

<https://doi.org/10.7236/JIIBC.2020.20.1.239>
JIIBC 2020-1-34

정칙화 항에 기반한 WGAN의 립슈츠 연속 안정화 기법 제안

Technique Proposal to Stabilize Lipschitz Continuity of WGAN Based on Regularization Terms

한희일*

Hee-Il Hahn*

요 약 최근에 제안된 WGAN(Wasserstein generative adversarial network)의 등장으로 GAN(generative adversarial network)의 고질적인 문제인 까다롭고 불안정한 학습과정이 다소 개선되기는 하였으나 여전히 수렴이 안되거나 자연스럽지 못한 출력물을 생성하는 등의 경우가 발생한다. 이러한 문제를 해결하기 위하여 본 논문에서는 분별기가 실제 데이터 확률분포를 보다 정확히 추정할 수 있도록 표본화 과정을 개선하는 동시에 분별기 함수의 립슈츠 연속 조건을 안정적으로 유지시키기 위한 알고리즘을 제안한다. 다양한 실험을 통하여 제안 기법의 특성을 분석하고 성능을 확인한다.

Abstract The recently proposed Wasserstein generative adversarial network (WGAN) has improved some of the tricky and unstable training processes that are chronic problems of the generative adversarial network(GAN), but there are still cases where it generates poor samples or fails to converge. In order to solve the problems, this paper proposes algorithms to improve the sampling process so that the discriminator can more accurately estimate the data probability distribution to be modeled and to stably maintain the discriminator should be Lipschitz continuous. Through various experiments, we analyze the characteristics of the proposed techniques and verify their performances.

Key Words : Generative Model, Lipschitz Continuity, Training Stability, Wasserstein Distance, Wasserstein GAN

1. 서 론

최근 들어 신경망을 이용한 생성모델(generative model)에 대한 관심이 급증하여 다양한 연구결과들이 발표되고 있다. 이전의 딥러닝 분야에서는 분별모델

(discriminative model) 중심으로 연구가 진행된 반면, 최근에는 이의 역 과정인 생성모델을 학습하는데 많은 연구가 집중되는 경향을 보인다. 분별모델은 주어진 학습 데이터 집합과 레이블 정보를 기반으로 학습하는 지도학습에 주로 기반한다[12, 13]. 임의의 입력데이터가 분별

*정회원, 한국외국어대학교 공과대학 정보통신공학과
접수일자: 2019년 12월 27일, 수정완료: 2020년 1월 24일
게재확정일자: 2020년 2월 7일

Received: 27 December, 2019 / Revised: 24 January, 2020 /
Accepted: 7 February, 2020

*Corresponding Author: hihahn@hufs.ac.kr

Dept. Information and Communications Eng., College of
Engineering, Hankuk University of Foreign Studies, Korea

기에 입력되면 그 데이터가 속한 부류의 레이블을 매핑하는 응용분야는 딥러닝이 가장 성공을 거둔 대표적인 분야에 속한다. 최근에는 이의 역 과정인 생성모델(generative model)을 학습하는데 많은 연구가 집중되는 경향을 보인다.

생성모델은 일련의 랜덤변수 수열 x_n 에 대하여 x_n 의 확률분포 $F_n(x)$ 가 $n \rightarrow \infty$ 에서 $F(x)$ 로 수렴할 때 $x_n \rightarrow x$ 는 가장 약한 개념의 수렴(convergence in distribution)에 속한다. 따라서 임의의 사건 ξ 에 해당되는 $x_n(\xi)$ 에 대한 수렴이 항상 보장되지는 않으므로 생성기의 성능은 제한적일 수밖에 없다. 분별모델을 학습하기 위하여 학습데이터 집합의 확률분포를 굳이 알아낼 필요는 없지만 생성모델은 기본적으로 확률분포 추정이 선행되어야 학습이 가능하다.

최근에는 우도를 명시적으로 구하지 않으면서도 원하는 확률분포를 갖는 표본을 생성하는 기법들이 제안되고 있는데, 이의 대표적 모델이 GAN[1]이다. GAN은 게임이론에 기반하여 생성모델을 학습하기 위한 기법으로서 생성기는 랜덤잡음 벡터를 입력받아 원하는 데이터 분포를 갖는 표본을 생성하는 생성기(generator) 구현을 목표로 한다. 이 때 생성된 표본은 실제 표본과 육안으로도 구별할 수 없을 정도로 정교해야 한다. 이를 위해서는 생성된 표본과 실제 표본을 정확히 분류할 수 있는 분류기(discriminator)가 필수적인데 이의 출력이 생성기를 학습시키는 입력신호로 활용되므로 분류기의 정확도가 생성기의 성능을 결정한다고 볼 수 있다.

생성기의 확률분포 \mathbb{P}_g 를 학습하기 위해서는 우선 생성기의 입력신호로 이용될 랜덤잡음의 확률분포 $\mathbb{P}_z(z)$ 가 정의되어야 한다. 생성기는 함수 $G(z; \theta)$ 로 나타내는데, 여기서 G 는 미분가능하고 매개변수 θ 를 갖는 딥러닝 알고리즘으로 구현된다. GAN에서 사용하는 생성기는 가변 자동인코더(variational autoencoder)[2]와 기능면에서 상당히 유사하다. 랜덤변수 $z \sim \mathbb{P}_z(z)$ 를 적절히 표본화하여 이를 생성기 G 에 입력시킴으로써 원하는 데이터 분포 \mathbb{P}_d 를 갖는 출력을 생성하도록 학습한다. 반면에 또 다른 딥러닝으로 구현되는 분별기 $D(x; w)$ 는 학습데이터와 생성기 출력을 입력받으면 각각 1과 0을 출력하도록 학습된다. 이와 같이 생성기와 분별기는 모두 딥러닝으로 구현되는 비선형 함수이고 학습데이터의 확률분포는 두 함수 내에 근사적으로 내재된다. 문제는 매개변수가 수렴함에 따라 해당되는 확률분포도 함께 수렴

하는가에 있다.

Radford, *etal*[3]은 DMLP 대신 CNN을 사용한 DCGAN(deep convolutional GAN)을 제안하여 이미지 생성 등의 응용에 특히 성능이 향상되는 이득은 있으나 안정적인 학습을 어렵게 하는 GAN 고유의 문제점을 해결하지는 못하고 있다. 이는 GAN이 JS 발산(Jensen-Shannon divergence)을 최소화시키는 작업을 수행하는데, 이로 인해 매끄러운 수렴이 보장되지 않기 때문이다.

GAN은 여러 성공적인 응용에도 불구하고 학습의 불안정 원인을 해결하기 위한 대책은 아직까지 미흡한 상태이다. 이러한 GAN의 학습 안정성을 개선하기 위하여 Arjovsky, *etal*[4]은 Wasserstein 거리를 손실함수로 채택한 WGAN을 제안한다. Wasserstein 거리는 JS나 KL 발산(Kullback Leibler divergence)보다 약한 개념의 거리함수이기 때문이다. 거리함수가 약할수록 매개변수 열을 이에 대응한 확률분포 열로의 연속적인 매핑을 구하기 쉽고 따라서 분포 열이 보다 쉽게 수렴한다[14].

GAN을 학습시키는 과정은 매우 까다로워서 응용대상에 따라 다른 방법을 적용해야 할 정도로 일반화시키기 어려운 단점이 있다. 게다가 학습 데이터에 포함된 각 부류의 패턴을 골고루 생성하지 못하고 제한된 패턴만 생성하는 모드붕괴(mode collapse)와 모드 드로핑(mode dropping) 현상이 발생하는 것으로 알려져 있다[5]. WGAN은 GAN의 이러한 약점을 줄이고 분별기와 생성기에 대한 보다 안정적인 학습을 가능하게 하는 것으로 알려져 있다. 그런데 문제는 생성기 함수 G 가 립쉬츠 연속이어야 \mathbb{P}_d 와 \mathbb{P}_g 간의 Wasserstein 거리 $W(\mathbb{P}_d, \mathbb{P}_g)$ 가 연속임을 보장할 수 있다. 이를 위하여 Arjovsky, *etal*[4]는 딥러닝으로 분별기를 학습하는 모든 과정에서 가중치를 고정된 임계값으로 클리핑하는 기법을 도입한다.

이 방식은 가중치를 강제로 클리핑함으로써 분별기가 지나치게 간략화된 함수로 수렴하고 학습 데이터 분포의 고차원 모멘트를 왜곡하는 것으로 알려져 있다. Gulrajani, *etal*[6]는 함수의 미분크기가 1에서 벗어날수록 증가하는 정칙화(regularization) 항을 추가함으로써 1-립쉬츠 연속함수를 구하는 알고리즘(이하 WGAN-GP로 명명함)을 제안한다. 미분크기가 1 이하이면 1-립쉬츠 함수임에도 불구하고 함수의 미분크기가 1이 되도록 제한함으로써 립쉬츠 연속을 안정적으로 보장하지 못할 확률이 높다.

WGAN-GP 방식은 실제 데이터와 생성기 출력물 간

의 임의의 한 점 \hat{x} 을 선택해서 그 점에 대해서만 정칙화항을 적용하기 때문에 그 외의 다른 영역에서 립쉬츠 연속을 기대하기 어려운 문제가 있다. 또한, 동일한 학습회수 내에서도 각 미니배치(mini-batch)마다 계산된 립쉬츠 연속 확률의 변화율이 매우 크므로 이를 줄이기 위하여 Wei, et al[7]는 WGAN-GP에서 채택된, 실제 데이터와 생성기 출력물 간의 임의의 한 점 외에도 드롭아웃을 통하여 실제 데이터 인근의 여러 점에서의 분별기 출력을 추가로 이용하는 방법(이하 WGAN-CT로 명명함)을 제안한다.

이 방식은 데이터 확률분포 \mathbb{P}_d 를 보다 조밀하게 표본화함으로써 보다 안정적인 립쉬츠 연속을 구할 수 있는 장점이 있다. 하지만 기본적으로 WGAN-GP와 동일한 정칙화항을 이용하므로 성능의 한계를 피할 수 없는 문제를 보인다. 본 논문에서는 이러한 문제를 해결하기 위한 기법을 제안하며 실험을 통하여 기존의 대표적인 알고리즘과 그 성능을 비교분석한다.

본 논문의 구성은 다음과 같다. II장에서는 관련연구를 리뷰하고, 제안모델을 III장에서 설명한다. IV장에서는 기존의 대표적인 WGAN 알고리즘들과 제안 알고리즘의 성능에 대하여 실험결과를 기반으로 비교분석한다. 마지막으로 V장에서는 결론을 맺고 향후 연구 진행방향에 대하여 논의한다.

II. 관련 연구

1. GAN

GAN은 학습하기가 매우 까다롭고 대부분의 응용에서는 heuristics에 의존하는 형편이다. GAN은 다음 식에 주어진 비용함수를 최적화시키기 위하여, 분별기는 $\log(D(x))$ 를 최대화시키고 동시에 생성기는 $\log(1 - D(G(z)))$ 를 최소화하는 학습을 수행한다[2].

$$\min_G \max_D \left\{ \begin{array}{l} E_{x \sim \mathbb{P}_d} [\log D(x)] + \\ E_{x \sim \mathbb{P}_g} [\log(1 - D(\tilde{x}))] \end{array} \right\} \quad (1)$$

이 때 최적의 분별기 $D^*(x)$ 는 다음과 같이 유도된다[2].

$$D^*(x) = \frac{\mathbb{P}_d(x)}{\mathbb{P}_d(x) + \mathbb{P}_g(x)} \quad (2)$$

$D^*(x)$ 가 구해지면 이에 따라 식 (1)을 최소화시키는 생성기 G 를 구한다. 그런데 실제로 학습해보면, 분별기가 잘 학습됨에 따라 생성기는 오히려 성능이 저하되는 불안정한 현상이 많이 노출된다. 분별기가 생성기보다 먼저 최적적으로 학습되어 있는 상태에서 손실함수를 최소화시키는 학습을 진행하면 \mathbb{P}_d 와 \mathbb{P}_g 간의 JS 발산(Jensen-Shannon divergence)을 최소화시키는 작업으로 귀결된다[1]. 생성기 출력의 확률분포 \mathbb{P}_g 는 학습을 통하여 매개변수를 최적화시킴으로써 데이터집합의 확률분포 \mathbb{P}_d 에 안정적으로 수렴하여야 한다. 즉, 매 학습단계마다 조정되는 θ 의 열 $\theta_1, \theta_2, \dots$ 가 θ^* 로 수렴함에 따라 이에 대응한 확률분포의 열 $\mathbb{P}_g^1, \mathbb{P}_g^2, \dots$ 가 연속적으로 $\mathbb{P}_g (\approx \mathbb{P}_d)$ 에 수렴하여야 좋은 성능을 기대할 수 있다. 이를 위해서는 두 분포 \mathbb{P}_d 와 \mathbb{P}_g 간의 거리개념인 손실함수가 매개변수 θ 에 연속이어야 한다. 그런데 GAN이 최소화시키는 손실함수인 JS 발산은 수렴의 연속성을 보장하지 못하기 때문에 이의 안정적 학습이 어렵고 성능을 저해하는 원인으로 작용한다. GAN에서는 식 (1)을 최적화시키기 위한 알고리즘은 결국 \mathbb{P}_d 와 \mathbb{P}_g 간의 JS 발산을 최소화시키는 것과 동치임이 증명되었다[2]. 하지만 JS 발산은 수렴의 연속성을 보장하지 못하기 때문에 이러한 문제를 개선하기 위하여 Wasserstein 거리를 손실함수로 채택한 WGAN이 제안된다.

2. WGAN

Wasserstein 거리함수 $W(\mathbb{P}_d, \mathbb{P}_g)$ 는 다음과 같이 정의된다.

$$W(\mathbb{P}_d, \mathbb{P}_g) = \inf_{\gamma \in \pi(\mathbb{P}_d, \mathbb{P}_g)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (3)$$

여기서 $\pi(\mathbb{P}_d, \mathbb{P}_g)$ 는 모든 조인트 확률분포 $\gamma(x, y)$ 의 집합으로서 $\mathbb{P}_d = \int \gamma(x, y) dy$ 이고 $\mathbb{P}_g = \int \gamma(x, y) dx$ 이다. \mathbb{P}_d 와 \mathbb{P}_g 의 확률측도(probability measure)를 각각 μ 와 ν 로 정의하면 식 (3)은 Kantorovich-Rubinstein 이중성[8]에 의해 다음과 같이 나타낼 수 있다.

$$W(\mathbb{P}_d, \mathbb{P}_g) = \sup_{\|f\|_L \leq 1} \{ \mathbb{E}_{s \sim \mu} [f(s)] - \mathbb{E}_{t \sim \nu} [f(t)] \} \quad (4)$$

여기서 $\|f\|_L \leq 1$ 는 함수 f 가 1-립쉬츠 연속임을 의미한다. k -립쉬츠 연속함수란 정의구역 내에서 다음 식을 만족시키는 함수를 의미한다.

$$|f(y) - f(x)| \leq k|y - x| \quad (5)$$

결과적으로 식 (4)를 만족시키는 1-립쉬츠 연속함수 f 를 구하면 되는데 신경망을 통하여 구현할 수 있다. Arjovsky, et al[4]는 함수 f 가 1-립쉬츠 연속임을 유지하기 위하여 f 의 가중치를 다음과 같이 일방적으로 클리핑시킨다.

$$w = \begin{cases} c & w \geq c \\ -c & w \leq -c \end{cases} \quad (6)$$

여기서 $c=0.01$ 로 고정된다. 이 방식은 가중치를 강제로 클리핑함으로써 분별기가 지나치게 간략화된 함수로 수렴하고 학습 데이터 분포의 고차원 모멘트를 왜곡하는 것으로 알려져 있다[5]. 이러한 문제점을 보완하기 위하여 Gulrajani, et al[6]는 식 (4)에 주어진 함수에 부호를 변경하여 손실함수로 변환한 다음 정칙화 항을 다음과 같이 추가함으로써 1-립쉬츠 연속함수를 구하는 WGAN-GP 방식을 제안한다.

$$\mathbb{E}_{t \sim \mathbb{P}_y}[f(t)] - \mathbb{E}_{s \sim \mathbb{P}_r}[f(s)] + \lambda \mathbb{E}_{v \sim \mathbb{P}_v}[(\|\nabla_v f(v)\| - 1)^2] \quad (7)$$

위 식에서 v 는 s 와 t 를 연결하는 직선 상에서 랜덤하게 선택한 점, 즉 $v = \alpha s + (1 - \alpha)t$, $0 \leq \alpha \leq 1$ 이고 \mathbb{P}_v 는 그러한 v 의 확률분포이다. 우식의 앞 두 항은 식 (4)의 부호를 변경한 것이고 세 번째 항은 f 가 1-립쉬츠 연속이기 위한 정칙화 항이다. 이 항을 자세히 살펴보면 f 의 미분크기가 1이 아니면 이에 비례한 페널티 (gradient penalty)를 부과함으로써 f 의 미분크기가 1을 유지하도록 하는 역할을 수행한다.

III. 제안 모델

함수 f 의 미분크기가 1 이하를 유지하면 그 함수는 1-립쉬츠 연속함수이다. 그런데 WGAN-GP는 함수의 미분크기가 1이 되도록 제한함으로써 립쉬츠 연속을 안

정적으로 보장하지 못할 확률이 높다. 이를 해결하기 위하여 본 논문에서는 함수의 1-립쉬츠 연속을 보다 안정적으로 보장하는 기법을 적용하여 WGAN의 성능을 향상시키는 방법을 제안한다. f 의 미분크기가 1을 유지하면 f 가 1-립쉬츠 연속이기는 하지만, 1 이하에서도 1-립쉬츠 연속이므로 이 방식 또한 1-립쉬츠 연속을 안정적으로 보장하기 어렵다. 이를 개선시키기 위하여 본 논문에서는 식 (7)에서 제시한 정칙화 항 $\mathbb{E}_{v \sim \mathbb{P}_v}[(\|\nabla_v f(v)\| - 1)^2]$ 대신에 다음과 같은 정칙화 항을 제안한다.

$$\mathbb{E}_{v \sim \mathbb{P}_v}(\|\nabla_v f(v)\| - 1)^2 u(\|\nabla_v f(v)\| - 1) \quad (8)$$

여기서 $u(\cdot)$ 는 단위계단함수이다. $\|\nabla_v f(v)\| = 1$ 을 유지하도록 학습하는 것보다 $\|\nabla_v f(v)\| \leq 1$ 이 되도록 학습하는 것이 보다 식 (4)에 주어진 이론에 부합할 뿐만 아니라 쉽고 안정적인 학습을 기대할 수 있다. 또한, WGAN-GP 방식은 실제 데이터와 생성기 출력물 간의 임의의 한 점 v 를 선택해서 그 점에 대해서만 페널티를 적용하기 때문에 그 외의 다른 영역에서 립쉬츠 연속을 기대하기 어려운 문제가 있다.

이를 개선하기 위하여 Wei, et al[7]는 식 (7)에 주어진 WGAN-GP의 손실함수에 보조 정칙화 항(consistency term)을 추가하여 다음과 같은 손실함수를 갖는 WGAN-CT를 제안한다.

$$\mathbb{E}_{t \sim \mathbb{P}_y}[f(t)] - \mathbb{E}_{s \sim \mathbb{P}_r}[f(s)] + \lambda_1 \mathbb{E}_{v \sim \mathbb{P}_v}[(\|\nabla_v f(v)\| - 1)^2] + \lambda_2 CT(s_1, s_2) \quad (9)$$

여기서 보조 정칙화 항은 다음과 같이 정의된다.

$$CT(s_1, s_2) = \mathbb{E}[d(D(s_1), D(s_2)) + d(D(s_1^*), D(s_2^*))] \quad (10)$$

위 식에서, $D(s_i)$, $D(s_i^*)$ 는 학습데이터 s 에 대하여 서로 다른 드롭아웃을 적용하였을 때의 분별기 출력을 나타내는데, 이를 두 번 반복 적용하여 이들 간의 거리 평균을 별도의 정칙화 항으로 이용한다. 위 식의 의미는 s 와 s_i , s_i^* 는 서로 인접한 것으로 간주하여 동일한 드롭아웃을 적용한 분별기 출력 간의 거리도 충분히 작아야 한다는 조건을 추가함으로써 분별기 함수를 최대한 매끄

러운 함수로 제한하기 위함이다.

본 논문에서는 식 (8)과 식 (9)의 항을 결합하여 다음과 같은 손실함수를 제안한다.

$$\begin{aligned} & \mathbb{E}_{t \sim \mathbb{P}_g} [f(t)] - \mathbb{E}_{s \sim \mathbb{P}_d} [f(s)] + \\ & \lambda_1 \mathbb{E}_{v \sim \mathbb{P}_g} (\| \nabla_v f(v) \| - 1)^2 u(\| \nabla_v f(v) \| - 1) \\ & + \lambda_2 CT(s_1, s_2) \end{aligned} \quad (11)$$

IV. 실험 결과

본 논문의 모든 실험은 동일한 조건에서 WGAN-GP, WGAN-CT와 제안모델을 비교하기 위하여 WGAN-GP와 기본적으로 동일한 구조와 파라미터를 이용하며 λ_1 과 λ_2 는 WGAN-CT에 따라 10, 2로 각각 지정한다. 제안 알고리즘의 성능을 확인하기 위하여 학습데이터로 MNIST[9] 숫자 이미지와 CIFAR-10[10]을 이용한다. WGAN-GP[6]과 WGAN-CT[7] 알고리즘을 학습시키는 데 있어서 생성기와 분별기의 구조, 학습률 등의 매개변수(hyperparameter) 등은 해당 논문에서 제시한 최적의 결과를 그대로 채택한다. 제안 알고리즘은 립스즈 연속을 보장하기 위한 정칙항 외에는 기본적으로 WGAN-GP와 동일하므로 이의 구조와 매개변수 등을 공유한다. 각 모델의 학습 반복횟수는 50,000으로 고정한다.

1. MNIST 실험결과

MNIST 데이터 집합은 70,000개의 필기체 숫자 이미지로 구성되어 있고 그 중에서 10,000개의 이미지는 테스트 용으로 제공된다. 본 논문에서 제안한 알고리즘과 WGAN-GP, WGAN-CT를 비교하기 위하여 1,000 개의 이미지를 테스트 이미지로 이용한다.



(a) WGAN-GP



(b) WGAN-CT



(c) 제안모델

그림 1. MNIST에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 생성 이미지

(a) WGAN-GP (b) WGAN-CT (c) 제안모델
 Fig. 1. The generated images of WGAN-GP, WGAN-CT and our proposed model at MNIST dataset.
 (a) WGAN-GP (b) WGAN-CT (c) Proposed model

그림 1은 MNIST 데이터집합에 대하여 50,000번의 학습과정 후에 WGAN-GP, WGAN-CT 및 제안 모델의 생성기가 출력시킨 이미지를 제시한다. 화질의 차이는 크지 않지만 제안모델의 출력이 보다 자연스럽게 배경과 전경 간의 컨트라스트가 더 강함을 확인할 수 있다. 그림 2는 식 (7), (9), (11)에 나타난 손실함수의 학습회수에 따른 크기를 그래프로 보여준다.

변화의 추이를 비교, 확인하기 위하여 5,000회의 학습회수 결과만 발췌한다. 세 모델의 손실함수는 서로 다르지만 그 차이는 상대적으로 크지 않으므로 이들을 학습시키는데 있어서의 수렴과정을 쉽게 파악할 수 있다. 이 그래프를 보면, 제안모델이 가장 빠르고 WGAN-CT, WGAN-GP의 순으로 수렴하는 것을 볼 수 있다.

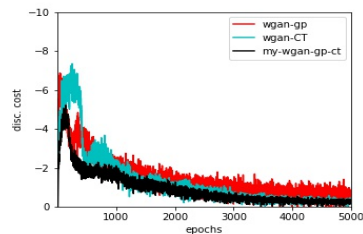


그림 2. MNIST에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 학습 손실함수

Fig. 2. Convergence curves of the discriminator cost of WGAN-GP, WGAN-CT and our proposed model when MNIST dataset is used.

그림 3은 학습데이터에 속하지 않은 1,000개의 랜덤으로 추출한 이미지에 대한 손실함수(validation loss)를 그래프로 보여준다.

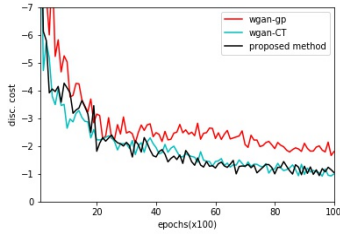


그림 3. MNIST에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 validation 손실함수 변화
 Fig. 3. The validation losses of WGAN, WGAN-GP and our proposed model when CIFAR-10 dataset is used.

그림 4는 각 모델에 대하여 학습회수에 따른 분별기의 립슈츠 연속 확률을 그래프로 나타낸 것이다.

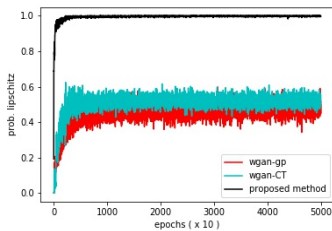
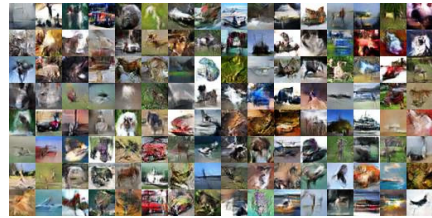


그림 4. MNIST에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 립슈츠 연속 확률
 Fig. 4. The probabilities of Lipschitz continuity of WGAN, WGAN-GP and our proposed model when MNIST dataset is used.

WGAN-GP와 WGAN-CT는 분별기 f 의 미분크기를 1로 유지하도록 학습함으로써 립슈츠 연속일 확률이 그리 높지 않으나 제안 모델은 학습초기부터 1에 가까운 확률로 립슈츠 연속을 유지하고 있다. 그림 4에 주어진 립슈츠 연속확률의 차이와 비교해보면 제안 모델이 보다 안정적으로 수렴하고 있다고 볼 수 있다. 이러한 특성을 명확히 확인하기 위하여 보다 스펙트럼이 복잡한 학습이미지를 이용하여 동일한 실험을 다음과 같이 수행한다.

2. CIFAR-10 실험결과

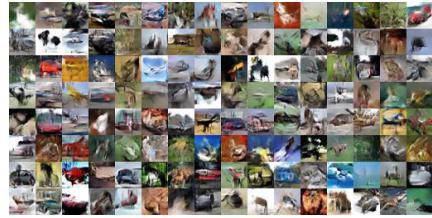
MNIST 보다는 스펙트럼이 복잡한 CIFAR-10 이미지 집합을 이용하여 위와 동일한 실험을 수행한다.



(a) WGAN-GP



(b) WGAN-CT



(c) 제안모델

그림 5. CIFAR-10에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 생성 이미지

(a) WGAN-GP (b) WGAN-CT (c) 제안모델
 Fig. 5. The generated images of WGAN-GP, WGAN-CT and our proposed model at CIFAR-10 dataset.
 (a) WGAN-GP (b) WGAN-CT (c) Proposed model

그림 5는 CIFAR-10 데이터집합에 대하여 50,000번의 학습과정 후에 WGAN-GP, WGAN-CT 및 제안 모델의 생성기가 출력시킨 이미지를 제시한다. 세 모델 모두 안정적으로 수렴하고 있고, 각 생성기가 발생시킨 이미지의 화질은 그 차이가 눈에 띌 정도로 크지는 않으나 제안 알고리즘이 생성한 이미지가 보다 선명하다고 볼 수 있다. 그림 6은 각 모델에 대한 분별기의 립슈츠 연속 확률을 보여준다. WGAN-GP는 여전히 매우 낮은 립슈츠 연속확률을 갖고 있고, WGAN-CT는 MNIST의 실험 결과와는 달리 1에 가까운 립슈츠 연속확률을 유지하고 있어서 제안모델의 결과와 거의 겹쳐 있다. 반면에 제안 모델은 MNIST의 실험결과와 거의 동일한 확률로 립슈츠 연속을 유지한다.

실험결과를 바탕으로 분석해보면 WGAN-GP는 안정적인 수렴이 보장되지 않고 WGAN-CT는 학습 이미지에 따라 립슈츠 연속을 유지하는 것이 쉽지 않음을 확인할 수 있다.

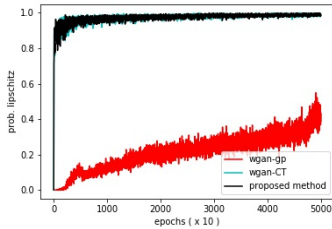


그림 6. CIFAR-10에 대한 WGAN-GP, WGAN-CT 및 제안 모델의 립쉬츠 연속 확률

Fig. 6. The probabilities of Lipschitz continuity of WGAN-GP, WGAN-CT and our proposed model when CIFAR-10 dataset is used.

생성된 화질의 객관적 지표로 그림 7에 제시한 바와 같은 인셉션 점수(inception score)[11]가 널리 이용된다. 이에 의하면 제안 모델과 WGAN-CT가 거의 일치하는 결과를 보이고 WGAN-GP 다소 느리게 따라가고 있어 그림 5와 그림 6을 통한 분석한 결과와 일치한다고 볼 수 있다.

이상의 실험결과로부터 제안 알고리즘은 WGAN-GP와 WGAN-CT보다 학습의 수렴속도와 생성기가 발생시킨 이미지의 화질 등의 관점에서 성능이 향상됨을 확인할 수 있다. 이를 토대로 분석해보면 모델의 분별기 함수가 립쉬츠 연속을 보다 높은 확률로 유지할수록 안정성이 증가되어 수렴속도와 더불어 출력 이미지의 화질향상을 얻을 수 있을 것으로 기대된다.

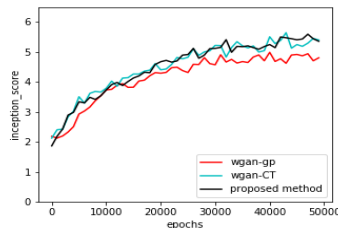


그림 7. CIFAR-10에 대한 WGAN, WGAN-GP 및 제안 모델의 인셉션 점수

Fig. 7. The inception scores of WGAN, WGAN-GP and our proposed model when CIFAR-10 dataset is used.

V. 결 론

WGAN의 수렴과 성능을 향상시키기 위해서는 분별기 함수의 립쉬츠 연속을 안정적으로 보장하는 것이 필수적인 것으로 알려져 있는데, 본 논문에서 제안한 모델은 이

를 적절히 만족시키고 있음을 다양한 실험으로 확인하였다. 분별기 함수의 미분크기를 1로 유지하기 위한 정칙화 항을 추가하는 기법이 WGAN의 개선책으로 널리 활용되고 있으나 이 방법으로는 안정적인 립쉬츠 연속을 보장할 수 없다. 반면에 제안 알고리즘은 학습데이터에 거의 영향을 받지 않고 립쉬츠 연속을 매우 안정적으로 유지하여 수렴속도가 향상됨을 실험으로 확인하였다.

향후에는 학습데이터의 분포를 보다 조밀하게 표본화하는 기법을 개선하여 립쉬츠 연속을 보다 안정적으로 유지하기 위한 연구를 계속 진행할 계획이다.

References

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," In *Advances in Neural Information Processing*, 2014.
- [2] D.P. Kingma and M. Welling, "Auto-encoding variational bayes," In *Proceedings of the International Conference on Learning Representations(ICLR)*, 2014.
- [3] A. Radford, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv:1511.06434v2 [cs.LG], Jan. 2016.
- [4] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," arXiv:1701.07875v3 [stat.ML], Dec. 2017.
- [5] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," arXiv:1701.04862v1 [stat, ML], Jan. 2017.
- [6] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. Courville, "Improved training of Wasserstein gans," arXiv:1704.00028v3 [cs.LG], Dec. 2017.
- [7] X. Wei, B. Gong, Z. Liu, W. Lu and L. Wang, "Improving the improved training of Wasserstein gans," arXiv:1803.01541v1 [cs.CV], Mar. 2018.
- [8] C. Villani, *Optimal Transport: old and new*, Vol. 338, Springer Science & Business Media, 2008.
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2234, 1998.
DOI: <http://dx.doi.org/10.1109/5.726791>
- [10] A. Krizhevsky, "Learning multiple layers of features from tiny images," Technical Report TR-2009, University of Toronto, 2009.
- [11] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gan," In *Advances in Neural Information Processing Systems*, pp. 2226-2234, 2016.

- [12] S. Lee, H. Kim, H. Seok, and J. Nang, "Comparison of fine-tuned convolutional neural networks for clipart style classification," International Journal of Internet, Broadcasting and Communication, Vol. 9, No. 4, pp. 1-7, 2017.
DOI: <https://doi.org/10.7236/IJIBC.2017.9.4.1>.
- [13] B.S. Kim and I.H. Lee, "Retinal blood vessel segmentation using deep learning," Journal of KIIT, Vol. 17, No. 5, pp. 77-82, 2019.
DOI:<http://dx.doi.org/10.14801/jkiit.2019.17.5.77>
- [14] C.I. Woo and E.H. Goo, "A study on integrity verification and tamper detection of digital image," Journal of the Korea Academia-Industrial Cooperation Society, Vol. 20, No. 10, pp. 203-208, 2019. D
OI: <https://doi.org/10.5762/KAIS.2019.20.10.203>

저 자 소 개

한희일(정회원)



- 1984년 서울대학교 제어계측공학과 학사 졸업.
 - 1988년 서울대학교 제어계측공학과 석사 졸업.
 - 1995년 University of Arizona 전기 및컴퓨터공학과 박사 졸업.
 - 2019년 현재 한국외국어대학교 정보통신공학과 교수.
- 주관심분야 : 신호처리, 컴퓨터비전, 머신러닝, 기하학 및 토폴로지

※ 본 논문은 2019년도 한국외국어대학교 교내 학술연구지원에 의하여 연구되었음