인공지능 기반 이상감지 기술

강지영
정보관리기술사
zhining@naver.com

도입

코로나 팬데믹 시대에 우리의 일상엔 많은 변화가 있었다. 소소하게는 마스크 착용부터 시작해서 비자발적인 직업의 변화를 겪기도 했고, 비대면 수업, 재택 근무와 같이 예전에 하던 활동이라도 다른 모습으로 진행하고 있다. 집 밥을 먹는 횟수가 늘어나면서 요리 실력 혹은 배달 음식 주문 스킬이 날로 향상되는 것은 덤이다. 사람들 간의 물리적인 거리는 멀어졌지만 전자 기기와는 훨씬 더 가까워졌다. 이처럼 급격히 변화한 흐름에 발 맞추어 가기 위해서는 아무리 사람 냄새 나는 아날로그 감성이 좋다고 하더라도, 디지털을 배척할 수는 없다. 남녀노소 직업을 불문하고 스마트폰을 활용한 콘텐츠 제작, SNS 홍보, 실시간 커뮤니케이션 등 온라인 세상의 활동을 즐길 줄 알고 디지털 변환을 꾀하는 사람만이 소통할 수 있는 시대가 되었다.

사실 팬데믹은 좀 더 빠르게 우리 생활에 디지털 변환을 스며들게 했을 뿐, 세상은 계속해서 변화하고 있었다. 4차 산업혁명의 핵심이 되는 기술인 인공지능, 클라우드, IoT 기술이 발달하고, 5G로 초연결 시대가 펼쳐지면서 모두가 연결되고 빠르게 처리할 수 있는 기반 기술은 끊임없이 발전하고 있다. 정말 아이러니하게도 디지털화

가 가속되는 세상에서 각광을 받는 기술은 '사람처럼' 행동할 수 있는 기술이다. 사람처럼 사물을 인식하고, 사람처럼 말하고, 사람처럼 판단할수록 우리는 그 기술에 열광한다. 결국 기술을 사용하는 것은 사람이기 때문이다. 그렇기 때문에 현실 세계에 대한 명확한 이해가 있어야 하고, 사람에 대한 깊이 있는 연구가 선제적으로 진행이 되어야만 쓸모 있는 기술을 만들 수 있다.

따지고 보면 기계를 사람처럼 공부시키는 기계학습(Machine Learning)도, 신경망을 기계에 추상화시켜 만들어 놓은 딥러닝(Deep learning)의 복잡한 알고리즘도 결국 사람의 두뇌를 컴퓨터에 만들어 놓은 것이다. 갓난아이가 수만 번 '엄마'라는 단어를 들으며 '엄마'를 인지하는 것처럼, 인공지능 기술은 컴퓨터에게 수만 개의 '강아지' 사진을 주고 '강아지'를 인지하도록 지도학습(Supervised Learning) 시킨다. 또한 인공지능은 당근과 채찍을 적절히 사용해서 알파고에게 바둑 두는 법을 알려주었고, 한 단어의 주변에 나오는 문맥을 학습하면서 자연어를 이해하는 능력도 향상시켰다. 최근에 인공지능 분야에서 나오는 SOTA(State of the art)를 갱신하는 논문을 보면, '아차, 사람도 이런 식으로 판단했었지!'하며 나 자신을 성찰해보는 시간을 갖기도 한다.

리처드 도킨슨의 '이기적 유전자'를 보면, '진화란 유전자가 오류를 막기 위해 모든 노력을 하고 있음에도 불구하고 막무가내로 생겨난 일'이라고 한다. 생물체의 유전자는 유리한 방향이라면, 즉 생존을 위해서 또는 안정성을 증가시키기 위해서 자기 복제를 통해 그 능력을 극대화 시키려고 한다. 그 과정에서 오류가 발생하고, 발생한 오류로 인해 새로운 종으로 진화도 할 수 있는 것이다.

필자는 여기서 '오류'에 대한 부분이 참 흥미롭게 느껴진다. 인류는 지금까지 수많은 기술을 만들어 왔고, 대부분의 기술은 인류가 더 나은 삶을 영위하기 위한 방법을 찾기 위해 개발되었을 것이다. 그럼에도 불구하고 예상치 못한 문제가 항상 발생한다. IT 시스템이라면 누군가 악의적인 목적을 가지고 시스템의 빈틈을 찾아서 오류를 발생시킬 수도 있고, 악의적이지 않더라도 예상치 못한 상황의 조합이 문제를 발생시키기도 한다. 이처럼 의도하지 않은 정상적인 시스템의 동작을 '이상(anomaly) 현상'이라고 한다. 어찌 보면 '이상 현상'은 기술의 진화를 촉진시키는 역할을 하는 것일지도 모르겠다. 물론 '이상 현상'을 마냥 긍정적으로 바라만 볼 수는 없다. 이상 현상을 막기 위해 모든 노력을 기울여야만 하고, 그 과정 속에서 자연스럽게 기술이 진화되는 것이 순리일지도 모르니 말이다.

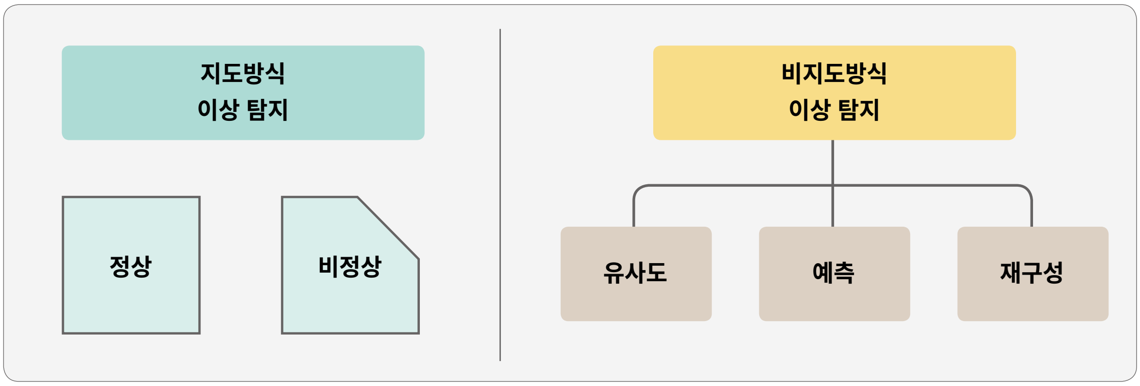
그렇다면 '이상 현상'을 해결하는 가장 첫 단계는 무엇일까? 바로 지금 발생한 현상이 문제, 즉 이상 상태임을 인지하는 것이다. 이상 상태는 정상이 아닌 상태다. 정상적인 상황과는 그 양상이 너무 달라 문제를 쉽게 인지할 수 있는 경우도 있지만, 정상과 굉장히 유사한 듯 보여서 정상이 아니라는 것을 인지하기도 어려운 경우도 있다.

나날이 인공지능 알고리즘의 성능이 좋아지고, 정교한 문제도 해결할 수 있지만 이와 더불어 이상 현상도 함께 발생한다. 다양한 이상 탐지(anomaly detection) 기술에 대해서 살펴보겠다.

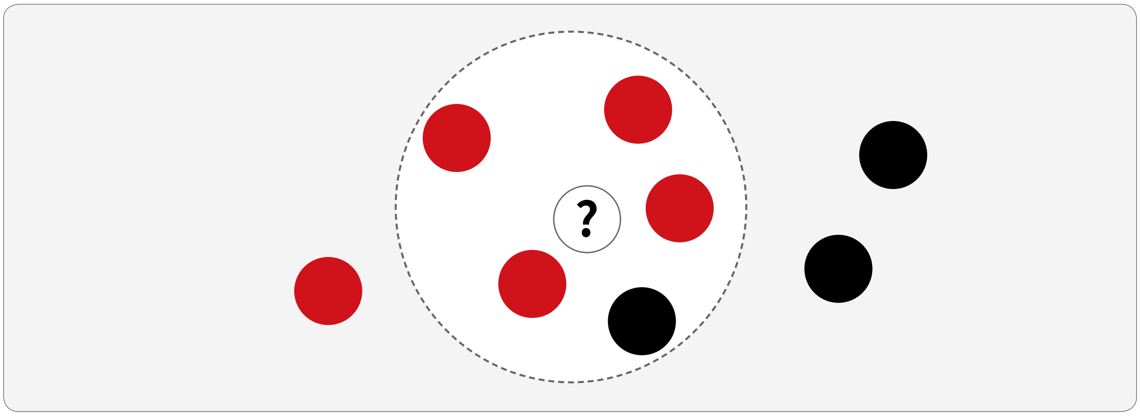
비지도(Unsupervised) 방식의 이상 탐지 방법

이상 현상은 다양한 분야에서 발생할 수 있는 만큼, 특정 비즈니스 도메인에 특화되어 이상 탐지를 위한 연구가 진행되었다. 정상(normal) 상황과 이상(anomaly)인 상황을 명확히 규정할 수 있다면, 특정 이벤트가 발생했을 때에 이상적으로 이상(anomaly)을 탐지할 수 있을 것이다. 하지만 이상 상황은 공식처럼 명확하게 규정할 수 없다. 이전에 발생했던 이상(anomaly) 상황에 대해서는 정의할 수 있겠지만, 예상치 못한 이상(anomaly) 상황이 항상 존재할 수밖에 없다. 이는 우리가 현실 세계

[그림 1] 지도방식의 이상 탐지 방법과 비지도 방식의 이상탐지 방법 개념도



[그림 2] 유사도 기반 이상 탐지 방법: KNN 알고리즘



출처 : https://algobeans.com/2016/09/14/k-nearest-neighbors-anomaly-detection-tutorial/

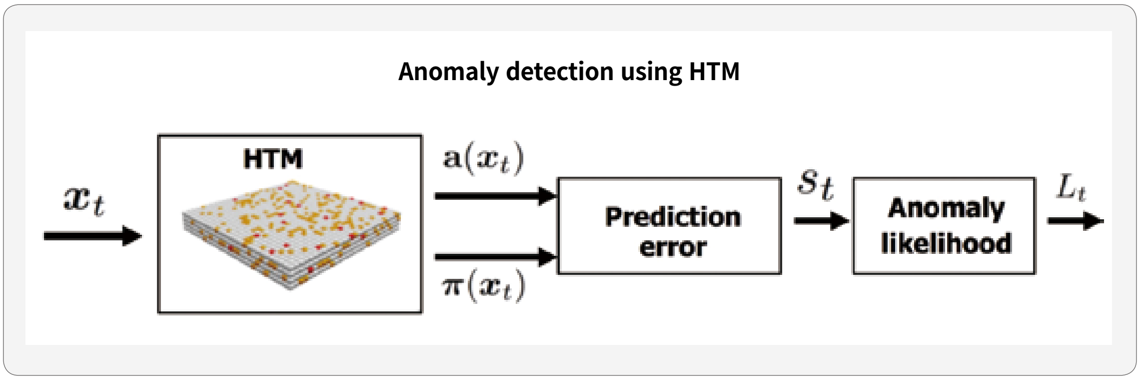
에서 발생하는 모든 이벤트를 다 규칙으로 만들 수 없기 때문이다.

그렇기 때문에 이상탐지 분야는 지난 몇 년 동안 다양한 데이터 유형 및 애플리케이션 시나리오에 대해 비지도(unsupervised) 방식으로 연구되었다. 비지도(unsupervised) 방식이란 정답을 알지 못하는 상태에서 데이터의 숨겨진 패턴과 특성을 발견하여 분석하는 방식이다. 이처럼 비지도 방식 기반의 이상탐지 방법은 크게 (1)유사도 기반 판별 방식, (2)예측 기반 판별 방식, (3)재구성 기반 판별 방식으로 나눌 수 있다.

(1) 유사도 기반 판별 방식

유사도 기반 판별 방식은 개체 간의 거리를 측정하여 개체 간의 유사성을 정량화하고, 일정 임계치 이상 거리가 떨어져 있다면 이상으로 간주한다. 이러한 감지 유형은 K-Nearest Neighbor(KNN)와 같은 거리 기반 알고리즘으로 구현할 수 있다. [그림 2]에서 정상 이벤트와 비정상 이벤트의 분포가 각각 빨간색, 검정색으로 나타나 있다고 가정할 때, 새로 발생한 이벤트(? 로 표시)가 k=5를 기준으로 이상(anomaly) 여부를 판별한다고 한다면,

[그림 3] 예측기반 이상 탐지 방법 : HTM



출처 : https://www.sciencedirect.com/science/article/pii/S0925231217309864

정상으로 판별하게 된다. 즉, KNN 알고리즘은 클러스터링을 위한 이웃(neighbor) 수를 어떻게 지정하는지가 중요한데, 알고리즘의 정확도를 높이기 위해서는 해당 도메인에 대한 충분한 사전 지식이 필요하다. 또한 KNN 알고리즘은 데이터의 분포가 특정 부분에 밀집해 있는 경우, 즉 지협적으로 이상(anomaly)이 발생한 경우에는 판별하기 어렵다는 한계가 있다.

다른 방법으로는 지정된 반경을 사용하여 정의한 객체의 근처에 위치한 이웃 수를 기준으로 하여 이상 점수를 계산하는 밀도(density) 기반 판별 방법이 있다. 밀도 기반 판별 방식은 치우친 데이터 분포도 고려하여 클러스터링 할 수 있지만, 여전히 이상(anomaly)에 대한 임계치를 정하는 것이 어려운 문제가 있다.

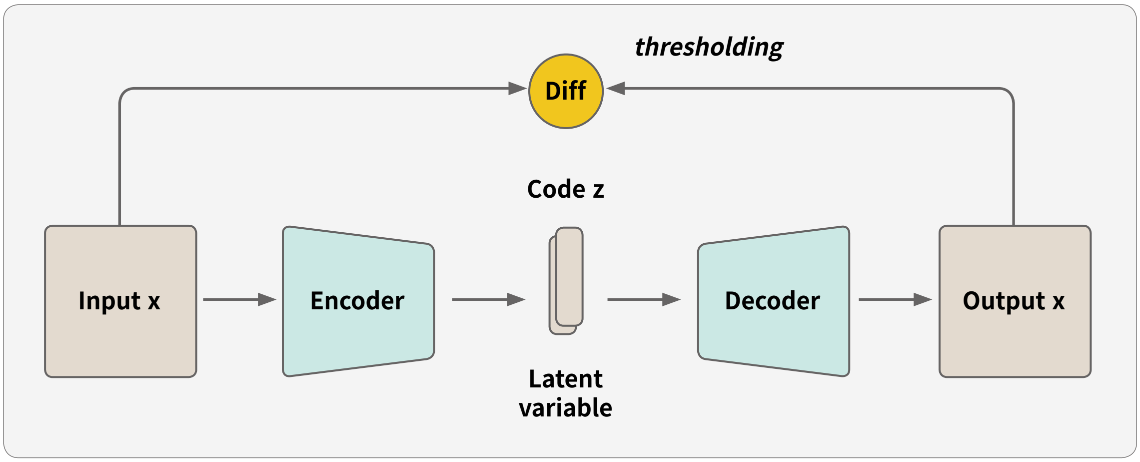
(2) 예측 기반 판별 방식

예측 기반 판별 방식은 미래에 발생할 데이터를 예측하는 모델을 사용한다. 이 모델에서 예측한 데이터와 실제 발생한 데이터가 일정 임계값을 초과하는 경우를 이상으로 식별한다. 예측 기반 판별 방식은 앞으로 발생할 이벤트를 예측하기 때문에 시계열 분석에 효과적인 알고리즘을 사용하여 구현할 수 있다. 대표적인 알고리즘으로 ARIMA, Holt-Winters, FDA, HTM (Hierarchical Temporal Memory), vanilla RNN, LSTM등으로 구현할 수 있다. [그림 3]은 HTM을 사용하여 이상 탐지를 하는 과정을 나타낸다. HTM 네트워크는 입력 데이터 를 사용하여 현재 입력을 나타내는 벡터 와 예측을 출력하는 벡터 를 출력한다. [와] 를 사용하여 예측 오차 측정 값 계산하고, [의] 확률 모델을 사용하여 시스템이 비정상 상태일 가능성()을 계산한다.

(3) 재구성 기반 판별 방식

재구성 기반 판별 방식은 이상을 판별할 데이터를 저차원 형태의 잠재 구조(latent structure)를 획득하고, 그 이후 인위적으로 재구성한 데이터를 생성하기 위한 모델을 사용한다. 잠재 구조를 활용하는 이유는 이상(anomaly) 현상을 저차원으로 매핑하면 정보가 손실되어 효과적으로 재구성할 수 있기 때문이다. 이렇게 고차원의 데이터를 저차원으로 전환하는 대표적인 방법으로 주성분 분석(PCA)이 있다. 하지만 PCA는 선형 재구성으로 제한되며 상관 관계가 높고 가우스를 따르는 데이터 분포에만 적

[그림 4] 재구성 기반 이상 탐지 방법



출처 : https://hoya012.github.io/blog/anomaly-detection-overview-1/

용이 가능하다는 한계가 있다. 최근에는 딥러닝 기반기술인 Auto-Encoder(AE), Variational Auto-Encoder(VAE), LSTM 기반 인코더(encoder)-디코더(decoder) 구조를 활용한 알고리즘을 활용한 재구성 기반 이상(anomaly) 판별 방식이 많이 소개되었다.

[그림4]는 AE기반의 이상탐지 알고리즘의 개념도를 나타낸다. 원본 입력(x)과 오토인코더로 재구성한 데이터(x') 간의 차이(diff)를 계산하여 이상(anomaly) 여부를 판단한다. 오토인코더의 인코더(encoder)를 통해 저차원의 잠재 변수(latent variable)를 얻을 수 있다. 이 과정에서 원본 입력(x)이 이상(anomaly) 데이터라면 잠재 변수에 값이 손실된 값을 얻게 되어 디코딩 과정에서 제대로 복원이 어려워진다.

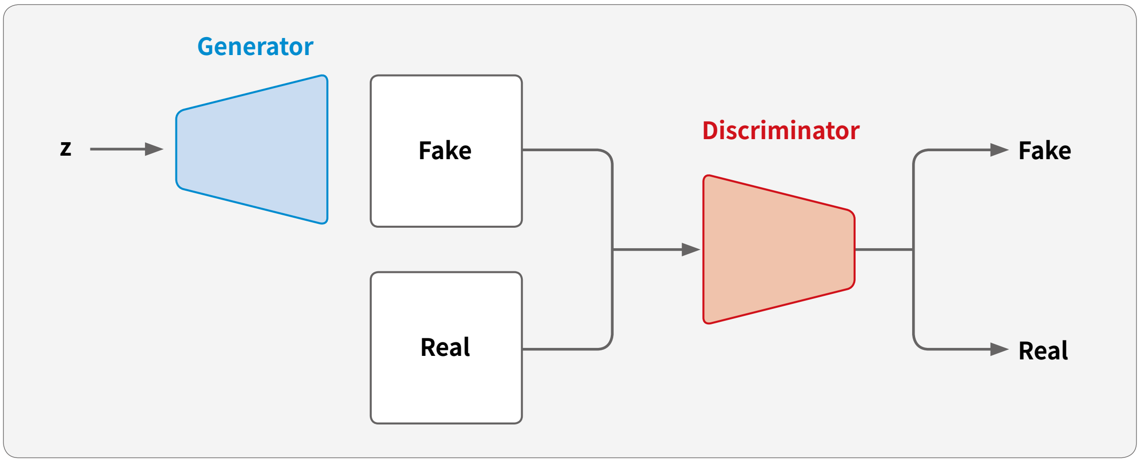
GAN 기반 이상 탐지 방법

이상(anomaly) 현상은 정상적인 현상에 비해 상대적으로 발생할 가능성이 매우 적다. 그렇기 때문에 이상 현상에 대한 특징을 정의하는 것은 쉽지 않다. 또한 예상치 못한 이상(anomaly) 현상에 대해서도 탐지할 수 있도록 대응하기 위해서는 새로운 접근 방법이 필요하다. 생성적 적대 네트워크(Generative Adversarial Networks, GAN)는 이 두 가지 문제점을 해결할 수 있는 방법이다.

(1) GAN(Generative Adversarial Networks)이란?

생성적 적대 네트워크(GAN)란 생성자(generator)와 판별자(discriminator)가 적대적으로 학습시키는 인공지능 알고리즘이다. [그림 5]는 GAN의 동작을 나타내는 개념도다. 생성자는 노이즈 벡터(z)를 사용하여 가짜 데이터(Fake)를 만들어낸다. 이해를 돕기 위해 여기서 생성하는 데이터는 이미지라고 가정하겠다. 판별자는 생성자가 만든 가짜 이미지와 진짜 이미지를 입력으로 받아 어떤 것이 진짜인지 판별하는 역할을 수행한다. GAN은 생성자가 만든 가짜 이미지가 진짜 이미지와 비슷해져서 판별자가 진위를 판별하지 못할 때까지 알고리즘을 개선하는 방식으로 학습을 진행한다. 이처럼 GAN으로 학습한 생성자는 진짜 같은 가짜 데이터를 만들어낸다. 유명

[그림 5] GAN의 개념도



출처 : https://kjhov195.github.io/2020-03-09-generative_adversarial_network/

화가의 화풍을 입힌 이미지, 유명인 목소리와 똑같은 음성 변조 파일, 소설, 시, 만화, 동영상 등 다양한 콘텐츠 생성 분야에서 활용하고 있다.

(2) GAN 기반 이미지 데이터 이상 탐지 방법

이처럼 GAN은 일반적으로 생성자를 학습시켜 새로운 데이터를 만들어내는 것을 목적으로 한다. 이상 탐지 기술에서는 GAN의 생성자가 잠재 벡터(z)를 활용하여 데이터를 잘 생성할 수 있도록 이상적인 잠재 데이터를 추출하는 것을 목적으로 한다.

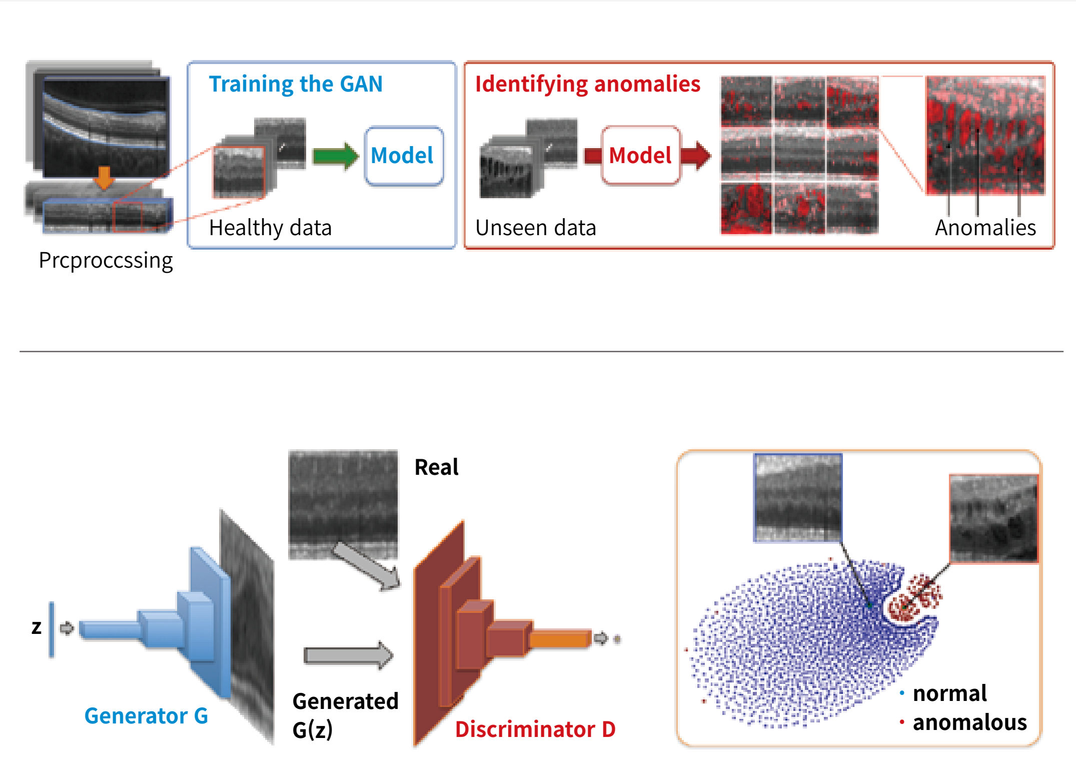
최근 이미지의 이상 탐지를 위한 GAN의 연구결과가 많이 발표되었다. 2017년 발표된 AnoGAN[1]은 의료 분야의 이미지에서 이상을 감지하는데 좋은 성능을 보이는 알고리즘이다.

1 Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery(https://arxiv.org/abs/1703.05921)

[그림6]은 AnoGAN의 동작을 나타낸다. AnoGAN은 정상 데이터의 주요 특징(z)을 입력 받아 정상 이미지를 생성하는 생성자(Generator G)를 학습시킨다. 학습이 진행될수록 생성자는 잠재 공간(latent space)에 있는 값을 사용하여 진짜 같은 가짜 정상 이미지를 생성한다. 판별자(Discrimicator D)는 진짜 정상 이미지(Real)와 생성자가 만든 가짜 정상 이미지(G(z))를 비교하여 잔차(residual)를 계산한다. 이후 생성자와 판별자의 가중치(weight)는 고정시키고, 판별자가 계산하는 잔차의 값이 최소화가 되도록 잠재 공간에 있는 벡터(z)값을 조절하는 과정을 거친다. 이 과정을 통해 기존에 알지 못했던 이상 현상(anomaly event)를 탐지할 수 있을 뿐만 아니라 잔차를 판별하는 과정에서 이상 현상이라고 판별하는 부분에 대해서도 나타낼 수 있다. AnoGAN은 정상데이터에 비해 비정상(abnormal) 데이터가 부족한 데이터 불균형 문제를 극복했을 뿐만 아니라 라벨링이 어려운 이미지 데이터를 전처리하는 데에 소요되는 시간과 리소스를 단축시킨 효과가 있다. 하지만 큰 데이터나 실시간 어플리케이션에서는 활용이 어렵다는 한계가 있다.

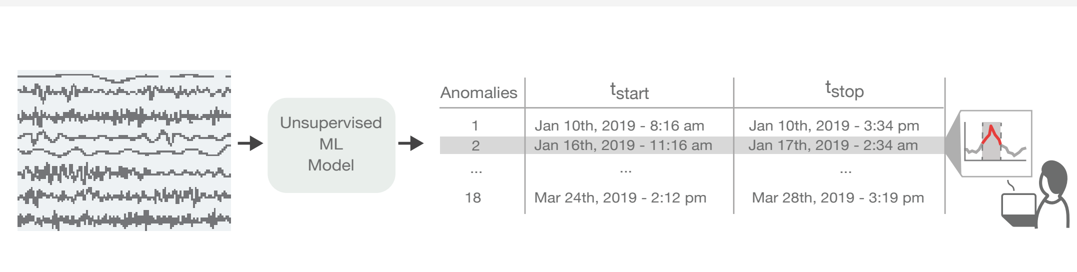
이후, 이러한 문제를 개선한 bi-directional GAN,

[그림 6] AnoGAN의 개념도



출처 : https://arxiv.org/pdf/1703.05921.pdf

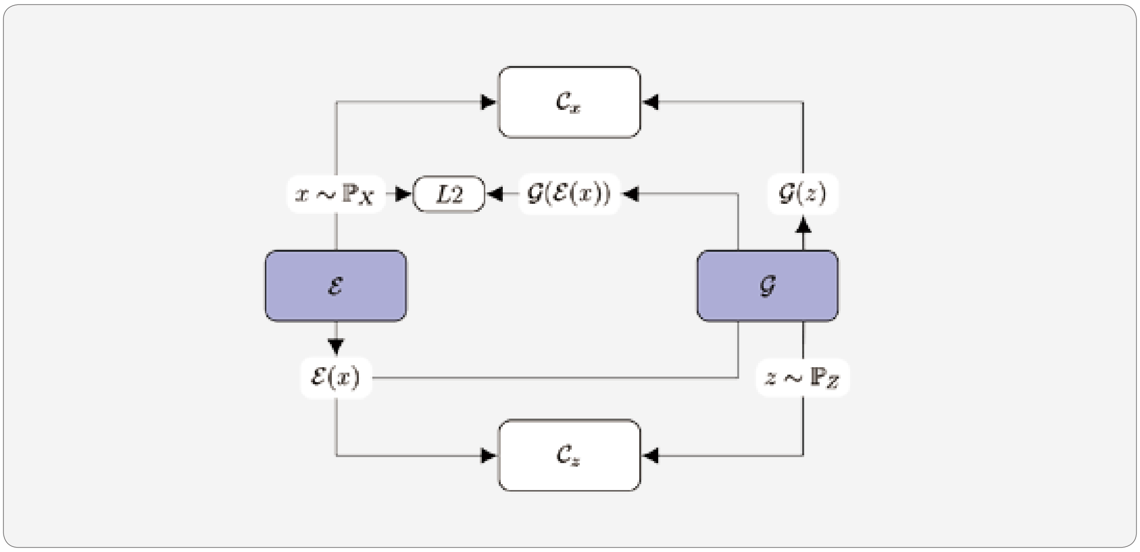
[그림 7] 시계열 데이터의 이상 탐지 사례



Anomalies	t_{start}	t_{stop}
1	Jan 10th, 2019 - 8:16 am	Jan 10th, 2019 - 3:34 pm
2	Jan 16th, 2019 - 11:16 am	Jan 17th, 2019 - 2:34 am
...	...	...
18	Mar 24th, 2019 - 2:12 pm	Mar 28th, 2019 - 3:19 pm

출처 : https://arxiv.org/pdf/2009.07769.pdf

[그림 8] TadGAN의 구조



출처 : https://arxiv.org/pdf/2009.07769.pdf

cycle-consistent GAN, BeatGAN, TadGAN 등 이상 탐지에 GAN을 활용한 다양한 연구 결과가 지속적으로 발표되는 추세다.

(3) GAN 기반 시계열 데이터 이상 탐지 사례

최근 실시간으로 데이터를 수집하고, 처리하여 의사를 결정하는 어플리케이션이 많아짐에 따라 시계열 데이터의 이상 탐지에 대한 수요가 증가했다. 국방, 에너지, 금융, 의료 등 여러 비즈니스 영역에서 활용이 가능하기 때문에 발생할 수 있는 이상(anomaly)의 유형도 다양하다.

시계열 데이터의 이상 현상은 시간, 즉 시스템이 비정상적으로 작동하는 시점(point) 또는 기간(duration)으로 정의된다. 시점(point) 이상은 비정상적인 값에 도달한 단일 시점 또는 연속적인 일련의 시점의 집합으로 정의할 수 있다. 기간(duration) 이상은 특정 시점은 비정상적이지 않을 수 있으나, 특정 구간을 전반적으로 검토하였을 때에 이상 현상이 발견되는 경우다.

시계열을 특정 길이의 하위 시퀀스로 분할하고 클러스터링 알고리즘을 적용하여 이상 현상을 탐지한다. 하지만 대부분의 시계열 데이터의 이상 현상은 임계치를 초과하지 않고, 신호 자체가 불분명한 경우가 많아 임계치를 활용하는 것이 어렵고, 통계적인 기법을 적용하기 위해서는 많은 양의 도메인 지식을 요구하고 있어 구현하는 것이 쉽지 않다.

이처럼 시계열 데이터는 데이터 자체가 기존 이상 탐지 분석 방법을 적용하기 어렵고, 실제로 많은 이상 탐지 알고리즘이 이상 현상에 대한 탐지뿐만 아니라 오탐하는 오류를 많이 발생하고 있다. 이상 탐지 알고리즘은 이상 현상을 정상적으로 탐지하는 것도 중요하지만, 정상 현상을 이상 현상이라고 오탐하지 않는 것도 중요한 성능 지표다. 일반적으로 시계열 데이터의 이상 탐지 알고리즘은 오탐(L2 에러)을 발생하는 문제가 많이 발생한다.

2020년 발표한 TadGAN은 GAN 기반 시계열 데이터 이상 탐지 알고리즘이다. [그림 8]은 TadGAN의 구조를

나타냅니다. TadGAN은 두 개의 생성자 ε와 g, 두 개의 판별자 Cx와 Cz가 있다. 생성자 ε는 시계열 시퀀스를 잠재 변수에 매핑하는 인코더 역할을 하고, 생성자 g는 잠재 변수를 재구성된 시계열 시퀀스로 변환하는 디코더 역할을 한다. 판별자 Cx는 X의 실시간 시계열 시퀀스와 g(z)에서 생성된 시계열 시퀀스를 판별하고, 판별자 Cz는 잠재 공간에 대한 매핑의 정도를 판별한다.

생성자	ε(인코더)	시계열 시퀀스(x) → 잠재 변수(z)
	g(디코더)	잠재 변수(z) → 시계열 시퀀스(g(z))
판별자	Cx	시계열 시퀀스(x)와 g(z)의 판별
	Cz	잠재 변수(z)의 매핑 정도

동적인 시간 래퍼(DTW, Dynamic time wrapping), 시점 기반 차이, 오탐(L2 norm)을 고려한 점수 등 기존 시계열 데이터 이상 탐지 알고리즘의 한계를 극복하기 위한 이상 탐지 점수(anomaly score) 계산 방법을 사용하였다.

마무리

지금까지 다양한 이상 탐지 방법에 대해서 살펴보았다. 지금 이 시점에도 수많은 곳에서 이상 현상은 발생하고 있을 것이다. 아무리 주의를 기울여서 시스템을 설계하고, 소프트웨어를 만든다고 하더라도 이상 현상은 발생할 수밖에 없다. 실시간으로 기존에 발생하지 않은 비정상 상황까지 탐지할 수 있는 알고리즘이 나왔다는 것은 매우 고무적인 일이라고 생각한다. 물론 이러한 알고리즘을 실제 시스템에 적용할 때에 예상치 못한 또 다른 문제점이 발견될 것이다.

현재의 IT 시스템이 진일보하기 위해서는 지속적으로 이상 현상을 탐지하고, 이를 대응하기 위한 알고리즘을 연구해야 한다. 그리하여 악의적으로 혹은 예측치 못한 시스템의 동작으로 인해 이상 현상을 야기시키는 알고리즘에 대응하여 더 강해지고 살아남는 것이 필요하다. 이것이 바로 이기적인 소프트웨어 유전자가 진화하는 방향이다.

참고문헌

Anomaly detection using forecasting methods arima and hwds (E. H. Pena, et al., 2013)

Lof: identifying density-based local outliers (M. M. Breunig, et al., 2000)

Variational autoencoder based anomaly detection using reconstruction probability (J. An, et al., 2015)

Unsupervised anomaly detection with generative adversarial networks to guide marker discovery (T. Schlegl et al., 2017)

TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks (A. Geiger, et al., 2020)