경매 메커니즘을 이용한 다중 적대적 생성 신경망 학습에 관한 연구

심주용 [1], 최진성 [1], 김종국 [2]
[1] 고려대학교 전기전자학과 박사과정
[2] 고려대학교 전기전자공학과 교수

shimjoo@korea.ac.kr, garangg@korea.ac.kr, jongkook@korea.ac.kr

A Study on Auction-Inspired Multi-GAN Training

Joo Yong Shim[1], Jean Seong Bjorn Choe[1], Jong-Kook Kim[1]
[1]Dept. of Electrical and Computer Engineering, Korea University

Abstract

Generative Adversarial Networks (GANs) models have developed rapidly due to the emergence of various variation models and their wide applications. Despite many recent developments in GANs, mode collapse, and instability are still unresolved issues. To address these problems, we focused on the fact that a single GANs model itself cannot realize local failure during the training phase without external standards. This paper introduces a novel training process involving multiple GANs, inspired by auction mechanisms. During the training, auxiliary performance metrics for each GANs are determined by the others through the process of various auction methods.

1. Introduction

Various studies of Generative Adversarial Networks (GANs) [1] models were developed in a short period of time. Nevertheless, the limitations of GANs are still unresolved [2],[3]: mode collapse and convergence failure. There have been a lot of efforts to solve these problems, which tried to improve the learning of GANs [2],[3],[4] or enforcing GANs to cover diverse modes [5],[6]. Although the problem has been alleviated to some extent, development challenges still remain.

We focus on resolving the mode collapse problem which is a phenomenon that a generator produces only the same image or a small subset of images without diversity. It occurs when a generator captures only a few modes of data distribution, which means that generator maps input noises to the same or only few points.

In this paper, we propose a novel approach to find the failure in advance and prevent during the training stage. Our approach gives the proper reference during training, and thus recognize the signs of failure and helps to get back on the right track. Usually, identification of the failure mode of GANs is done by monitoring the loss curves or generated images manually. To the best of our knowledge, this is the first attempt to diagnose the failure mode of GANs during training.

It is difficult for a single GANs model itself to realize local failure during the training phase. Thus, we thought it would be helpful to give an appropriate external reference value while training. Simply using the optimal generator or discriminator are not helpful for the adversarial training due to its nature characteristic [7]. It is important to keep them in the same phase to get the proper reference value.

Our approach includes auxiliary training that uses the reference value of the best scored GANs among multiple GANs in the same training phase. The score is determined by using the values obtained by the auction-like process.

The main contributions of the proposed method can be summarized as follows: (1) This paper provides the novel idea on training multiple GANs effectively by adopting auction approach, which is expected to reduce the mode collapse problem. (2) We adopt new update process (auxiliary training) to properly guide the training performance of multiple GANs. (3) The impacts of training multi-GANs with proposed algorithm are examined by comparing with single-trained GANs.

2. Proposed Approach

We consider a set of n GANs $\{(G_1,\ D_1), (G_2,\ D_2), \ldots, (G_N,\ D_N)\}$, which are trained simultaneously. Each GANs are trained to maximize their performance through auction based two-stage update algorithm. The overall algorithm is described in Algorithm 1.

Algorithm 1: Overall Algorithm

for *number of iterations* **do**
    for all $GANs$**do**
        Ordinary Train
    end for
    do *Auction* and **select** the *Best GAN*
    for all $GANs$ *except the Best GAN* **do**
        Auxiliary Train
    end for
end for

The first-step update is done individually in the same way as the general GANs training process. For example, if we use vanilla-GANs, each GANs is updated according to the objective function:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1-D(G(z))],$$

where $x\sim p_{data}$ is image sample from the real data distribution, $z\sim p_z$ is a prior input noise, which are generated randomly.

The second-step update is done through the auxiliary training process, which refers to the best GANs in the same learning phase. It is expected the performance of the GANs can be calibrated during training, by calculating the relative loss that refers to the outputs of the best GANs.

The best GANs is obtained by using the bid value submitted by participants for images presented by the auctioneer in the auction process. In the auction, all generators become auctioneers and present a set of generated images as a lot for the auction, and all discriminators become participants and submits bids for images.

Auctions are held for each generator(auctioneer), thus a total of *n* auctions will be held. In every auction phase, all of the discriminators(participants) output a numerical value that determine how real is the generated images and use them as bidding values. The winner of each auction can be decided by using winner selection methods, but we skip this process because the purpose of this auction process is to get the proper valuation of the generated images. Auction phase can be summarized as:

1) The auctioneer *i* presents a lot I^i, which is the group of *k* items generated by the generator G_i

$$I^i = (I_1^i, I_2^i, \dots, I_k^i)$$
$$I_k^i \leftarrow G_i(z_k).$$

2) Each bidder *j* (*j =1, …., N, j≠i*) is asked to evaluate *k* items from auctioneer *i*. The value of the item is the output of the discriminator D_j. The final bid X^{ij} is represented as the mean of values for all items.

$$X^{ij} = \frac{1}{K}(X_1^{ij}, X_2^{ij}, \dots, X_k^{ij})$$
$$X_k^{ij} \leftarrow D_j(I_k^i)$$

Based on the auction result, the score of the $GANs\ i = (G_i, D_i)$ is determined with score function $S(i)$. We use the mean difference of the bid prices received from all other participants for images that G_i created, and the bid price submitted by D_i for images presented by all other auctioneers. The score $S(i)$ is formulated as following:

$$S(i) = \frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N} X^{ij} - \frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N} X^{ji}$$

Here, $\sum \boldsymbol{X^{ij}}$ is the sum of the values evaluated by all other discriminators D_j ,($j \in \{1, \dots, N\},\ j\neq i$) for the images generated by G_i. This represents the sum of the probability values that discriminators of other GANs thought that the images generated by G_i are from the real. Images generated by relatively poor GANs are more likely to be determined as fake images, that is, the value of $D_j(I_k^i)$ is more likely to be low, thus resulting lower X^{ij} values. $\sum \boldsymbol{X^{ji}}$ is the sum of the values that discriminator D_i evaluated for the images generated by all other generators G_j ,($j \in \{1, \dots, N\},\ j\neq i$). Similarly, this means the sum of probability values that D_i thought as the real for the images generated by other generators. If the GANs *i* has relatively poor in performance, it is highly likely that image generated by other GANs $I_k^i \leftarrow G_i(z_k)$ will be judged as relatively good images, that is, $D_i(I_k^j)$ is more likely to be high. Therefore, the worse the GAN, the higher the value of X^{ji} will be measured. Taken together, the score is designed as the above formula *S(i),* and the better the performance, the higher the score.

When the best GAN is determined, the auxiliary training process is conducted. Firstly, we define relative real loss L_{real} which compare the value of actual image evaluated by the best GAN's discriminator $D^*(X)$ with the value of the actual image evaluated by itself $D(X)$. It is expected to help the performance of the discriminator *D* by providing a calibration value of how the output differs from the output determined by the best GAN. Similarly, we also define relative fake loss L_{fake} which uses the discriminator output on generated images. The discriminator, which is trained with its paired generator, tend to be easily deceived with only one-sided data generated by its paired generator and fall into a mode collapse. However, other discriminators, which are trained with their own-paired generators, are expected that the distribution of data that is determined as the real is likely to be elsewhere, so they do not easily deceive by one-sided data. Using these characteristics, the discriminator can correct its value by referring the value determined by the best GAN's discriminator, so that the generator can also refer to the value determined by the best discriminator. Auxiliary Training process is described in Algorithm 2.

Algorithm 2: Auxiliary Train
Sample $X = \{x^{(1)}, \dots, x^{(m)}\}$ from the real data distribution p_{data} Calculate *Relative real loss:* $L_{real} = L(D(X), D^*(X))$ Sample $Z = \{z^{(1)}, \dots, z^{(m)}\}$ from the noise prior p_z Generate images $\bar{X} \leftarrow G(Z)$ Calculate *Relative fake loss:* $L_{fake} = L(D(\bar{X}), D^*(\bar{X}))$ Update D with *Relative Loss* $L = L_{real} + L_{fake}$

Here, we used MSE loss for relative losses, thus the real relative loss will be as:

$$L_{real} = L\big(D(X), D^*(X)\big)$$
$$= \frac{1}{m}\sum_{i=1}^{m}\big(D(x^{(i)}) - D^*(x^{(i)})\ \big)^2$$

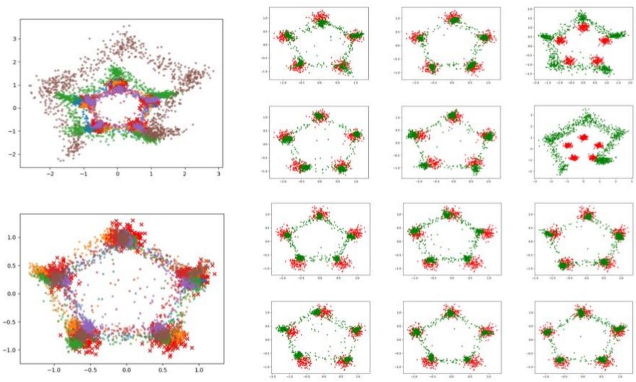
And similarly, the fake relative loss will be as:

$$L_{fake} = L(D(\bar{X}), D^*(\bar{X}))$$
$$= \frac{1}{m}\sum_{i=1}^{m}\big(\ D\big(G(z^{(i)})\big) - D^*\big(G(z^{(i)})\big)\big)^2$$

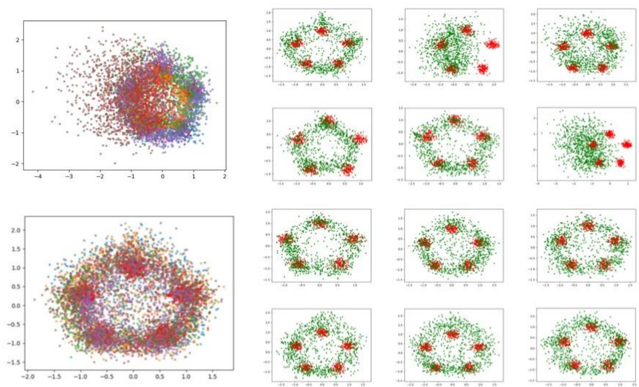
3. Results

We present analysis of proposed algorithm on two classical GANs models: vanilla GANs [1] and WGANs [8]. We used simple 2D-Gaussian Dataset for the experiment, because mode collapse is more evident in this simple dataset.

In the experiment we trained 6 GANs with proposed algorithm. As shown in the Fig.1 and Fig.2, the GANs trained with our method applied cover various points without falling into mode collapse, while GANs simply trained in the standard way are often biased to one side or cannot find data points in some areas.



(Figure 1) Result on multiple vanilla GANs trained without our method(top) and with our method applied(bottom).



(Figure 2) Result on 2D-Gaussian Dataset of multiple vanilla GANs trained without our method(top) and with our method applied(bottom).

4. Conclusion and Discussion

In this paper, we proposed a novel idea on training multiple GANs. In our approach, new auxiliary training process is adopted. We selected the best GAN through the auction like valuation process and is used as the reference. The performance of the discriminator is calibrated by referencing the value of the best GANs.

Although we have solved mode collapse and instability problem, there is much more space to improve. Compared to the other studies such as [4], [5], [9], our method needs more computational efforts. Performance alleviation and implementation on other GANs models will be the future work. Better bid function or using some other auction mechanisms might be helpful. Updating with relative loss also needs to be further investigated. We leave all of the limitations as a future work to be done.

References

[1] Goodfellow, Ian J., Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron C., and Bengio, Yoshua. “Generative adversarial nets.” In Advances in Neural Information Processing Systems, 2014.

[2] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. “Improved techniques for training gans.” In Advances in Neural Information Processing Systems, 2016.

[3] M. Arjovsky and L. Bottou. “Towards principled methods for training generative adversarial networks.” In International Conference on Learning Representations, 2017.

[4] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein. “Unrolled generative adversarial networks.” In International Conference on Learning Representations, 2017.

[5] Ghosh, Arnab, et al. "Multi-agent diverse generative adversarial networks." In Computer Vision and Pattern Recognition. 2018

[6] Choi, J., and Han, B. "Mcl-gan: Generative adversarial networks with multiple specialized discriminators." In Advances in Neural Information Processing Systems, 2022.

[7] Farnia, F., and Ozdaglar, A. “Do GANs always have Nash equilibria?”. In International Conference on Machine Learning, 2020.

[8] Arjovsky, M., Chintala, S., and Bottou, L. “Wasserstein generative adversarial networks.” In International Conference on Machine Learning, 2017.

[9] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. “Improved training of wasserstein gans.” In Advances in Neural Information Processing Systems, 2017.