

Chapter
02생성모델(Generative Model)
연구 동향 및 금융에서의 활용

임선규_금융결제원 과장

I. 서론

최근 딥러닝 기술 중 생성 모델(Generative Model) 관련 연구가 활발하게 이루어지고 있다. 지난 몇 년간 생성 모델은 GAN(Generative Adversarial Networks), VAE(Variational Autoencoder) 등에 대한 연구가 주를 이루었고, 2021년 상반기 국제 딥러닝 컨퍼런스 ICLR(International Conference on Learning Representations) 2021에서는 Score-based 모델 등에서 의미 있는 연구결과가 제시되었다[1].

가트너는 최근 심포지엄에서 ‘Top Strategic Technology Trends for 2022’라는 주제로 트렌드 기술 12가지를 발표하였는데, 이 기술들 중 하나가 ‘Generative AI’이다[2]. 가트너는 현재 시장에서 가장 강력하고 가시적인 인공지능 기술 중 하나가 ‘Generative AI’라고 하였고, 2025년이면 세계적으로 생성되는 데이터의 10%가 ‘Generative AI’에 의해 만들어 질 것이라고 전망하였다[3].

생성 모델은 다양한 산업 분야에서 활용되고 있는데, 금융 산업에서도 생성 모델을 활용하는 연구가 계속되고 있다. 이상거래탐지를 위한 FDS(Fraud Detection System)에서 데이터 불균형 문제를 해결하기 위해 GAN 모델을 활용하는 연구, VAE 모델을 활용하여 이상탐지를 하는 연구 등이 그것이다.

본 고에서는 다양한 생성 모델들 중에 그간 연구가 활발했던 GAN 모델과 VAE 모델을 살펴보고, 이들 모델의 최근 연구 동향 및 금융 산업에서의 활용에 대해 살펴보고자 한다.

본 고의 구성은 다음과 같다. II장에서는 생성 모델의 기본 개념과 모델의 종류를 살펴본

* 본 내용은 임선규 과장(☎ 02-531-1306, sklim@kftc.or.kr)에게 문의하시기 바랍니다.

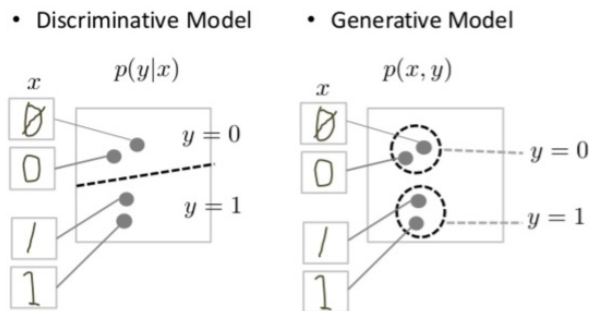
** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

다. III장과 IV장에서는 각각 GAN 모델과 VAE 모델에 대해 살펴본다. V장에서는 연구 동향 및 금융 산업에서의 활용에 대해 살펴보고, 마지막으로 VI장에서는 향후 전망을 제시한다.

II. 생성 모델

머신러닝(Machine Learning)에서 데이터의 분류(Classification)를 위한 모델은 판별 모델(Discriminative Model)과 생성 모델(Generative Model)로 구분할 수 있다. 판별 모델은 데이터 x 를 분류하는 레이블 y 를 결정하기 위해 조건부 확률 $p(y|x)$ 를 극대화하는 y 를 찾는다. 즉, 모델에서 직접적으로 $p(y|x)$ 를 학습하여 y 를 결정한다. 반면, 생성 모델은 데이터 x 의 레이블 y 를 결정하기 위해 간접적으로 결합확률 $P(x,y)$ 를 학습하거나, 만약 레이블이 없다면 $p(x)$ 를 학습한다.

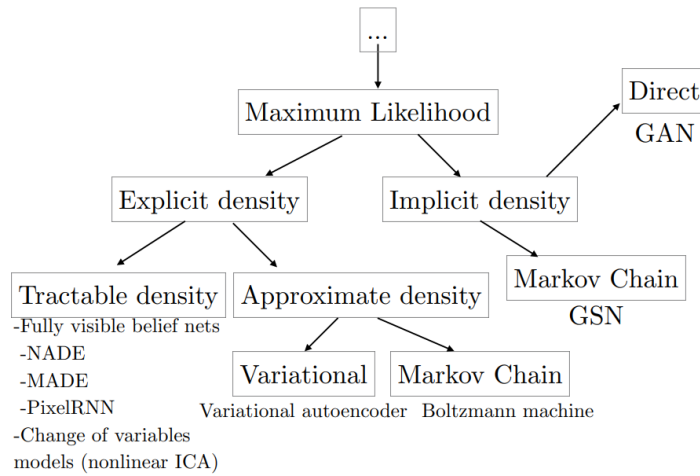
다시 말해, 판별 모델이 레이블의 차이를 적절하게 구분하는 결정경계(Decision Boundary)를 학습하여 데이터가 어떤 레이블을 갖는지 결정하는 모델이라면, 생성 모델은 데이터의 분포(Distribution)를 학습하여 주어진 데이터가 어떤 분포에 속할 확률이 높은지 결정하는 모델이다. [그림 1]은 판별 모델과 생성 모델의 이러한 특징을 개념적으로 보여준다.



〈자료〉 "Generative Adversarial Networks," <https://developers.google.com/machine-learning/gan/generative>, 2021.

[그림 1] Handwritten Digit를 인식하는 판별 모델과 생성 모델

생성 모델은 모델의 구성방식에 따라 다양하게 분류할 수 있다. [그림 2]는 생성 모델의 분류를 나타낸다. 우선 학습 데이터의 확률 분포를 알고 있는 상태에서 추정하는 Explicit Model과 확률 분포를 모르는 상태에서 샘플링을 통해 추정하는 Implicit Model로 분류할



〈자료〉 Goodfellow, Ian, "Nips 2016 tutorial: Generative adversarial networks," arXiv preprint arXiv:1701.00160, 2016.

[그림 2] 생성 모델(Generative Model)의 분류

수 있다. Explicit Model은 다시 확률 분포를 정확하게 정의할 수 있는 Tractable Model과 확률 분포를 정확하게 정의하는 대신 이에 대한 모델을 근사하는 Approximate Model로 나눌 수 있다. [표 1]은 생성 모델의 분류 중 Explicit Model과 Implicit Model의 주요 특징을 설명한다.

지난 몇 년간 다양한 생성 모델 중에서 Explicit & Approximate Model 중 하나인 VAE와 Implicit Model에 속하는 GAN에 대한 연구 및 활용이 활발하게 진행되었다. 따라서 본 고에서도 GAN 모델과 VAE 모델에 대해서 구체적으로 살펴보고자 한다.

[표 1] Explicit Model과 Implicit Model

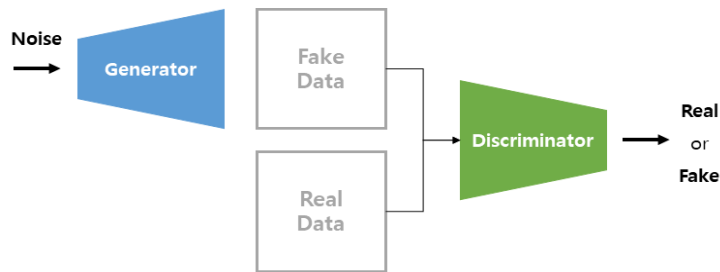
분 류	설 명	
Explicit Model	데이터의 확률 분포를 알고 있는 상태에서 추정하는 모델	
	Tractable Model	데이터의 확률 분포를 정확하게 정의하는 모델
	Approximate Model	데이터의 확률 분포를 근사하는 모델
Implicit Model	데이터의 확률 분포를 모르는 상태에서 샘플링으로 추정하는 모델	

〈자료〉 금융결제원 자체 작성

III. GAN

GAN는 최근 가장 주목받고 있는 생성 모델 중 하나이다. 딥페이크의 근간이 되는 기술로 대중적으로 많이 알려졌으며, Ian Goodfellow가 2014년 기본 모델을 제시한 이후 다양한 형태로 파생 모델들이 연구되었고, 이러한 모델들은 여러 산업에서도 활용되고 있다.

[그림 3]은 GAN 모델의 주요 구성요소 및 기본 개념을 나타낸다. GAN은 생성자(Generator)와 판별자(Discriminator)로 구성되는데, 생성자는 판별자가 페이크(Fake) 데이터를 식별할 수 없는 확률을 높이는 방향으로 학습하며, 이와 반대로 판별자는 생성자의 페이크 데이터를 리얼(Real) 데이터와 비교해 이를 식별해 내는 확률을 높이는 방향으로 학습한다. 쉽게 말해, 생성자는 판별자를 속이기 위해 노력하며, 판별자는 생성자에게 속지 않기 위해 노력한다.



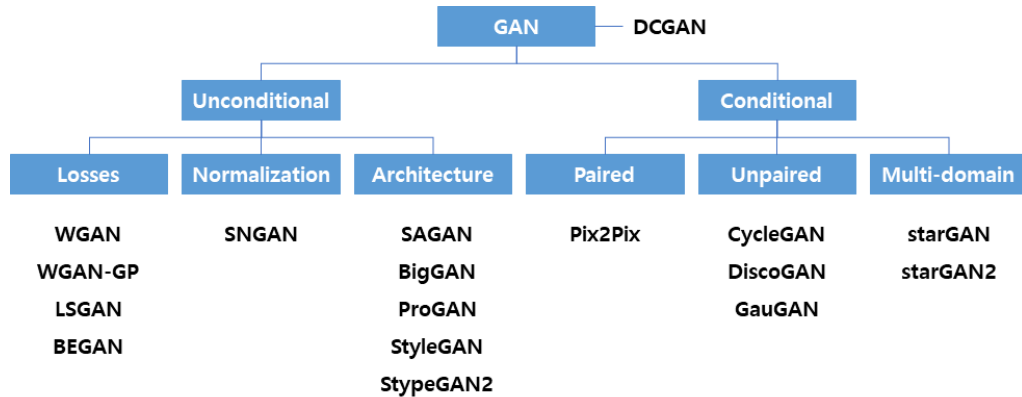
〈자료〉 금융결제원 자체 작성

[그림 3] GAN 모델의 기본 개념도

GAN의 이론적 기반은 Min-max 문제에서 경쟁자가 상호 대응에 따라 최선의 선택을 하면서 결과적으로 자신의 선택을 변경하지 않는 내쉬 균형(Nash Equilibrium)을 찾는 것이다. 생성자와 판별자가 서로 합리적인 판단을 하는 수학적 균형점을 찾아가는 과정에서 서로 충돌하며 발생하는 차이를 수정한다. 이러한 이론적 기반은 GAN의 목적함수(Objective Function)에 포함되어 있다.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log \{1 - D(G(z))\}]$$

판별자 D는 목적함수 V(D, G)가 최대가 되길 바라며, 이는 리얼 데이터 x를 정확하게 식별하고(D(x)=1), 페이크 데이터 G(z) 또한 정확하게 식별할 수 있도록(D(G(z))=0) 노력한



〈자료〉 “Deep Learning GAN Tutorial,” <https://ysbsb.github.io/gan/2020/06/17/GAN-newbie-guide.html>, 2020.

[그림 4] GAN 모델의 분류

다는 것을 의미한다. 반대로 생성자는 목적함수가 최소가 되길 바라며, 판별자가 리얼 데이터를 정확하게 식별하지 못하고($D(x)=0$), 페이크 데이터 또한 정확하게 식별하지 못하도록($D(G(z))=1$) 노력함을 의미한다.

GAN 모델은 이미지, 비디오, 텍스트 등의 데이터를 생성 및 활용하는 다양한 분야에서 사용되고 있다. Original GAN 모델은 기존의 단일 네트워크 학습 모델과 달리 구조적인 문제로 인해 발생하는 학습의 불안정성 문제가 있었다. 이는 CNN(Convolutional Neural Networks) 적용을 통해 성능이 대폭 향상된 DCGAN(Deep Convolutional GAN)이 제시되면서 해결되었고, 이후 GAN은 더욱 주목을 받으며 다양한 파생 모델들이 연구되면서 급속도로 발전하였다. DCGAN으로부터 파생된 모델들을 분류하면, [그림 4]와 같이 대표적으로 랜덤하게 데이터를 생성하는 Unconditional GAN과 원하는 데이터를 생성하기 위해 특정 조건(나이, 성별 등)들을 부가적으로 활용하는 Conditional GAN으로 나눌 수 있다. Conditional GAN은 입력과 출력 이미지 사이의 관계 정보의 포함여부에 따라 Paired와 Unpaired 등으로 분류할 수 있다. [표 2]는 이미지 생성 관련 주요 GAN 모델을 설명한다.

그 동안 GAN 모델에 대한 연구는 이미지 데이터 생성 분야에서 활발히 진행되어 왔는데, 최근에는 이미지뿐만 아니라 동영상, 자연어 처리(텍스트 및 음성) 등 다양한 데이터를 대상으로 한 활용에도 연구가 진행되고 있다.

[표 2] 이미지 데이터 생성 관련 주요 GAN 모델

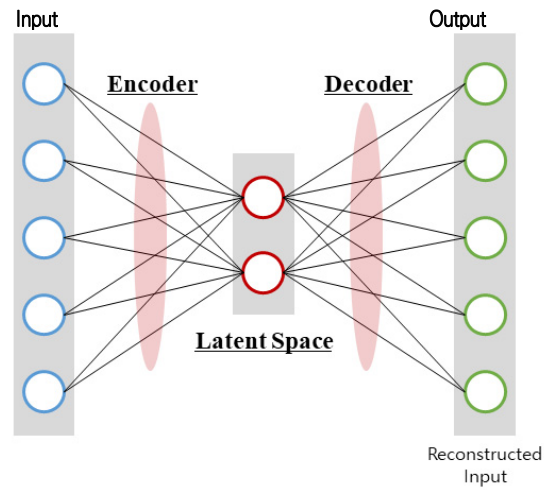
GAN 모델	설명
WGAN(2017)	GAN 개선을 위해 Wasserstein Distance 기반 손실함수를 재정의한 모델
WGAN-GP(2017)	WGAN 개선을 위해 Gradient Penalty 기반 판별자를 최적화한 모델
LSGAN(2017)	GAN 개선을 위해 판별자에 Least Square Loss를 적용한 모델
BEGAN(2017)	손실함수를 이미지 데이터의 분포 대신 자체적인 분포를 대상으로 변경
SNGAN(2018)	판별자 학습 안정화를 위해 Spectral Normalization을 제안한 모델
SAGAN(2018)	모든 Feature Location에서 획득한 정보 기반의 이미지 생성 모델
BigGAN(2019)	SAGAN을 기반으로 배치 크기 및 Depth 등 스케일 증가한 모델
ProgressiveGAN(2018)	고해상도 얼굴 이미지 생성을 위해 초저해상도 이미지 활용 모델
SinGAN(2019)	ProGAN 응용 및 단계별 생성자/판별자 사용으로 이미지 생성/변형
StyleGAN(2019)	ProGAN의 발전된 모델, 얼굴 이미지에서 헤어색상, 성별 등 변형 가능
Pix2Pix(2017)	입출력 이미지 매핑을 기반으로 이미지를 변환하는 최초의 모델
Pix2PixHD(2019)	Pix2Pix에서 겹쳐있는 동일 물체의 차이 표현을 위해 정보 추가
LostGAN(2019)	물체의 모양에 대한 정보 없이 위치 정보(Layout)로 이미지 생성
CycleGAN(2017)	입출력 이미지 매핑 정보 없이 이미지를 변환하는 모델
DiscoGAN(2017)	상이한 두 도메인의 관계를 발견해 도메인 간 데이터를 변환하는 모델
GauGAN(2019)	배치 정규화 계층을 보완해 이미지의 Semantic과 Style 정보 강화
StarGAN(2018)	단일 네트워크를 기반으로 이미지에서 원하는 부분만 변환 가능
Deepfill(2018)	이미지에서 유실 또는 삭제된 부분을 복원하는 모델
SRGAN(2017)	저해상도 이미지를 고해상도로 변환하는 Super Resolution 모델

〈자료〉 금융결제원 자체 작성

IV. VAE

VAE(Variational Autoencoder) 모델을 살펴보기에 앞서 모델의 구조를 먼저 이해하기 위해 AE(Autoencoder) 모델을 참고로 살펴볼 필요가 있다. AE 모델은 입출력 데이터 사이에 저차원의 Latent Space를 두어 Bottleneck 구조로 구성된다.

AE 모델에서는 고차원의 입력 데이터를 Encoder를 통해 저차원의 Latent Space로 압축하며, 이렇게 압축된 정보를 Decoder를 통해 다시 고차원의 출력 데이터로 복원한다. 즉, 데이터를 압축하고 다시 재구성할 때 발생하는 오류를 최소화할 수 있는 Latent Space를 생성할 수 있도록 학습하는 모델이다. [그림 5]는 AE 모델의 이러한 구조를 나타낸다.



〈자료〉 금융결제원 자체 작성

[그림 5] AE 모델의 기본 구조

대표적인 AE 모델은 [표 3]과 같이 복잡한 부호화가 가능하도록 다수의 저차원 Latent Space를 대칭적으로 쌓은 Stacked AE 모델, 입력 데이터에 Noise를 추가하여 Noise가 제거된 원본의 입력데이터를 복원하도록 학습하는 Denoising AE 모델, 입력 데이터에서 보다 유용한 특징을 학습하기 위해 활성화된 노드 수를 감소시키는 Sparse AE 모델 등이 있다.

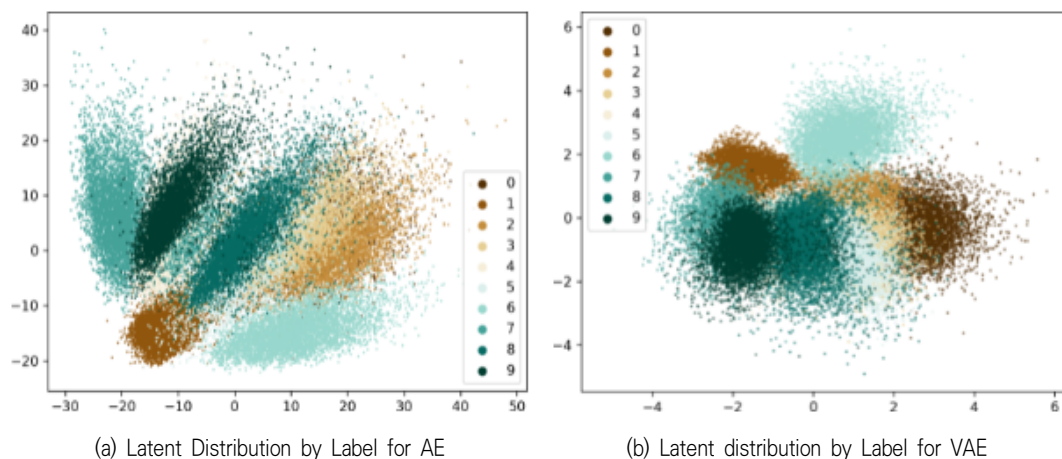
VAE 모델은 AE 모델과 구조적으로 유사하지만 확률적(Probabilistic) 및 생성적(Generative) 개념이 추가로 활용된다. AE 모델은 Single Vector로 표현할 수 있는 Latent Space를 학습하는데 반해 VAE 모델은 Latent Space의 확률 분포에 대한 평균과 표준편차를 학습한다.

[표 3] 대표적인 AE 모델의 종류

종류	설명
Uncomplete AE	기본적인 AE 모델로, 고차원의 입력 데이터를 저차원 공간으로 변환하며 중요한 특징을 학습
Stacked AE	Encoder/Decoder에 대칭으로 Hidden Layer를 추가해 복잡한 부호화가 가능한 모델
Denoising AE	데이터에 Noise가 추가되었을 경우 이를 제거하여 원본 데이터를 추출하는 모델
Sparse AE	입력 데이터에서 좀 더 의미있는 특징을 학습하기 위해 희소성(Sparsity)을 이용한 모델

〈자료〉 금융결제원 자체 작성

[그림 6]은 MNIST 데이터에 대해 AE 모델과 VAE 모델의 Latent Space의 분포를 시각화한 것이다. 각 포인트의 색상은 0부터 9까지의 숫자 이미지 레이블을 의미한다. [그림 6]을 통해 VAE 모델이 AE 모델보다 분포 측면에서 군집화가 더 잘 이루어진다는 것을 알 수

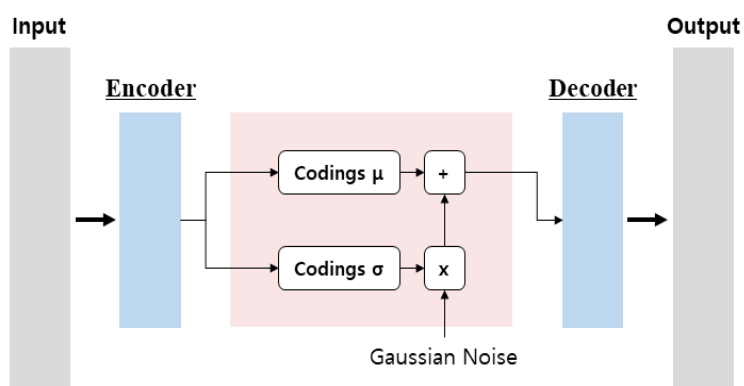


〈자료〉 T. Spinner et al. "Towards an interpretable latent space: An intuitive comparison of autoencoders with variational autoencoders," in Proc. Workshop Vis. AI Explainability(Berlin, Germany), Oct. 2018.

[그림 6] AE 모델과 VAE 모델의 Latent Distribution

있다. 즉, VAE 모델이 AE 모델보다 저차원의 Latent Space로 압축된 데이터를 고차원의 원본 데이터로 더 잘 복원할 수 있다고 해석할 수 있다.

VAE 모델을 좀 더 구체적으로 살펴보면, VAE의 Decoder는 Encoder가 생성한 Latent Space의 사후확률을 학습하는 것이 목적이다. 즉, 입력 데이터 x 와 Latent Space z 의 사후 확률 $p(z|x)$ 를 학습하는 것이다. 하지만 이 사후확률은 직접적으로 계산하기 어려워 변분추론(Variational Inference)을 통해 근사적으로 학습한다. Latent Space z 의 분포 $q(z)$ 로



〈자료〉 금융결제원 자체 작성

[그림 7] VAE 모델의 기본구조

근사하기 위해 $p(z|x)$ 와 $q(z)$ 사이의 KL Divergence를 계산하고, 이 값이 감소하는 방향으로 $q(z)$ 를 갱신하는 절차를 수행한다.

[그림 7]을 보면 앞서 언급하였듯이 VAE 모델은 Latent Space를 학습할 때 이에 대한 확률 분포인 평균(μ)과 표준편차(σ)를 학습한다.

V. 연구 동향 및 금융에서의 활용

최근 국제 딥러닝 컨퍼런스 ICLR 2021에서는 생성 모델에 대한 연구들 중에서 Score-based 모델에서 의미 있는 연구 결과가 제시되었다.

Score-based 모델은 데이터의 Noise Scale에 따라 이를 추가하면서 사전확률을 만드는 데, 각 단계에서 Log Likelihood Gradient인 Score를 추정하고, Back-propagation에서 Score를 이용하여 사전확률로부터 데이터를 생성하는 모델이다. ICLR 2021에서 발표된 논문 “Score-based generative modeling through stochastic differential equations” [4]에서는 사전확률을 구하기 위해 Noise를 추가하는 과정을 Stochastic Process로 보고 해당 과정이 포함된 미분방정식인 SDE(Stochastic Differential Equation)를 풀어낸다[1].

본 고에서 제시된 아이디어는 Score-based 모델로는 처음으로 High Fidelity를 갖는 1024×1024 이미지를 생성할 수 있음을 보이면서 Score-based 생성 모델이 실용적으로 사용될 수 있는 가능성을 보였다[5].

또한, 컴퓨터 비전과 패턴 인식 컨퍼런스인 CVPR(Computer Vision and Pattern Recognition) 2021에서는 생성 모델의 상용화에 주요 걸림돌 중 하나인 오류를 수정할 수 있는 기술이 발표되었다. “Automatic Correction of Internal Units in Generative Neural Networks”에서는 생성 모델 내부에서 문제를 일으키는 뉴런을 찾아 제거하는 알고리즘을 고안하였다. XAI(eXplainable AI) 기술을 활용해 생성 모델 내부의 어떤 위치에 결함이 존재하는지 파악하는 기술이다[6].

다양한 산업 분야에서도 생성 모델을 활용하고 있는데, 예를 들어, 의료 산업에서는 양질의 데이터 부족, 개인정보 재식별 위험 등 실제 의료데이터의 한계를 보완하기 위한 목적으로 합성 의료데이터 생성 시 활용하고 있으며[7], 방송 산업에서는 아나운서, 기상캐스터, 광고모델 등 가상 인간을 생성하여 방송 콘텐츠의 경쟁력을 강화하는데 활용하고 있다.

금융 산업에서도 생성 모델을 활용하기 위한 연구들이 진행되고 있다. 이상거래를 탐지하는 FDS(Fraud Detection System)에서는 단순한 규칙 기반의 탐지에서 고도화한 행위 기반의 탐지를 위해 딥러닝 기술을 활용할 수 있다.

여기서 문제는 금융거래의 특성상 일반적으로 대부분의 거래는 정상 상황의 거래이고, 비정상 거래는 상당히 소수에 해당한다는 것이다. 따라서 정상 거래와 비정상 거래의 데이터 불균형 문제가 발생할 수 있다. 이 경우, GAN 모델을 활용한 비정상 거래의 오버샘플링을 통해 데이터 불균형 문제를 해결함으로써 머신러닝 기반의 FDS를 고도화할 수 있다.

FDS에서 비정상 거래의 오버샘플링에 GAN 모델을 적용한 연구들을 보면 GAN, cGAN(conditional GAN), WGAN(Wasserstein GAN), WCGAN(Wasserstein Conditional GAN), ACGAN(Auxiliary Classifier) 등 다양한 모델들이 활용되었다[8]-[11]. [8]에서는 각 분류기(DNN, CNN, Random Forest)마다 오버샘플링에 활용할 다양한 GAN을 적용하여 성능을 실험하였는데, 대체로 CNN 모델에서 좋은 성능을 보였다. [9]에서는 신용 데이터 불균형 문제를 개선하기 위해 데이터를 증폭하는 SMOTE(Synthetic Minority Over-sampling Technique)와 데이터를 생성하는 ACGAN을 적용해 실험하였는데, SMOTE 기법이 더 나은 성능을 보였다.

FDS에서 데이터 불균형 문제를 해결하는 용도뿐만 아니라, 이상거래를 탐지하는데 직접적으로도 생성 모델이 적용될 수 있다. 기본적인 방법으로 AE 모델의 Latent Space를 정상적인 금융 거래들로 학습한 후 FDS에 적용하면, 비정상 거래의 경우 AE 모델의 Decoder가 복원한 결과에서 상당한 오차가 발생하여 탐지를 할 수 있다. 또한, RaPP(Reconstruction along Projection Pathway) 기반의 이상탐지 모델[12]을 적용할 경우 Latent Space에서의 변화까지 파악하여 인지할 수 있어 보다 좋은 성능으로 이상거래를 탐지할 것으로 예상된다.

VI. 결론

지난 몇 년간 생성 모델에 대한 연구는 매우 활발하게 진행되었다. 일반 사용자들에게는 딥페이크라 불리는 이미지 합성 기술로 알려지면서, 범죄에 악용되는 경우도 많았고, 재미를 위한 콘텐츠를 생성하는 서비스들에서도 활용되었다. 연구 영역에서도 마찬가지로 이미지의 생성, 변환 등을 중심으로 다양한 모델들이 제시되어 왔다.

향후 생성 모델은 생성의 정확도 등 성능 측면에서 지속적으로 개선될 것으로 예상되며, 이에 따라 의료, 교통, 금융, AI윤리 등 다양한 분야에서 활용될 것이다. 예를 들어, 의료 산업에서는 의료영상에서 Noise를 제거하는데 적용될 수 있고, 교통 분야에서는 자율주행 모델의 학습을 위해 데이터를 증강할 수 있다[13]. 금융 산업에서도 앞서 언급했던 바와 같이 이상거래를 탐지하기 위한 FDS에 더욱 고도화된 방식으로 적용될 수 있다.

가트너에서 발표한 2022년 주요 기술에서 ‘Generative AI’를 주목한 이유처럼 다가오는 미래에는 우리가 활용하는 많은 데이터들이 생성 모델을 통해 만들어질 수도 있을 것이다.

● 참고문헌

- [1] Sanghyu Yoon, “[ICLR 2021] 2편: ICLR 2021 속 Generative model 트렌드,” 2021.
- [2] Gartner, “Gartner Identifies the Top Strategic Technology Trends for 2022,” 2021.
- [3] NextDaily, “가트너, 2022 전략기술 트렌드 톱12 발표,” 2021. 10. 22
- [4] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B., “Score-Based Generative Modeling through Stochastic Differential Equations”, arXiv e-prints, 2020.
- [5] SIA, “Review Score-based Generative Modeling by Diffusion Process,” 2021.
- [6] Tousi, Ali, et al., “Automatic Correction of Internal Units in Generative Neural Networks,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [7] 김보라, 권오연, “인공지능을 활용한 ‘합성 의료데이터’: GAN 기술 중심으로”, 한국보건산업진흥원, 2020.
- [8] 김예원, 유예림, 최홍용. “생성적 적대 신경망과 딥러닝을 활용한 이상거래탐지 시스템 모형”, Information Systems Review 22.1, 2019, pp.59-72.
- [9] 황정현, 김강석, “이상거래탐지를 위해 GAN 을 사용한 효율적인 도메인 적응 방법”, 아주대학교, 2021.
- [10] 송실대학교, “핀테크 서비스 금융사기 방지를 위한 비대면본인확인 및 이상거래필터링 기술”, 과학기술정보통신부, 2021.
- [11] 한정수, 이채현, 고경찬, 우종수, 홍원기, “GAN 을 활용한 비트코인 불법거래 과표본 추출 기법”, 통신망운용관리 학술대회, 2020.
- [12] Kim, K. H., Shim, S., Lim, Y., Jeon, J., Choi, J., Kim, B., & Yoon, “Rapp: Novelty detection with reconstruction along projection pathway,” International Conference on Learning Representations, 2019.
- [13] 매일경제, “‘AI 편향성’ 논란 뜨거운데…대안으로 떠오른 ‘생성모델’”, 2021. 5. 6.