GAN 모델에서 손실함수 분석

이초연*1), 박지수**, 손진곤*2)
*한국방송통신대학교 대학원 정보과학과
**동국대학교 융합교육원
e-mail:qecche77@knou.ac.kr

A Study on the Loss Functions of GAN Models

Cho-Youn Lee*, JiSu Park**, Jin Gon Shon*
*Dept. of Computer Science, Graduate School, Korea National Open University
**Convergence Institute, Dongguk University

요 약

현재 딥러닝은 컴퓨터 분야에서 이미지 처리 방법으로 활용도가 높아지면서 딥러닝 모델 개발 연구가 활발히 진행되고 있다. 딥러닝 모델 중에서 이미지 생성모델은 대표적으로 GAN(Generative Adversarial Network, 생성적 적대 신경망) 모델을 활용하고 있다. GAN은 생성기 네트워크와 판별기 네트워크를 이용하여 진짜 같은 이미지를 생성한다. 생성된 이미지는 실제 이미지와의 오차를 최소화해야 하며 이때 사용하는 함수를 손실함수라고 한다. GAN에서 손실함수는 이미지를 생성하는 학습이 불안정하여 이미지 품질이 떨어지는 문제가 있다. 개선된 GAN 관련 연구가 진행되고 있지만 완전한 문제 해결에는 부족하다. 본 논문은 7개의 GAN 모델에서 사용하는 손실함수를 분류하고 특징을 분석한다.

1. 서론

현재 4차 산업 확산으로 인공지능에 대한 관심이 커지면서 딥러닝(deep learning) 모델에 관한 연구가 광범위하게 진행되고 있다. 딥러닝은 컴퓨터 비전, 음성 인식, 게임 및 멀티모달과 같은 다양한 인공지능에 활용되고 있다. 그러나 딥러닝을 학습하는 것은 복잡하며 다양한 샘플과 수천 개의 카테고리가 있는 대규모 데이터 세트에서 매개변수를 정교하게 조정해야 한다[1]. 딥러닝은 데이터를 스스로 학습하는 신경망(neural network)을 이용하여 패턴인식에 대한 빅데이터의 문제를 해결할 수 있다. 이미지 인식이나 분류, 예측을 위한 DNN(심층신경망), CNN(합성곱 신경망), RNN(순환 신경망) 등이 있으며, 생성모델은 GAN(생성적 적대 신경망)이 대표적이다[2].

CNN은 컨볼루션(convolutional) 필터와 차원을 줄이면서 특징을 선택 출력하는 풀링(pooling) 방법으로 이미지 인식의 정확도가 높아 많이 사용하는 모델이다. RNN은 시계열 데이터를 분석하고 미래의 값을 예측하는 방법으로 사용된다[3]. GAN은 많은 난해한 확률 계산을 근사화하는 어려움을 회피하기 위해서 생성모델을 훈련하기 위한 대체 프레임워크로 도입되었다. GAN은 시그모이드 크로스 엔트로피 손실함수(sigmoid cross entropy loss function)를 사용하기 때문에 학습이 불안정하고 그라디언트(vanishing gradient)가 사라진다. 수렴을 보장하지만 최소화와 최대화의 과정에서 이론적 가정이 깨지면서 불안정한 결과가 출력된다[4][5].

GAN은 비지도 학습이며 생성기(generative) 네트워크와 판별기(discriminator) 네트워크를 사용하여 미니맥스2인 게임(two-player minimax game)으로 진짜 같은 이미지를 생성한다. 생성 이미지와 실제 이미지와의 오차를 최소화하기 위한 손실함수(loss function)는 모델 성능의 최적화를 위하여 매우 중요하다[5].

본 논문은 GAN 계열 신경망에서 모델별로 개선된 손실함수들을 분류하고 특징을 분석한다.

2. 손실함수

손실함수는 신경망에서 학습하여 얻은 예측 값(y)과 정답 레이블(t)과의 오차를 나타내는 함수를 나타낸다. 예측 값은 $y=wx+b$이며 w는 가중치, b는 편향이다. 손실함수의 값을 최소화하기 위해서 최적화된 매개변수 w와 b를 구해야 하며 경사하강법(gradient descent)을 사용한다. 손실함수는 일반적으로 평균제곱오차(mean squared error)와 크로스 엔트로피 오차(cross entropy error)를 사용한다. 식은 (1)(2) 와 같다.

$$E=\frac{1}{2}\sum(y_k-t_k)^2 \quad (1)$$

$$E=-\sum t_k \log y_k \quad (2)$$

식(1)에서 k는 데이터의 차원 수이며, 예측 값(y)과 정답 레이블(t)과의 오차를 계산하고 마이너스를 피하기 위해 제곱하고 그 합을 더한다. 식(2)는 확률 로그함수를 사

1) 한국방송통신대학교 대학원 재학생
2) 교신저자

용하며 log는 밑이 e인 자연로그이며, t에서 1에 해당하는 y값을 대입하여 크로스 엔트로피 오차를 계산한다.

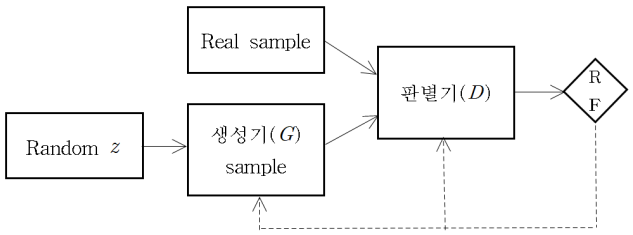
3. GAN 모델에서 손실함수 분석

3.1 GAN

GAN은 적대적 과정을 통해 생성모델을 평가하기 위한 프레임워크이다. 손실함수는 식(3)과 같다.

$$\min_G \max_D V(D,G) = E_{x\sim pdata(x)}[\log D(x)] + E_{z\sim pz(z)}[\log(1-D(G(z)))] \quad (3)$$

식(3)에서 GAN의 목적은 랜덤 데이터 $pz(z)$를 실제 데이터 $pdata(x)$의 분포와 유사하게 만드는 것이다. 균일 분포 [-1,1]로부터 샘플링된 z 벡터를 이용하여 이미지를 생성한다. V는 가치함수, D는 판별기, G는 생성기, E는 기댓값, x는 실제 데이터의 샘플 이미지, $D(x)$는 실제 이미지로 판단할 확률, $G(z)$는 생성기에서 얻은 샘플 이미지, $D(G(z))$는 생성기가 생성한 이미지로 판단할 확률이다. $\log(1-D(G(z)))$를 최소화한다. 따라서 생성기 입장에서는 판별기가 real이라고 판단할 수 있도록 최소의 손실함수 값이어야 한다. 손실함수의 전역최소값(global minima)을 찾기 위해 연속적인 그라디언트를 사용한다. GAN은 시그모이드 크로스 엔트로피 손실함수를 사용하여 학습이 불안정하고 이미지 품질이 떨어지는 단점이 있다. 시그모이드는 활성화 함수이며 출력값이 0에서 1사이이며 그래프가 S자 모양의 곡선이다[5]. GAN의 구조는 (그림 1)과 같다.



(그림 1) GAN의 구조

3.2 CGAN(Conditional GAN)

조건부 CGAN은 생성되는 데이터를 제어할 수 있는 모델이다. 스케치 기반 얼굴 생성과 이미지 자동태그 등에 응용된다. 손실함수는 식(4)과 같다.

$$\min_G \max_D V(D,G) = E_{x\sim pdata(x)}[\log D(x|y)] + E_{z\sim pz(z)}[\log(1-D(G(z|y)))] \quad (4)$$

기존 GAN의 생성기(G)와 판별기(D) 모두의 입력레이어에 조건으로 원하는 데이터 y를 입력하여 구성할 수 있다. y는 클래스 레이블이나 다른 데이터와 같은 보조 정보를 나타낸다. 생성기에서 입력 노이즈 $pz(z)$와 y는 결합되고, 판별기에서 x와 y는 입력 및 차별적 함수로 표시된다. CGAN은 이미지에 있는 객체를 자연어로 묘사할 수 있다. 이미지 특징에 조건부 tag-vectors의 분포를 생성한다. 조건 입력으로 인한 과적합에 주의해야 한다[6].

3.3 WGAN(Wasserstein GAN)

WGAN은 Earth-Mover(EM)거리의 근사를 최소화하는 방법을 이용하여 wasserstein 거리의 하한(inf)을 적용한 모델이다. 기존 GAN보다 학습 안정성 향상과 생성기와 판별기(WGAN에서는 비평가(critic)라고 함)의 훈련에 있어서 균형을 유지할 필요가 없다. 모드붕괴현상(mode dropping phenomenon)이 감소한다. WGAN은 판별기를 최적으로 훈련하여 이동하는 양의 EM 거리를 지속적으로 추정 할 수 있다. 손실함수는 식(5) 와 같다.

$$W(P_r,P_g) = \inf_{\gamma\in \Pi(P_r,P_g)} E_{(x,y)\sim\gamma}[\| x-y \|] \quad (5)$$

inf(infimum)는 최대하한, 어떤 실수의 부분집합의 가장 작은 경계값, 기댓값은 $\| x-y \|$의 최소값, $\Pi(P_r,P_g)$는 모든 결합분포 $\gamma(x,y)$의 집합, $\gamma(x,y)$는 실제 데이터 분포 P_r를 생성기의 분포 P_g로 변환하기 위해 x에서 y로 이동되어야하는 mass의 양을 나타낸다. Adam (1> 0)과 같은 옵티마이저를 사용하거나 높은 학습 속도를 사용할 때 WGAN 학습이 불안정 해지는 경우가 있다. 가중치 클리핑으로 인해 샘플생성 오작동의 단점이 있다[7].

3.4 WGAN-GP

WGAN에서 가중치 클리핑(clipping weighting) 때문에 샘플을 생성할 때 오작동을 한다. 이때 그라디언트가 사라지거나 폭발하기 때문에 대안으로 비평가(critic)에서 그라디언트의 놈(norm, 이동거리)에 패널티(penalty)를 적용한다. WGAN-GP의 손실함수는 식(6)과 같다.

$$L = \underset{\tilde{x}\sim P_g}{E}[D(\tilde{x})] - \underset{x\sim P_r}{E}[D(x)] + \lambda \underset{\hat{x}\sim P_{\hat{x}}}{E}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2] \quad (6)$$

판별기(D)의 목적 함수에 패널티를 추가, 1에서 멀어지면 패널티 λ를 적용한다. 랜덤샘플 $\hat{x}\sim P_{\hat{x}}$에 대한 그라디언트 norm에 패널티가 있는 제약 조건이다[8].

3.5 EBGAN(Energy-based GAN)

에너지 기반 EBGAN은 총 변동 거리에 대한 생성 접근법으로 판별기(D)를 에너지 함수로 본다. 판별기를 에너지 함수로 보는 것은 로지스틱 출력과 함께 일반적인 이진 분류기외에도 오토인코더(auto-encoder) 아키텍처를 사용하고 손실함수를 사용할 수 있게 한다. GAN보다 안정적이고 단일 스케일 아키텍처(single-scale architecture)가

고해상도 이미지(high-resolution images)를 생성하도록 한다. 손실함수는 식(7)(8) 과 같다.

$$L_D(x,z) = D(x) + [m - D(G(z))]^+ \quad (7)$$

$$L_G(z) = D(G(z)) \quad (8)$$

주어진 양의 마진 m을 감안할 때, 데이터 샘플 x 및 생성된 샘플 $G(z)$, 판별기 손실(L_D)과 생성기 손실(L_G)에 대한 정의, 여기서 $[\bullet]^+ = \max(0, \bullet)$. G의 매개변수와 관련하여 L_G를 최소화하는 것은 L_D의 두 번째 항을 최대화하는 것과 유사하다. $D(G(z)) \geq m$ 일 때와 동일하지만 0이 아닌 그라디언트가 있다. D의 최적화가 어렵고 불완전한 그라디언트로 제한된다[9].

3.6 BEGAN(Boundary Equilibrium GAN)

경계평형 BEGAN은 오토인코더 손실을 도입한 다음 생성된 이미지와 실제 이미지의 오토인코더 손실 분포 사이의 wasserstein 거리에 대한 하한 값을 최적화하는 것이 목적이다. 판별기(D)에 오토인코더를 사용하고 경계평형 방법을 제안하여 훈련 중에 생성기(G)와 판별기(D)의 균형을 유지한다. 손실함수는 식(9)과 같다.

$$L(v) = |v - D(v)|^{\eta} \; where \begin{cases} D: R^{N_x} \mapsto R^{N_x} & is\ the\ autoencoder\ function. \\ \eta \in 1, 2 & is\ the\ target\ norm. \\ v \in R^{N_x} & is\ a\ sample\ of\ \mathrm{dimension}\ N_x. \end{cases}$$

$$W_1(\mu 1, \mu 2) = \inf_{\gamma \in \Gamma(\mu 1, \mu 2)} E_{(x1,x2)\sim\gamma}[\,\| x1 - x2 \|\,] \quad (9)$$

여기서 $\mu 1, \mu 2$는 생성된 이미지와 실제 이미지의 오토인코더 손실 분포이다.

$D: R^{N_x} \mapsto R^{N_x}$는 오토인코더로 설계된 컨볼루션 심층신경망이다. $N_x = H \times W \times C$ 는 x의 크기에 대한 약식이며 H, W, C 는 높이, 너비, 색상이다[10].

3.7 LSGAN(Least Squares GAN)

LSGAN은 판별기에 최소제곱 손실함수(least squares loss function)를 사용한다. 가짜샘플에 패널티를 적용하여 최소제곱 손실함수의 결정 경계 방향으로 샘플을 생성한다. 손실함수는 식(10)과 같다.

$$\min_D V_{LSGAN}(D) = \frac{1}{2} E_{x \sim pdata(x)}[(D(x) - b)^2] + \frac{1}{2} E_{z \sim pz(z)}[(D(G(z)) - a)^2] \quad (10)$$

$$\min_G V_{LSGAN}(G) = \frac{1}{2} E_{z \sim pz(z)}[(D(G(z)) - c)^2]$$

LSGAN의 손실함수는 a는 가짜 데이터, b는 진짜 데이터, c는 생성기(G) 입장에서 판별기(D)가 c를 가짜 데이터라고 믿는 값, 이진코드 a=0, b=c=1 이다. 기존 GAN과는 달리 결정경계(decision boundary)에서 멀리 떨어진 샘플에 패널티를 준다. G가 이미지를 생성할 때 결정경계에 최대한 가까운, 즉 실제 이미지에 가깝게 생성하도록 한다. JS(Jensen-Shannon)거리는 모든 분포의 거리를 효과적으로 측정해주지 못한다. 생성된 샘플을 결정경계 쪽으로 끌어오는 대신 생성된 샘플을 실제 데이터로 직접 가져오는 연구가 필요하다[11]. GAN 모델의 손실함수 비교 분석은 <표 1>과 같다.

<표 1> GAN 모델에서 손실함수 비교 분석

	특 징	장 점	단 점
GAN(식3)[5]	• 크로스엔트로피 손실함수 사용 • 시그모이드 활성화 함수 사용	• 과적합(overfitting) 되지 않음 • 마르코프 체인 불필요	• 학습 불안정 • 지역 최소값에 빠짐 • 그라디언트 소실(vanishing gradients)
CGAN(식4)[6]	• 생성기(G)와 판별기(D)에 원하는 조건 y를 사용	• 생성되는 데이터 조절가능	• D에서 MLP로 다시 구현됨 • 과적합 됨 • 하이퍼파라미터 공간과 아키텍처에 대한 추가 탐구 필요
WGAN(식5)[7]	• 이동하는 양의 EM거리 이용 • 실수에서 부분집합의 가장 작은 경계값 inf(최대 하한) 사용 • 판별기(D)를 비평가(critic)라 함	• GAN 보다 학습안정 향상 • 그라디언트 소실감소 • 생성기(G)와 판별기(D)의 평형유지 불필요	• 높은 학습속도는 학습 불안정해짐 • 가중치클리핑에 의한 샘플생성 오작동
WGAN-GP (식6)[8]	• 비평가 손실에서 그라디언트에 패널티 적용	• WGAN보다 학습속도 향상 • 샘플 품질 향상 • 가중치 클리핑을 패널티로 해결	• 비평가에서 과적합 됨
EBGAN(식7,8)[9]	• 판별기(D)를 에너지 함수로 봄 오토인코더 적용	• GAN보다 학습안정 향상 • 고해상도 이미지 생성	• 판별기(D)의 최적화가 어렵다 • 불완전한 그라디언트 때문에 모드붕괴
BEGAN(식9)[10]	• 판별기(D)에 오토인코더 적용	• 학습안정 향상 • 생성기(G)와 판별기(D) 균형유지	• 오토인코더로 인한 압축 데이터 손실
LSGAN(식10)[11]	• 판별기(D)에 최소제곱 손실함수 사용 • 가짜 샘플에 패널티 적용	• GAN보다 학습안정 향상 • 고해상도 이미지 생성	• JS거리는 모든 분포의 거리를 효과적으로 측정해주지 못함 • 과적합 됨

4. 결론 및 향후 연구

본 논문은 GAN 모델의 이미지 처리에서 손실함수(loss function)의 필요성을 설명하고, 7개의 GAN 모델에서 사용하는 손실함수에 대하여 개선된 방법별로 분류하고 특징을 분석하였다. 손실함수는 생성된 이미지와 실제 이미지와의 오차를 최소화하기 위한 중요한 지표이다. 본 논문은 GAN의 이미지 분류에 사용하는 시그모이드 크로스 엔트로피 손실함수에 대한 단점을 개선한 CGAN, WGAN, WGAN-GP, EBGAN, BEGAN, LSGAN를 소개하고 모델별로 손실함수에 대하여 비교분석하였다. 비교분석한 결과 모든 판별기에 손실함수의 개선된 방법을 적용하고 있다. 손실함수에 사용하는 방법별로 분류하면 다음 <표 2>와 같다.

<표 2> 손실함수에 따른 방법별 분류

손실함수 분류	GAN 모델
조건부 적용	CGAN
inf(최대하한) 적용	WGAN, WGAN-GP, BEGAN
오토인코더 적용	EBGAN, BEGAN
패널티 적용	WGAN-GP, LSGAN

기존 GAN 모델은 최소화와 최대화의 과정에서 학습 불안정과 수렴, 시각적 품질 저하 등에 대한 문제가 발생하면서 개선된 GAN 모델에 관한 연구가 활발해졌다.

분석한 모델들의 특성을 갖는 이유는 CGAN에서 조건 입력 y는 클래스 레이블이나 보조 정보를 나타내기 때문에 생성되는 데이터 조절이 가능하다. inf는 어떤 실수의 부분집합의 가장 작은 경계값을 나타내기 때문에 inf를 적용하는 모델은 기존 GAN보다 학습이 안정되었다. 오토인코더는 일반적인 GAN의 트릭을 피하기 위해 사용한다. 인코더와 디코더의 과정으로 입력값을 출력값으로 매칭할 수 있기 때문에 GAN보다 학습이 안정되었지만 판별기의 최적화는 어렵다는 특성이 있다. 패널티를 적용한 모델들은 그라디언트의 norm과 결정경계에서 멀어진 샘플에 패널티를 적용하여 실제 이미지에 가깝게 이미지를 생성할 수 있도록 하기 때문에 이미지 품질이 향상되었다.

GAN 모델은 CCTV, 자연어처리, 이미지 디자인, 장애물에 가려진 이미지 재생성에 사용된다. GAN 모델을 활용한 CCTV에서 범죄자의 이미지를 생성한다면 결과에 대한 이유가 명확해야 한다. 인공지능 기반 의사결정이 잘못되지 않았음을 보장하기 위해서 GAN 모델의 블랙박스는 투명해야하며 수식과 설명이 필요하다. 따라서 손실함수들에 대한 분석이 도움이 될 것으로 기대한다. 향후 연구에는 개선된 다양한 GAN 계열 신경망과 분석한 손실함수들을 바탕으로 생성된 이미지와 실제 이미지와의 매칭에 에지 검출을 이용한 GAN 모델을 연구할 필요가 있다.

참고문헌

[1] S Wu, G Li, L.Deng, Liu Liu, Dong Wu, Yuan Xie, Luping Shi, “L1-Norm Batch Normalization for Efficient Training of Deep Neural Networks”, IEEE, pp. 2043-2051, 2018

[2] 김윤진, “딥러닝(Deep Learning)을 활용한 이미지 빅데이터(Big Data) 분석 연구”, 박사학위논문, 중앙대학교 대학원, 통계학과 통계학전공, 2017년 2월

[3] 유병인, 황원준, 한승주, 이선민, 김정배, 한재준, “인간 수준에 근접한 딥러닝 기반 영상 인식의 동향”, 한국정보과학회, 33권, 9호, 32-41쪽, 2015년 8월

[4] 최영주, “사실적인 얼굴 영상 생성을 위한 딥러닝 연구”, 서강 대학교 영상대학원 미디어공학과, 박사학위논문, 2017. 6

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozairy, Aaron Courville, Yoshua Bengioz, “Generative Adversarial Nets”, Advances in neural information processing systems, pp. 2672-2680, 2014

[6] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *Comput. Sci.*, pp. 2672-2680, Nov. 2014.

[7] M. Arjovsky, S. Chintala, and L. Bottou. “Wasserstein gan”, *arXiv:1701.07875*, 2017

[8] I.Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs", in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5767-5777. [Online]. Available:https://arxiv.org/pdf/1606.0349

[9] J. Zhao, M. Mathieu, and Y. Lecun, “Energy-based generative adversarial network,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Sep. 2016. [Online]. Available: https://arxiv.org/pdf/1609.03126

[10] D. Berthelot, T. Schumm, and L. Metz. “BEGAN: Boundary Equilibrium Generative Adversarial Networks”, arXiv preprint, arXiv:1703.10717, 2017.

[11] X. Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, Stephen Paul Smolley, “Least Squares Generative Adversarial Networks”, IEEE, pp. 2813-2821, 2017