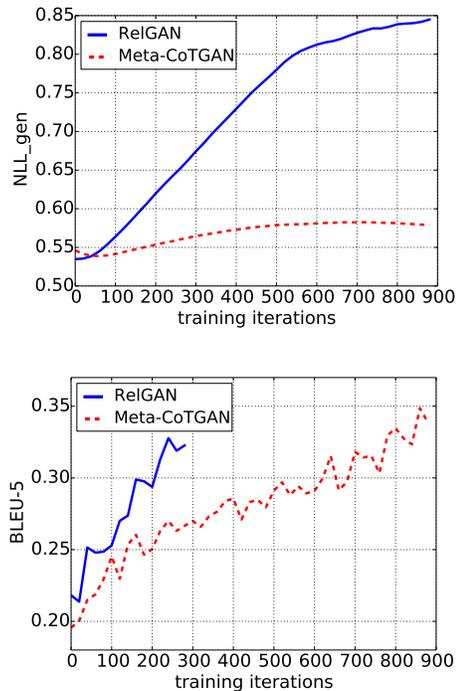


Figure 3: We demonstrate the quality-diversity trade-off for our method as well as the baseline RelGAN on COCO Image Captions dataset. Our model progressively achieves better BLEU-5 score than RelGAN with an apparently slow progress for *mode collapse*. The BLEU-5 for RelGAN is plotted up to the point when its corresponding NLL<sub>gen</sub> loss reaches its reported standard. Otherwise, the BLEU-5 score becomes no more meaningful since the model has turn into severe *mode collapse* (i.e., generating repeated sentences).

Method	BLEU-2	BLEU-3	BLEU-4	BLEU-5	NLL <sub>gen</sub>
RelGAN (100)	0.849 ± 0.030	0.687 ± 0.047	0.502 ± 0.048	0.331 ± 0.044	0.756 ± 0.054
Meta-CoTGAN (100)	<b>0.858 ± 0.003</b>	<b>0.692 ± 0.005</b>	<b>0.518 ± 0.007</b>	<b>0.363 ± 0.009</b>	<b>0.578 ± 0.036</b>
Meta-CoTGAN <sup>cot-off</sup> (100)	0.824 ± 0.011	0.647 ± 0.022	0.466 ± 0.028	0.315 ± 0.022	0.580 ± 0.031
Meta-CoTGAN <sup>meta-off</sup> (100)	0.835 ± 0.013	0.661 ± 0.016	0.487 ± 0.016	0.338 ± 0.014	0.587 ± 0.019
RelGAN (1000)	0.814 ± 0.023	0.634 ± 0.020	0.455 ± 0.023	0.303 ± 0.020	0.655 ± 0.048
Meta-CoTGAN (1000)	0.842 ± 0.011	0.675 ± 0.019	0.502 ± 0.026	0.349 ± 0.024	0.583 ± 0.028
Meta-CoTGAN <sup>cot-off</sup> (1000)	0.824 ± 0.007	0.643 ± 0.009	0.497 ± 0.013	0.324 ± 0.015	0.582 ± 0.017
Meta-CoTGAN <sup>meta-off</sup> (1000)	0.817 ± 0.021	0.638 ± 0.027	0.465 ± 0.025	0.319 ± 0.018	0.589 ± 0.022

Table 4: Ablation study result on COCO Image Captions dataset. We evaluate our proposed model when the cooperative training part and meta optimization have been turned off, respectively. Reported scores are derived from 6 random seeds.

We can see that our proposed method consistently outperforms all baselines in terms of all the BLEU metrics and NLL<sub>gen</sub>. Under the temperature setting of 100, our method outperforms the strong RelGAN baseline by 0.041/0.039 on BLEU-4/BLEU-5. Noticeably, the best BLEU scores for our method are obtained when the NLL<sub>gen</sub> loss is at a significantly lower level than RelGAN. This indicates that by conducting cooperative training, we could derive generator model with better *sample quality* and *sample diversity* simultaneously. Moreover, it shows that our method could robustly perform well in rather challenging and diverse real-world datasets like EMNLP. Meanwhile, the performance of our method is quite robust, consistently outperforming RelGAN under both temperature settings, over all the evaluation metrics. By investigating through the generated real samples, we observe that the generated sentences convey rather diverse semantics and the output consists of considerably long sentences, unlike the conventional adversarial text generators that would shortly fall to the phase of generating short and repeated sentences.

## Ablation Study

**Impact of Cooperative Training Language Model** We demonstrate the impact of using an online updated language model to conduct our proposed cooperative training process. To this end, a direct comparison is to use a pretrained language model not updated with cooperative training. We denote such baseline as Meta-CoTGAN<sup>cot-off</sup>. We demonstrate the result on COCO Image Captions dataset in Table 4. We could observe that when online update to the language model is turned off, the model still preserve comparable *sample diversity* in terms of NLL<sub>gen</sub>, since the cooperative training loss is still employed on the real data. However, under both temperature setting, the *sample quality* metrics could not perform as well as the full set of the proposed method. This shows that it is beneficial to update the language model jointly with the generator to let it offer a smoothly chanting target distribution to the generator.

**Impact of Meta Optimization** We also evaluate the impact of the meta optimization setup. To this end, we compare our approach with a principled way of engaging the cooperative training loss for optimizing the generator parameters, which is proposed in the form of linearly summing up the adversarial loss and the cooperative training loss in a weighted

manner, i.e.,  $\mathcal{L}_{adv}(\theta) + \lambda \mathcal{L}_{cot}(\theta)$ . We denote such baseline as Meta-CoTGAN<sup>meta-off</sup>. The results are shown in Table 4. Overall, Meta-CoTGAN<sup>meta-off</sup> obtain comparable scores for NLL<sub>gen</sub>. However, its performance in terms of the *sample quality* metrics is still much inferior than using full set of solution. Thus, we could conclude that meta optimization is an important ingredient for balancing the quality-diversity trade-off. Intuitively, our proposed meta optimization set-up offers an efficient way to ensure the generator parameters *after* the adversarial update would decelerate from *mode collapse*, which is critical to derive the superior performance.

## Conclusion and Discussion

We propose a meta cooperative training approach to facilitate the training of adversarial text generation models. Our method utilizes a cooperatively trained language model to effectively decelerate the *mode collapse* of adversarial training via distilling the prediction output distribution of the language model over the real data to the adversarial generator model. We evaluate our proposed method in both synthetic dataset and two real-world datasets with sequence length at a range from 7 to 51. As a result, our proposed method could consistently outperform the baseline algorithms on *sample quality* metrics and *sample diversity* metric simultaneously. Our proposed approach is general and could be promising to work with distinct RL-based or RL-free adversarial text generation algorithms as long as they face the issue of *mode collapse*. Our future work would be to apply meta cooperative training on more emerging RL-based/free GAN models.

## References

- [Al-Shedivat et al. 2018] Al-Shedivat, M.; Bansal, T.; Burda, Y.; Sutskever, I.; Mordatch, I.; and Abbeel, P. 2018. Continuous adaptation via meta-learning in nonstationary and competitive environments. In *ICLR*.
- [Arjovsky, Chintala, and Bottou 2017] Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *ICML*, 214–223.
- [Bahdanau, Cho, and Bengio 2015] Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.
- [Caccia et al. 2018] Caccia, M.; Caccia, L.; Fedus, W.; Larochelle, H.; Pineau, J.; and Charlin, L. 2018. Language gans falling short. *arXiv preprint arXiv:1811.02549*.

- [Che et al. 2017] Che, T.; Li, Y.; Zhang, R.; Hjelm, R. D.; Li, W.; Song, Y.; and Bengio, Y. 2017. Maximum-likelihood augmented discrete generative adversarial networks. *arXiv preprint arXiv:1702.07983*.
- [Chen et al. 2015] Chen, X.; Fang, H.; Lin, T.-Y.; Vedantam, R.; Gupta, S.; Dollár, P.; and Zitnick, C. L. 2015. Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*.
- [Chen et al. 2018] Chen, L.; Dai, S.; Tao, C.; Zhang, H.; Gan, Z.; Shen, D.; Zhang, Y.; Wang, G.; Zhang, R.; and Carin, L. 2018. Adversarial text generation via feature-mover’s distance. In *NeurIPS*, 4666–4677.
- [Fedus, Goodfellow, and Dai 2018] Fedus, W.; Goodfellow, I.; and Dai, A. M. 2018. Maskgan: better text generation via filling in the\_. In *ICLR*.
- [Finn, Abbeel, and Levine 2017] Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 1126–1135.
- [Goodfellow et al. 2014] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *NIPS*, 2672–2680.
- [Guo et al. 2018] 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*, 5141–5148.
- [Hinton, Vinyals, and Dean 2015] Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
- [Jang, Gu, and Poole 2017] Jang, E.; Gu, S.; and Poole, B. 2017. Categorical reparameterization with gumbel-softmax. In *ICLR*.
- [Kingma and Ba 2015] Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In *ICLR*.
- [Kusner and Hernández-Lobato 2016] Kusner, M. J., and Hernández-Lobato, J. M. 2016. Gans for sequences of discrete elements with the gumbel-softmax distribution. *arXiv preprint arXiv:1611.04051*.
- [Lamb et al. 2016] Lamb, A. M.; Goyal, A. G. A. P.; Zhang, Y.; Zhang, S.; Courville, A. C.; and Bengio, Y. 2016. Professor forcing: A new algorithm for training recurrent networks. In *NIPS*, 4601–4609.
- [Li et al. 2016] Li, J.; Monroe, W.; Ritter, A.; Galley, M.; Gao, J.; and Jurafsky, D. 2016. Deep reinforcement learning for dialogue generation. In *EMNLP*, 1192–1202.
- [Li et al. 2017] Li, P.; Lam, W.; Bing, L.; and Wang, Z. 2017. Deep recurrent generative decoder for abstractive text summarization. In *EMNLP*, 2091–2100.
- [Li et al. 2018] Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2018. Learning to generalize: Meta-learning for domain generalization. In *AAAI*, 3490–3497.
- [Li, Sun, and Li 2019] Li, X.; Sun, M.; and Li, P. 2019. Multi-agent discussion mechanism for natural language generation. In *AAAI*, 6096–6103.
- [Lin et al. 2017] Lin, K.; Li, D.; He, X.; Zhang, Z.; and Sun, M.-T. 2017. Adversarial ranking for language generation. In *NIPS*, 3155–3165.
- [Liu, Zhong, and Li 2012] Liu, Y.; Zhong, S.-h.; and Li, W. 2012. Query-oriented multi-document summarization via unsupervised deep learning. In *AAAI*.
- [Lu et al. 2018] Lu, S.; Zhu, Y.; Zhang, W.; Wang, J.; and Yu, Y. 2018. Neural text generation: past, present and beyond. *arXiv preprint arXiv:1803.07133*.
- [Lu et al. 2019] Lu, S.; Yu, L.; Zhang, W.; and Yu, Y. 2019. Cot: Cooperative training for generative modeling of discrete data. In *ICML*, 4164–4172.
- [Nie, Narodytska, and Patel 2019] Nie, W.; Narodytska, N.; and Patel, A. 2019. Relgan: Relational generative adversarial networks for text generation. In *ICLR*.
- [Radford, Metz, and Chintala 2015] Radford, A.; Metz, L.; and Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.
- [Rajeswar et al. 2017] Rajeswar, S.; Subramanian, S.; Dutil, F.; Pal, C.; and Courville, A. 2017. Adversarial generation of natural language. *arXiv preprint arXiv:1705.10929*.
- [Rusu et al. 2016] Rusu, A. A.; Colmenarejo, S. G.; Gulcehre, C.; Desjardins, G.; Kirkpatrick, J.; Pascanu, R.; Mnih, V.; Kavukcuoglu, K.; and Hadsell, R. 2016. Policy distillation. In *ICLR*.
- [Sun, Li, and Li 2018] Sun, M.; Li, X.; and Li, P. 2018. Logician and orator: Learning from the duality between language and knowledge in open domain. In *EMNLP*, 2119–2130.
- [Sutskever, Vinyals, and Le 2014] Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *NIPS*, 3104–3112.
- [Sutton et al. 2000] Sutton, R. S.; McAllester, D. A.; Singh, S. P.; and Mansour, Y. 2000. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*, 1057–1063.
- [Ulyanov, Vedaldi, and Lempitsky 2016] Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*.
- [Wen et al. 2015] Wen, T.-H.; Gasic, M.; Mrksic, N.; Su, P.-H.; Vandyke, D.; and Young, S. 2015. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. In *EMNLP*, 1711–1721.
- [Williams 1992] Williams, R. J. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning* 8(3-4):229–256.
- [Xu et al. 2018] Xu, J.; Ren, X.; Lin, J.; and Sun, X. 2018. Dp-gan: diversity-promoting generative adversarial network for generating informative and diversified text. *arXiv preprint arXiv:1802.01345*.
- [Yin and Pan 2017] Yin, H., and Pan, S. J. 2017. Knowledge transfer for deep reinforcement learning with hierarchical experience replay. In *AAAI*, 1640–1646.

- [Yu et al. 2017] Yu, L.; Zhang, W.; Wang, J.; and Yu, Y. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *AAAI*, 2852–2858.
- [Zhang et al. 2017] Zhang, Y.; Gan, Z.; Fan, K.; Chen, Z.; Henao, R.; Shen, D.; and Carin, L. 2017. Adversarial feature matching for text generation. In *ICML*, 4006–4015.
- [Zhu et al. 2018] Zhu, Y.; Lu, S.; Zheng, L.; Guo, J.; Zhang, W.; Wang, J.; and Yu, Y. 2018. Texus: A benchmarking platform for text generation models. In *SIGIR*, 1097–1100.