

Latent vector 분포 조정을 활용한 DCGAN 기반 이모지 생성 기법

송윤경¹, 하유진¹, 성아영², 김건우³
¹경상국립대학교 AI 융합공학과 석사과정
²경상국립대학교 컴퓨터과학부 학부생
³경상국립대학교 컴퓨터과학부 교수

Songyg1020@gnu.ac.kr, nrt5077@gnu.ac.kr, achieve00@gnu.ac.kr, gunwoo.kim@gnu.ac.kr

DCGAN-based Emoji Generation exploiting Adjustment of Latent vector Representation

Yun-Gyeong Song¹, Yu-Jin Ha¹, A-Yeong Seong², Gun-Woo Kim²
¹Dept. of AI Convergence Engineering, Gyeongsang National University
²Dept. of Computer Science & Engineering, Gyeongsang National University

요 약

최근 SNS의 발달로 인해 자신의 감정을 빠르고 효과적으로 전달할 수 있는 이모지의 중요성이 커지고 있다. 하지만 이모지를 수동으로 생성하기 위해서 시간과 비용이 많이 들고 자신의 감정에 맞는 이모지를 찾아야 하며 해당 이모지가 없을 수 있다. 기존 DCGAN을 활용한 이모지 자동 생성 연구에서는 부족한 데이터셋으로 인해 G(Generator)와 D(Discriminator)가 동등하게 학습하지 못해서 두 모델 간 성능 차이가 발생한다. D가 G보다 최적해에 빠르게 수렴하여 G가 학습이 되지 않아 낮은 품질의 이모지를 생성하는 불안정 문제가 발생한다. 이 문제를 해결하기 위해 본 논문에서는 Latent vector 분포를 데이터셋에 맞게 조정하여 적은 데이터로 G에서 안정적으로 학습할 수 있게 하는 G 구조와 다양한 이모지 생성을 위한 Latent vector 평균 조정 기법을 제안한다. 비교 실험 결과 불안정 문제를 개선하였고 FID와 IS 수치를 통해 성능 개선 효과를 검증했다.

1. 서론

스마트폰의 보급 이후 문자를 의사소통의 수단으로 하는 SNS (Social Network Service)가 개발되었고 문자를 통한 의사소통의 한계를 보완하기 위해 이모지가 발달하였다. 이모지는 문자를 통한 의사소통보다 감정 전달을 효과적으로 할 수 있으며 신속한 메시지 전달이 가능하다. 이러한 장점들로 이모지의 중요성이 증가하고 있다.[1]

현재 존재하는 이모지는 사람의 특정 감정에 대해 세분되지 않아 자신의 감정에 맞는 이모지를 찾기 힘들다. 그러므로 수많은 이모지 중에 사용자의 감정에 가장 유사한 이모지를 찾고 이후 비슷한 상황에 사용자의 감정에 더 유사한 다른 이모지가 있어도 새로운 이모지를 찾기보다는 이전에 사용하였던 이모지를 사용한다. 이러한 문제 때문에 다양한 이모지 생성 연구가 필요하다.

GAN(Generative Adversarial Network) [2]는 데이터 생성 모델이며 G(Generator)와 D(Discriminator) 2개의 모델이 적대적으로 학습하여 데이터를 생성한다. D는 실제 데이터인지 가짜 데이터인지 판별하고 G는 D loss를 통해 실

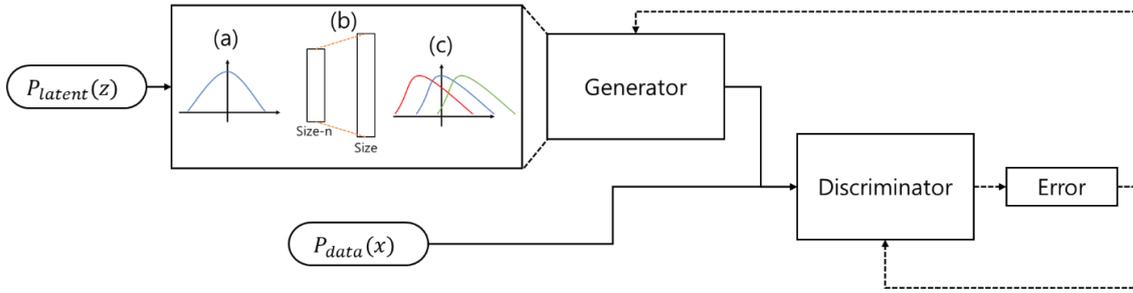
제와 유사한 데이터를 생성한다. GAN을 학습하기 위해서는 많은 데이터가 필요하며 데이터가 부족하면 D와 G 사이에 경쟁 관계에 이루어지지 않아 비교적 학습이 쉬운 D의 성능이 G보다 압도적으로 향상되어 G가 학습이 안 되는 불안정 문제가 심화된다.

본 논문에서는 적은 데이터셋에서 G가 실제 데이터의 분포를 빠르게 학습하는 Generator 구조와 다양한 이모지 생성을 위해 -1~1 사이의 균일분포에서 평균을 샘플링하여 정규분포를 생성하는 Latent vector 분포 조정 기법을 제안한다.

2. 관련 연구

2.1 DCGAN

DCGAN[3]은 GAN에 Convolution 구조를 결합한 것으로 GAN보다 높은 성능과 안정적인 학습을 보여준다. 하지만 DCGAN을 이용하여 이모지를 생성하면 적은 데이터셋으로 인해 G와 D가 경쟁 관계를 맺지 못하고 불안정 문제가 발생한다.



(그림 1) 제안된 Generator 구조

2.2 DeLiGAN

DeLiGAN[4]은 적은 데이터에서 다양한 데이터 생성을 위해 연구된 GAN 기반의 모델이다. Gaussian mixture model을 통해 다양한 컴포넌트를 만들고 무작위로 샘플링을 통해 Latent vector를 구한다. Latent의 평균과 분포를 변경하는 것으로 데이터의 다양성을 증가시킬 수 있고 선명한 데이터를 생성할 수 있음을 증명한다. 하지만 이 방법은 컴포넌트의 평균과 분산을 학습시켜야 하므로 학습 시간이 증가하고 정규분포에서만 샘플링을 하기 때문에 초기에 불안정 문제가 발생할 수 있다.

3. 방법

3.1. Latent vector 분포 조정 구조

본 연구에서는 정규분포를 따르는 Latent vector z 를 실제 데이터의 특징 확률 분포 일부가 포함되어도록 조정하는 Generator 구조를 제안한다.

$$(1)$$

(그림 1)(b)와 같이 size- n 의 크기를 가지는 Latent vector를 식(1)과 같이 size 크기를 가지는 layer로 조정하면 Latent vector에 실제 데이터의 공통된 특징 분포 일부분을 설명할 수 있는 분포를 포함할 수 있다.

N 은 하이퍼파라미터로 실제 데이터의 특징 분포를 얼마나 반영할지를 결정한다. N 은 1부터 1씩 더해 epoch 30이 넘어가도 불안정 문제가 발생하지 않는 N 을 구하여 사용한다. 불안정 문제가 발생하지 않는 시점부터 안정성을 올리기 위해 N 에 높은 숫자를 부여하면 실제 데이터의 공통된 부분이 아닌 세부적인 요소까지 반영하게 되며 다양한 데이터가 생성되지 않을 수 있다.

3.2. latent vector 평균 조정 기법

Latent 분포 조정 구조를 통해 확장된 latent vector에 -1과 1 사이의 균등 분포에서 샘플링된 평균을 더 하여 다양한 데이터를 생성하도록 하는 latent vector 평균 조정 기법을 제안한다.

$$z \sim p(z) \quad p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \quad (2)$$

정규분포 수식(2)에 고정된 평균과 분산으로 샘플링된 latent vector는 G 의 오차 함수가 가장 낮을 때의 특정 클래스의 데이터만 생성하는 mode collapse 문제가 발생할 확률이 높다.

$$z_i = z_i + \mu_i \quad (3)$$

그러므로 (그림 1)의 (b)와 같이 확장된 분포에 -1과 1 사이의 균일분포를 따르는 무작위 평균을 (3)과 같이 latent vector에 더하여 불규칙한 평균과 분산을 통해 다양한 데이터를 생성할 수 있다.

4. 실험

4.1. 평가지표

평가지표로는 IS(Inception Score)[5]와 FID(Fr chet Inception Distance)[6]를 사용하였다. IS는 Inception Network를 사용하여 생성된 데이터의 클래스를 예측하고 (4)을 통해 생성된 데이터의 품질과 다양성을 평가한다. IS가 높을수록 생성된 데이터의 품질과 다양성 측면으로 성능이 우수하다고 판단할 수 있다.

$$(4)$$

FID는 실제 데이터와 생성된 데이터의 다변량 정규분포를 추정하고 거리를 계산하여 두 데이터가 얼마나 비슷한지 (5)를 통해 나타내며 FID가 낮을수록 생성 데이터가 실제 데이터와 유사하다고 판단할 수 있다.

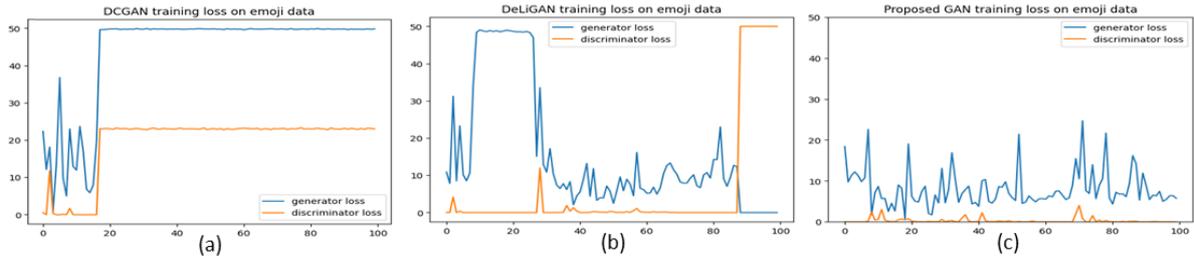
$$\frac{1}{w} \quad \frac{1}{w} \quad (5)$$

<표 1> DCGAN, DeLiGAN, Proposed GAN의 IS, FID 비교

	GAN	DeLiGAN	ProposedGAN
FID	0.30	0.34	0.24
IS	0.12	0.15	0.16

4.2. 데이터셋

데이터셋은 Emoji Generator 사이트에서 이모지를 크롤링하여 수집하였다. Positive, Negative 각각의 클래스에 402장의 이미지가 존재하고 총 804장의 Custom 이모지 데이터셋을 사용하였다.

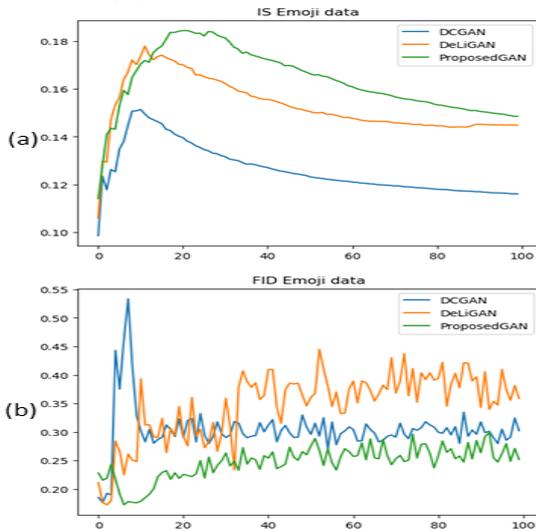


(그림 2) DCGAN, DeLiGAN, Proposed GAN 손실함수

4.3. 실험 및 결과

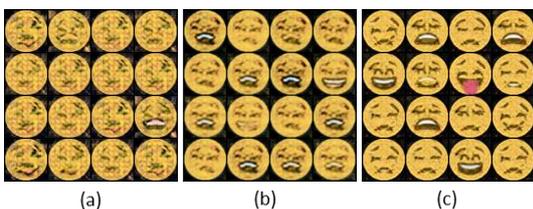
본 논문에서는 DCGAN, DeLiGAN, Proposed GAN 비교 실험을 진행하였다. Pytorch 버전은 1.12.1, 최적화 알고리즘은 Adam[7], Learning Rate 는 1e-2, Batch size 는 35 환경에서 학습을 진행하였다.

같은 환경, 데이터셋에서 각 모델의 결과는 (그림 2)와 같으며 DCGAN 은 부족한 데이터셋으로 인해 불안정 문제가 발생하는 것을 (a)를 통해 확인할 수 있다. DeLiGAN 은 epoch 90 까지 G 와 D 가 경쟁 구도를 생성하다가 이내 불안정 문제가 발생하는 것을 (b)를 통해 확인할 수 있다. Proposed GAN 은 초반에 실제 데이터 일부분의 분포를 학습하기 때문에 빠르게 경쟁구조를 생성하여 안정적으로 학습되는 것을 (C)를 통해 알 수 있다.



(그림 3) DCGAN, DeLiGAN, Proposed GAN 의 IS, FID

(그림 3)(a)는 각 모델의 IS, (b)는 각 모델의 FID 로 Proposed GAN 이 DCGAN 과 DeLiGAN 에 비해 실제 데이터와 유사하며 품질이 우수하고 다양한 데이터를 생성하는 것을 확인할 수 있다.



(그림 4) (a)DCGAN, (b)DeLiGAN, (c) Proposed GAN 이모지

5. 결론

본 논문에서는 다양하고 품질이 우수한 이모지를 생성하기 위한 latent vector 분포 조정 구조와 latent vector 평균 조정 기법을 제안한다. 생성된 이모지(그림 4)는 모델별 FID 가 낮은 지점에서 생성된 이모지이며 실험 결과 Proposed GAN 은 DCGAN 과 DeLiGAN 에 비해 안정적으로 학습하는 것을 보인다. 제안한 구조와 기법을 통해 DCGAN, DeLiGAN 각각 FID 는 81%, 71% 감소하였고 IS 는 129%, 106% 증가한 것을 <표 1>을 통해 알 수 있다. 제안하는 구조와 기법을 통해 DCGAN 과 DeLiGAN 보다 안정적으로 학습하면서 다양하고 품질이 우수한 데이터 생성 성능 검증을 입증했다.

6. Acknowledgement

본 논문은 2023 년도 정부(교육부)의 지원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2021R1G1A1006381)

참고문헌

- [1] 김선진. 모바일 메신저의 이모티콘 특성에 관한 비교 연구 - 국내, 해외 대표 앱(카카오톡, 라인, 페이스북, 왓츠앱)을 중심으로 -. 디지털디자인학연구, 14(1), 87-96. (2014).
- [2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. Generative adversarial networks. *Communications of the ACM*, 63(11), 1 39-144. (2020).
- [3] Radford, A., Metz, L., & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*. (2015).
- [4] Gurusurthy, S., Kiran Sarvadevabhatla, R., & Venkatesh Babu, R. Deligan: Generative adversarial networks for diverse and limited data. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 166-174). (2017).
- [5] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. Improved techniques for training gans. *Advances in neural information processing systems*, 29. (2016).
- [6] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30. (2017).
- [7] Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. (2014).