2차원 GAN 모델의 데이터 생성 성능 분석

김진욱[1], 김민재[2], 김예인[2], 권석주[2], 정주영[3], 이경률[4]
[1]국립목포대학교 정보보호기술학협동과정 석사과정
[2]국립목포대학교 정보보호학과 학부생
[3]국립목포대학교 컴퓨터학부 학부생
[4]국립목포대학교 컴퓨터학부 교수
wlsdnr0816@mokpo.ac.kr, alswo807@mokpo.ac.kr, yink4773@gmail.com,
ing9237@naver.com, jy0000@mokpo.ac.kr, carpedm@mnu.ac.kr

Analysis of the Data Generation Performance of Two-Dimensional GAN Models

Jinwook Kim[1], Minjae Kim[2], Yein Kim[2], Seokju Kwon[2], Juyeong Jeong[3], Kyungroul Lee[4]
[1]Dept. Interdisciplinary Program of Information & Protection, Mokpo National University
[2]Dept. of Information Security Engineering, Mokpo National University
[3,4]School of Computer Science and Engineering, Mokpo National University

요 약

디지털 시대에서 인공지능은 인간이 할 수 있는 글 작성, 그림 그리기와 같은 다양한 예술 분야에서 창작 활동을 위하여 활용되고 있다. 이러한 인공지능 기술을 생성형 AI라 불리며, 생성형 AI 기술 중에는 실제 데이터와 유사한 가짜 데이터를 생성하는 GAN 기술을 포함한다. GAN은 생성자와 판별자가 서로 경쟁하면서 데이터를 생성하고 구분하는 동작을 반복하며, 이를 통하여 실제와 매우 유사한 특징을 가지는 가짜 데이터를 생성한다. 이에 따라, GAN 모델의 키보드 데이터 보호의 활용 가능성을 검증하기 위하여, 다양한 GAN 모델을 기반으로, 2차원 데이터를 생성하고, 생성된 데이터의 성능을 분석하였다. 분석 결과, Random Forest 모델과 Gradient Boosting 모델에서 WGAN-GP 모델의 성능이 우수하지 않은 것으로 분석되었으며, CTGAN 모델은 다른 모델에 비하여 성능이 우수한 것으로 나타났다. 또한, 학습 횟수를 증가할 경우에는 Copula GAN의 성능이 가장 우수한 것으로 나타났다. 본 논문의 결과는 다양한 GAN 모델의 성능평가와 다양한 데이터로의 활용을 위하여 활용될 수 있을 것으로 사료된다.

1. 서론

인공지능(Artificial Intelligence) 기술이 발전함에 따라, 사람들은 글, 그림, 프로그래밍과 같은 다양한 콘텐츠 분야에서 적극적으로 활용하는 실정이다. 이와 같이, 콘텐츠를 창작하기 위하여 활용되는 인공지능 기술은 생성형 AI이며[1], 생성형 AI는 챗봇과 같이 인간과 유사한 대화를 생성하거나 딥페이크와 같이 사실처럼 보이는 가짜 정보를 생성할 수 있어 목적을 위하여 활용이 가능하다.

이처럼 데이터를 생성하는 모델 중 GAN(Generative Adversarial Networks)[2]은 생성형 AI 기술 중 하나로, 생성자와 판별자로 구성되어 실제 데이터를 기반으로 실제와 매우 유사한 가짜 데이터를 생성하는 것이 목적이다. 이는 딥페이크처럼 음성 데이터, 얼굴 이미지 합성과 같은 범죄 악용 가능성이 잠재적인 문제점으로 대두되지만, 의료 영상, 악성코드와 같은 분야에서의 부족한 데이터를 보완하고 분석하기 위한 데이터를 증강하는 장점을 가진다.

따라서 본 논문에서는 GAN 모델의 다양한 활용 가능성을 검증하기 위하여, 키보드 데이터를 보호하기 위한 목적으로, 키보드 데이터를 수집한 후, 수집된 키보드 데이터를 기반으로 2차원 가짜 데이터를 생성하고 성능을 분석한다[3]. 다양한 GAN 모델의 성능 분석 결과를 기반으로, 키보드로부터 입력되는 중요한 데이터를 보호하는 방안에서의 적용 가능성을 도출하고자 한다.

2. 배경 지식

2.1. GAN

2014년에 처음 제안된 적대적 생성 신경망이라 불리는 GAN은 생성자와 판별자로 구성된 두 개의 신경망이 서로 경쟁하면서 학습한다[2]. 여기서, 생성자는 실제 데이터를 기반으로 유사한 가짜 데이터를 생성하고, 판별자는 생성된 가짜 데이터와 실제 데이터를 구분하는 역할을 수행한다. 이와 같은 동작을 반복하면서, 생성자와 판별자는 서로 지속적으로 경쟁하며, 이를 통하여 성능을 향상시킴으로써 실제와 매우 유사한 특징을 가지는 가짜 데이터를 생성한다.

2.2. 2차원 데이터 생성 GAN 모델

본 논문에서는 2차원 데이터를 기반으로 데이터를 생성하는 GAN 모델들 중, CTGAN, Copula GAN, WGAN-GP, TabGAN을 대상으로 성능을 분석한다.

CTGAN은 표 형식의 데이터 분포를 모델링하고 샘플링하기 위하여 사용되는 GAN 기반 모델이다. 기존의 통계 및 딥러닝 모델들이 표 형식의 데이터를 제대로 모델링하지 못하는 문제를 해결하기 위하여, 조건부 생성기와 샘플링 기반 훈련 기법을 도입하였다. 또한, 모드별 정규화와 같은 구성 요소를 통하여 데이터 불균형 문제를 극복하였다[4].

Copula GAN은 CTGAN의 변형 모델로, 데이터 간의 의존 구조를 더욱 잘 반영하기 위하여, Copula 이론을 도입한 모델이다. Copula[5]는 다변량 분포 함수와 주변 분포 함수를 연결시키는 것으로, 개별 데이터의 분포와 데이터 간의 종속성 구조를 분리하여 모델링한다. 이때, 이 모델은 Copula 중에서도 Elliptical Copula 계열에 속한 Gaussian Copula를 활용한 변수 간 상관 구조를 학습하며 [5], SDV(Synthetic Data Vault) 라이브러리에 구현되어 있다.

WGAN-GP는 Wasserstein 거리가 도입된 WGAN(Wasserstein GAN)의 모델을 개선한 모델로, 판별자가 1-Lipschitz 함수가 되어야 하므로, 가중치 클리핑 방식이 적용되었지만, 이로 인하여 최적화에 어려움이 있다. 이러한 문제를 해결하기 위하여, 기울기 패널티를 도입하여 안정적인 훈련을 제공하는 모델이 WGAN-GP이다[6].

TabGAN은 WGAN 기반의 조건부 생성 구조와 수치형 변수 전처리를 위한 RQT(Randomized Quantile Transformation)를 활용하여 고품질의 가짜 데이터를 생성하는 프레임워크이다[7].

3. GAN을 활용한 2차원 데이터 생성

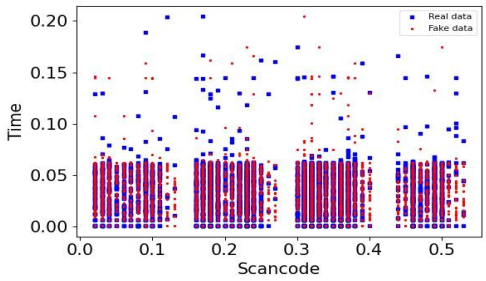
본 논문에서는 GAN 모델을 활용한 2차원 데이터 생성의 성능을 분석하기 위하여, 다양한 데이터 중 키보드 데이터를 기반으로, 데이터 수집 및 가짜 데이터 생성, 실제 데이터와 가짜 데이터의 유사도를 분석한다.

실험에 사용된 실제 데이터는 4,095개로, 실제 데이터와 가짜 데이터를 1:1 비율로 구성하고, 각 모델의 epoch를 50, 100, 200으로 설정하였다. 유사도 평가를 위하여, 실제 데이터와 가짜 데이터의 분포를 그래프로 표현하였으며, 파란색은 실제 데이터를 의미하고, 빨간색은 가짜 데이터를 의미한다.

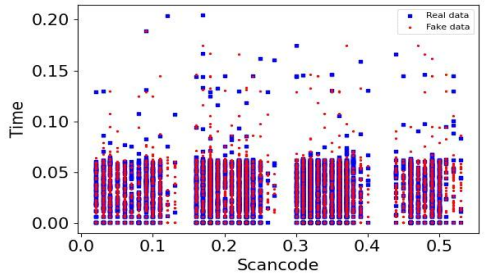
3.1. CTGAN 모델 데이터 분포

본 절에서는 CTGAN 모델을 기반으로, 실제 키보드 데이터와 생성된 가짜 키보드 데이터의 분포를 나타내고, 그 결과를 분석한다.

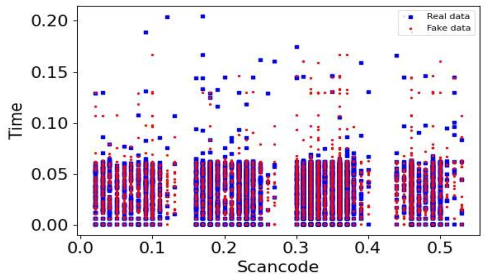
결과를 살펴보면, CTGAN 모델은 Epoch 50에서 가짜 데이터와 실제 데이터 간 차이가 나타났고, Epoch 100에서는 Epoch 50의 결과보다 분포가 더욱 개선되었으나, Time 축에서 차이가 나타났다. 마지막으로, Epoch 200은 실제 데이터와 가장 유사한 것으로 나타났다.



(그림 1) Epoch 50 기반 CTGAN 모델 데이터 분포



(그림 2) Epoch 100 기반 CTGAN 모델 데이터 분포

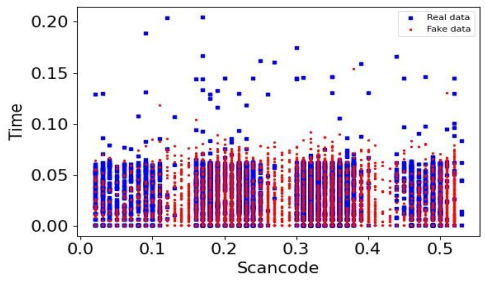


(그림 3) Epoch 200 기반 CTGAN 모델 데이터 분포

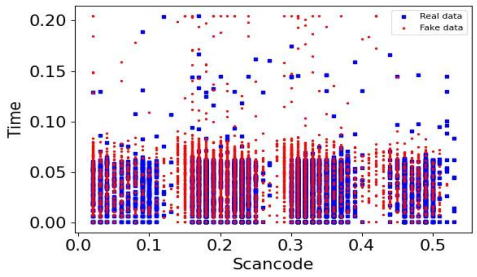
3.2. Copula GAN 모델 데이터 분포

본 절에서는 Copula GAN 모델을 기반으로, 실제 키보드 데이터와 생성된 가짜 키보드 데이터의 분포를 나타내고, 그 결과를 분석한다.

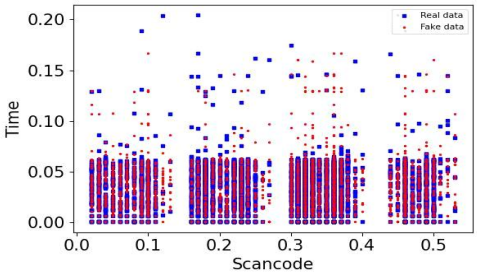
결과를 살펴보면, Copula GAN 모델은 Epoch 50에서 가짜 데이터와 실제 데이터 간 차이가 나타났고, 특정 구간에서는 실제 데이터보다 가짜 데이터가 과도하고 밀집한 양상이 나타났다. Epoch 100에서는 50보다 고르게 분포된 것으로 나타났으나, 노이즈성 데이터가 여전히 존재한다. 마지막으로, Epoch 200에서는 여전히 노이즈성 데이터가 일부 존재하지만, Scancode 구간별 가짜 데이터가 실제 데이터의 분포와 유사한 것으로 나타났다.



(그림 4) Epoch 50 기반 Copula GAN 모델 데이터 분포



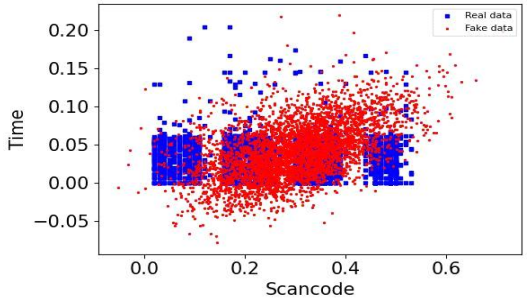
(그림 5) Epoch 100 기반 Copula GAN 모델 데이터 분포



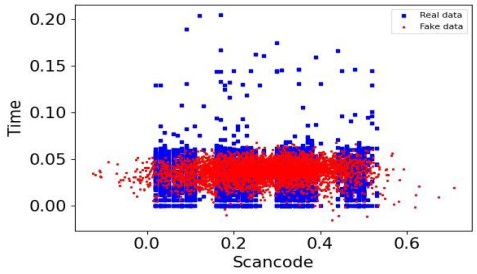
(그림 6) Epoch 200 기반 Copula GAN 모델 데이터 분포

3.3. WGAN-GP 모델 데이터 분포

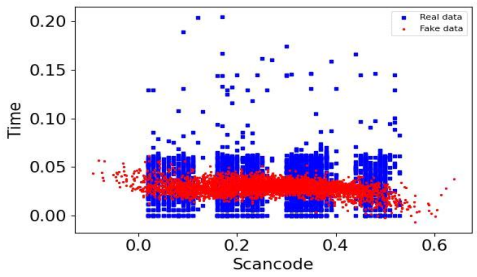
본 절에서는 WGAN-GP 모델을 기반으로, 실제 키보드 데이터와 생성된 가짜 키보드 데이터의 분포를 나타내고, 그 결과를 분석한다.



(그림 7) Epoch 50 기반 WGAN-GP 모델 데이터 분포



(그림 8) Epoch 100 기반 WGAN-GP 모델 데이터 분포

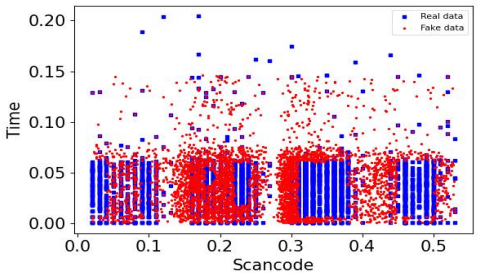


(그림 9) Epoch 200 기반 WGAN-GP 모델 데이터 분포

결과를 살펴보면, WGAN-GP 모델은 Epoch 50에서 넓은 범위로 데이터가 구성되고, Scancode 구간별 실제 데이터와 유사하지 않은 것으로 나타났다. Epoch 100에서는 분포의 범위가 50보다 좁아지고 안정적으로 보이지만, Scancode 구간에 포함되지 않는 데이터가 생성되었으며, 여전히 실제 데이터와 유사하지 않은 것으로 나타났다. 마지막으로, Epoch 200은 100보다 가짜 데이터가 더욱 좁은 범위에 집중된 것으로 나타났다.

3.4. TabGAN 모델 데이터 분포

본 절에서는 TabGAN 모델을 기반으로, 실제 키보드 데이터와 생성된 가짜 키보드 데이터의 분포를 나타내고, 그 결과를 분석한다.

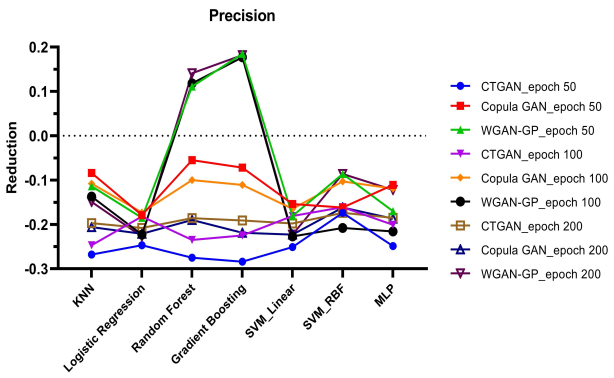


(그림 10) TabGAN 모델 데이터 분포

결과를 살펴보면, TabGAN 모델은 전체적인 데이터의 생성 분포가 실제 데이터의 범위에 포함되지 않는 데이터가 생성되는 것으로 나타났다. 이는 TabGAN 모델이 2차원 데이터 생성에 적합하지 않는 것으로 판단되며, 이러한 이유로 성능을 분석하기 위한 모델에서 제외하였다.

4. GAN 모델별 성능 분석

본 장에서는 상기 GAN 모델에서 생성한 가짜 데이터의 성능을 분석하기 위하여, 머신러닝을 활용한 성능의 증감 수치를 비교한다. 분류 성능을 평가하기 위하여 사용한 머신러닝 모델들은 KNN(K-Nearest Neighbors), Logistic Regression, Random Forest, Gradient Boosting, SVM_Linear/_RBF, MLP(Multi-Layer Perceptron)이며, 정밀도를 기준으로 성능 분석 결과를 (그림 11)에 나타내었다.



(그림 11) GAN 모델별 성능 분석 결과

그림을 살펴보면, Epoch 50에서는 CTGAN 모델이 다른 모델에 비하여 분류 성능을 감소시켰고, Epoch 100에서는 CTGAN 모델이 성능 감소가 높으며, 일부 모델에서는 WGAN-GP 모델이 성능 감소가 높은 것으로 나타난다. Epoch 200에서는 Copula GAN 모델이 성능 감소가 높게 나타나며, 대부분의 모델이 성능이 감소하는 것으로 보이지만, WGAN-GP 모델의 경우에는 Random Forest 모델과 Gradient Boosting 모델이 성능을 감소시키지 못하는 것으로 나타났다.

더욱 상세한 성능 분석을 위하여, 전체 GAN 모델들의 Epoch별 평균 성능을 비교하였으며, 그 결과를 <표 1>에 나타내었다.

<표 1> 전체 GAN 모델들의 평균 성능 분석 결과

학습	GAN 모델	정확도	정밀도	재현율	F1점수	AUC
랜덤	X	9.10	0.765	0.926	0.836	0.957
학습 50	CTGAN	0.539	0.515	0.677	0.582	0.584
		-0.371	**-0.250**	-0.249	**-0.254**	**-0.373**
	Copula GAN	0.664	0.648	0.651	0.649	0.725
		-0.246	-0.117	**-0.274**	-0.187	-0.232
	WGAN	0.731	0.702	0.809	0.740	0.765
		-0.179	-0.063	-0.117	-0.097	-0.192
학습 100	CTGAN	0.582	0.56	0.592	0.575	0.599
		-0.328	**-0.205**	**-0.334**	**-0.261**	**-0.358**
	Copula GAN	0.686	0.639	0.791	0.707	0.745
		-0.224	-0.126	-0.135	-0.130	-0.211
	WGAN	0.695	0.663	0.861	0.743	0.734
		-0.215	-0.102	-0.064	-0.094	-0.223
학습 200	CTGAN	0.597	0.574	0.630	0.599	0.626
		-0.313	-0.191	-0.296	-0.237	-0.331
	Copula GAN	0.580	0.564	0.561	0.561	0.614
		-0.330	**-0.201**	**-0.365**	**-0.275**	**-0.343**
	WGAN	0.698	0.696	0.648	0.658	-0.731
		-0.212	-0.069	-0.277	-0.179	-0.226

표를 살펴보면, 랜덤으로 생성된 가짜 데이터의 성능을 기준으로, Epoch 50에서는 CTGAN 모델, Epoch 100에서는 CTGAN 모델, Epoch 200에서는 Copula GAN 모델이 가장 성능이 우수한 것으로 나타났다. Epoch 100의 경우에는 분류 모델별 성능 분석에서 WGAN-GP의 성능이 우수하였으나, 일부 우수하지 못한 모델로 인하여, 전체 평균 성능이 낮아지는 것으로 나타났다. 모든 모델들로부터 생성된 데이터는 머신러닝의 분류 성능을 감소시키며, 실제와 유사한 가짜 데이터를 생성함으로써 키보드 데이터를 보호하기 위한 기술에서 효과적으로 활용이 가능함을 검증하였다.

5. 결론

생성형 AI 기술을 활용하여 다양한 콘텐츠를 효과적으로 제작할 수 있다. 생성형 AI 기술 중 하나인 GAN은 의료, 음악과 같은 다양한 분야에서 활용이 가능하지만, 음성 위조, 얼굴 합성을 통하여 악의적인 용도로 사용될 가능성도 존재한다. 이에 따라, GAN의 긍정적인 활용 방안을 도출하기 위하여, 다양한 GAN 모델들의 특징을 분석하고, 키보드 데이터를 기반으로 성능을 분석하였다.

분석 결과, Random Forest 모델과 Gradient Boosting 모델에서 WGAN-GP 모델의 성능이 우수하지 않은 것으로 나타났으며, CTGAN 모델이 가장 성능이 우수한 것으로 나타났다. 또한 학습 횟수를 증가할 경우에는 Copula GAN의 성능이 가장 우수한 것으로 나타났다.

이를 통하여, 키보드 데이터를 보호하기 위하여 실제와 유사한 가짜 데이터를 생성할 경우, 실제 키보드 데이터와 가짜 키보드 데이터의 머신러닝 분류 성능을 가장 감소시키는 CTGAN 모델이 효과적으로 데이터를 보호할 것으로 사료된다. 향후 더욱 다양한 GAN 모델과 다양한 데이터의 성능분석을 통하여 효과적으로 데이터를 생성하는 GAN 모델을 연구할 예정이다.

사사표기

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2022-NR066642)

참고문헌

[1] 이기석, 이승욱, 윤민성, 유정재, 오아름, 최인문, 김대욱, "디지털 에셋 창작을 위한 생성형 AI 기술 동향 및 발전 전망", 한국전자통신연구원 전자통신동향분석, 제39권, 제2호, pp. 33-42, 2024.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks", Communications of the ACM, Vol. 63, No. 11, pp. 139 - 144, 2020.

[3] J. Lee, W. Jeong, and K. Lee, "Keyboard Data Protection Technique Using GAN in Password-Based User Authentication: Based on C/D Bit Vulnerability", Sensors, Vol. 24, No. 4, 2024.

[4] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, "Modeling Tabular data using Conditional GAN", Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 7335 - 7345, 2019.

[5] 정준영, "Copula 함수를 이용한 결합분포 함수 추정 및 연관성 비교", 국내석사학위논문, 고려대학교 대학원, pp. 1-34, 2014.

[6] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs", Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 5769 - 5779, 2017.

[7] Rustad, Arne, "tabGAN: A Framework for Utilizing Tabular GAN for Data Synthesizing and Generation of Counterfactual Explanations", Master's thesis, Norwegian University of Science and Technology, pp. 1-135, 2022.