

# Conditional Generative Adversarial Nets, CGAN (2014) 논문리뷰

changdaeoh · 2021년 6월 25일

❤ 0

CGAN

Deep Learning

Generative Model

gan

기계학습

딥러닝

생성모델

## Paper Review



▼ 목록 보기

7/25



# Paper Review

기본 GAN에 추가적인 정보를 함께 주입하여 생성 결과물을 통제할 수 있는 연구에 관한 논문이다.

[original paper](#)

분량이 매우 적은편에 속하며 내용도 직관적이라 GAN에 대한 사전지식이 있으면 가볍게 읽으실 수 있을 듯.

## Contents

- 0. Abstract
- 1. Introduction
- 2. Related Work
- 3. Conditional Adversarial Nets
- 4. Experimental Results

Future work는 생략

# 0. Abstract

본 연구에서는 2014년 제안된 참신한 생성모델 **GAN의 조건부 버전을 다룬다**. 이 모델은 단순히 데이터와 보조변수  $y$ 를 함께discriminator와 generator 둘 모두에게 전달함으로써 구축할 수 있으며, 조건부 이미지생성은 물론 이미지태깅, multi-modal model의 학습에 이용될 수 있다.

# 1. Introduction

intractable한 확률론적 계산의 근사에 대한 어려움을 피해갈 수 있는 GAN은 Markov chain이 요구되지 않으며 backpropagation을 통해 end to end로 학습이 가능하며, 다양한 요소들이 모델에 쉽게 통합될 수 있다는 장점이 있다. 생성하는 이미지의 품질 또한 기존의 생성모델들에 비해 훨씬 real한 모습을 모였다.

GAN을 다른 부가적인 정보를 이용하여 conditioning하면 **데이터 생성 과정(data generating process)를 원하는대로 통제**할 수 있다.

부가 정보로써 활용할 수 있는 변수는 다양한데, 쉽게 생각할 수 있는 것은 class label이며, 다른 modality로부터의 데이터/특성 등도 활용이 가능하다!

## 2. Related Work

### 2.1 multi-modal learning for image labeling

현대에 supervised neural network들이 크게 성공을 이루었지만, (1) 예측타겟 범주 수가 매우 많은 경우로의 확장이 어려우며, (2) 일대일 매핑을 학습시키는 것은 쉬우나 일대다의 매핑(이미지 레이블링 등)은 어려운 문제이다.

(1)을 해결하기 위해 도움이 될 수 있는 방법은 **다른 modalities로부터 추가적인 정보를 가져다 쓰는 것**이다. 이를 통해 본래의 목적 task에서 더 좋은 performance를 달성할 수 있으며, 일반화 측면에서도 더 나아질 수 있다고 한다.

(2)를 해결하기 위한 방법으로는 **조건부 생성모델을 사용하는 것**이다. 이 경우 입력은 조건부 변수로 간주되며, 일대다 매핑의 학습은 조건부 분포를 예측하는게 된다.

## 3. Conditional Adversarial Nets

### 3.1 Generative Adversarial Nets

오리지널 GAN에 대한 remind. <생략>

궁금하신분들은 [GAN리뷰](#)를 참고해주세요!

일반 GAN의 목적함수는 아래와 같습니다.

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

### 3.2 Conditional Adversarial Nets

GAN의 discriminator와 generator가 모두 어떤 추가적인 정보 y에 의해 컨디셔닝된다면 GAN은 조건부 모델로 확장될 수 있다. 이때 추가적인 정보로 간주되는 y는 어떠한 종류의 auxiliary information이라도 될 수 있다. 예를 들어 class label이나 다른 모달리티로부터의 데이터가 될 수 있겠다.

저자들은 이 보조변수  $y$ 를 discriminator와 generator 각각에 별도의 input layer를 두어 추가적으로 주입하였으며, 이 두 layer의 output이 하나의 layer에 전달되어 **original input과 auxiliary variable의 joint hidden representation**을 학습하도록 한다!!

참고로 이 joint hidden representation이 구성되는 방식은 자유도가 매우 높아 여러 변형을 시도해볼 수 있겠다. (일반 input과 조건이되는 변수의 representation을 각각 따로 학습시켜 concatenate하던지 원소별 덧셈/곱셈을 하든지 등등도 가능할 듯!)

조건부 GAN의 train objective는 아래와 같다.

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})))]$$

그냥 일반 GAN의 목적함수에 input 변수  $x$ ,  $z$ 만 조건부 변수  $x|y$ ,  $z|y$ 로 바뀐 모습이다.

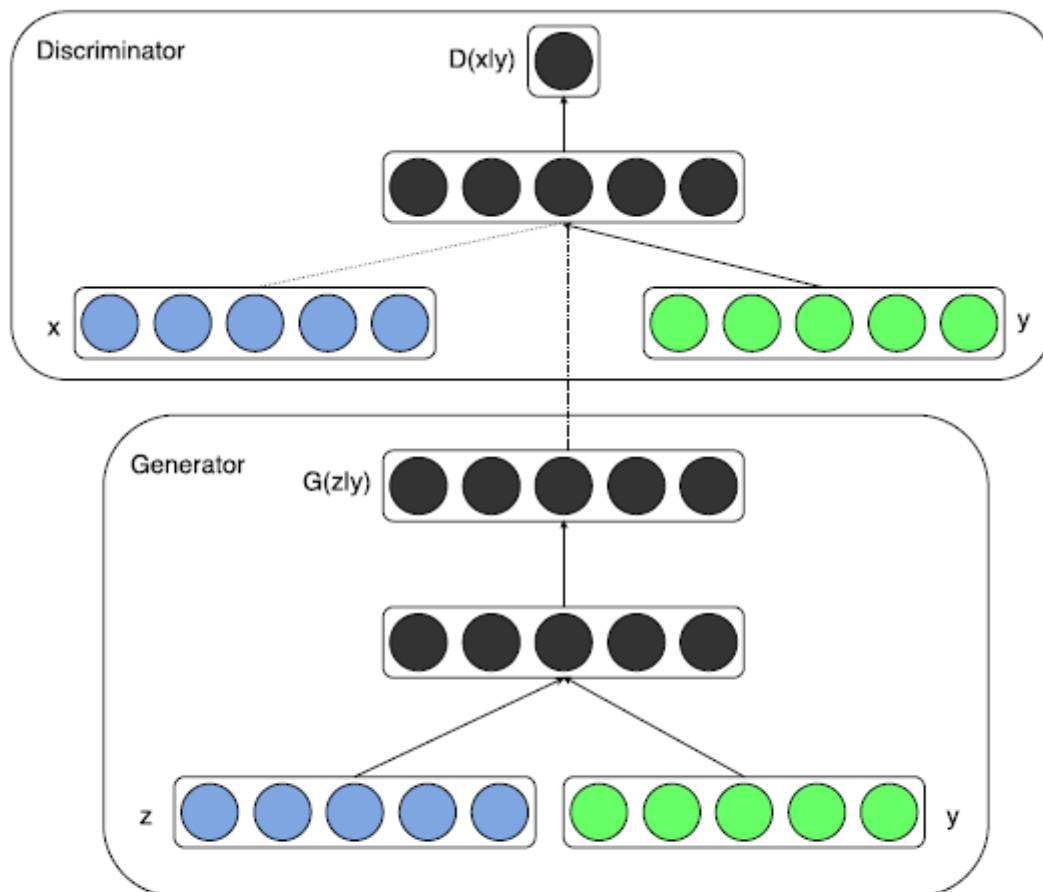


Figure 1: Conditional adversarial net

## 4. Experimental Results

### 4.1 Unimodal

CGAN에 대한 Unimodal case의 실험으로 MNIST 데이터셋의 숫자 label을 보조정보 변수  $y$ 로써 one-hot 인코딩하여 주입한다.

generator에서는 noise prior로부터의 잠재변수  $z$ 와  $y$ 를 각각 서로다른 hidden layer로 전달하고 그 이후 하나의 레이어에 함께 전달한다. 최종적으로  $28 \times 28$ 차원 MNIST 샘플을 생성한다. 활성화 함수로 ReLU를 사용하며 출력층에서는 sigmoid를 사용한다.

Discriminator도 마찬가지로  $x$ 와  $y$ 를 각각 다른 hidden layer에 우선 전달 한 뒤 이후에 하나의 레이어에 함께 전달된다. 이후 최종 아웃풋 층은 sigmoid 결과값을 반환하는 하나의 유닛이다. output layer이전에는 활성화함수로 maxout을 사용한다.

SGD with Momentum 옵티마이저를 사용하며 LR decay, Dropout 기법등을 함께 활용한다.

Model	MNIST
DBN [1]	$138 \pm 2$
Stacked CAE [1]	$121 \pm 1.6$
Deep GSN [2]	$214 \pm 1.1$
Adversarial nets	$225 \pm 2$
Conditional adversarial nets	$132 \pm 1.8$

**Table 1:** Parzen window-based log-likelihood estimates for MNIST. We followed the same procedure as [8] for computing these values.

Test set에 대한 로그가능도 추정치를 구해보면 위와 같은데, CGAN가 그리막 우수한 건 아닌것처럼 보인다. 저자들은 이 지표를 효과의 입증보다는 개념의 증명을 위해 사용하는바라고 밝혔으며, 하이퍼파라미터 최적화와 구조에 대한 더 심도있는 탐색을 진행하면 위 지표에 대한 스코어를 더 향상시킬 수 있을거라 기대한다고함.

예전 논문들을 보면 **생성모델의 평가지표로 테스트 셋에 대한 근사된 로그가능도를 이용하는 경우가 많은데**, 좋은 척도라고 확신할 순 없지만 그렇다고 다른 좋은 metric이 있는것도 아니라서 그냥 쓰는 정도라고 들은 것 같다.

요 강의에서(CMU 11785 lec23)

근사 로그가능도 기반 evaluation의 아이디어를 내가 이해한대로 설명하자면,

- 먼저 train set으로 훈련된 generative model을 이용해 충분히 많은 샘플들을 생성한다.
- 생성된 샘플들에 대해 Gaussian Parzen window(커널밀도추정이 더 익숙한 이름임.)방법을 이용해 density function을 추정할 수 있다.
- 추정된 분포는  $p_{model}$ 가 되며 이 값이 실제 데이터의 분포  $p_{data}$ 와 최대한 비슷할수록 생성모델이 잘 학습된 것이다.
- test set의 데이터 샘플들을 이 추정된 모델에 의한 분포  $p_{model}(x)$ 에 하나씩 집어넣고 로그를 취해 모두 더한다.
- 결과값은  $\sum \log p_{model}(x_{test})$ 이며 이는 테스트셋에 대한 로그가능도 추정치가 된다.  
(이 값이 높을수록 generator가 실제 데이터(testset)에 가까운 데이터를 생성한 것으로 간주할 수 있다.)

레이블을 주어 조건부로 생성한 이미지들은 아래와 같다.



Figure 2: Generated MNIST digits, each row conditioned on one label

## 4.2 Multimodal

노이즈 변수  $z$ 에 더하여 이미지에 관한 feature들을 보조변수로서 주입하여 tag-vector를 생성하는 실험을 진행함. 즉, input되는 정보는 image 도메인인데, 출력은 word token이라는 점에서 Multimodal learning이라고 볼 수 있음.

MIR Flickr 25000 데이터셋을 활용했다고 하며, 이 데이터셋은 이미지 + UGM(user-generated metadata)를 포함하는 데이터셋인 듯 함.

이미지와 tag의 feature를 extract하기 위해 convolutional model과 language model을 함께 사용했다고 하며, ImageNet에 pretrain된 CNN에 MIR Flickr의 이미지를 흘려보내 출력이전의 마지막 fc layer로부터 output된 feature를 generator에 input되는 보조변수로 활용하였다고 한다.

language model은 MIR Flickr의 tag 들에 대해 feature를 extract하는데 사용했다고 하는데, tag에 관한 feature들은 어디에 사용되었는지 논문을 봐도 모르겠다.. tag feautre도 image feature이랑 같이 generator/discriminator에 input되는 건가..? 아시는분 댓글좀  $\pi\pi\pi$

아무튼 어찌어찌 학습된 generator에  $z + \text{image feature vector}$ 를 주입하여 생성한 이미지 태그들 중 cosine 유사도가 높은 애들 몇개를 추린 결과는 아래와 같다.





	User tags + annotations	Generated tags
	montanha, trem, inverno, frio, people, male, plant life, tree, structures, transport, car	taxi, passenger, line, transportation, railway station, passengers, railways, signals, rail, rails
	food, raspberry, delicious, homemade	chicken, fattening, cooked, peanut, cream, cookie, house made, bread, biscuit, bakes
	water, river	creek, lake, along, near, river, rocky, treeline, valley, woods, waters
	people, portrait, female, baby, indoor	love, people, posing, girl, young, strangers, pretty, women, happy, life

Table 2: Samples of generated tags



본 논문의 결과는 매우 preliminary한 것이라 저자들도 인정하고있음. 그러나 Conditional GAN의 엄청난 잠재력을 보였으며, 다양한 application으로의 확장가능성도 보여주었음.

일단 구현 자체가 너무 간단한데, 생성되는 데이터를 원하는 대로 통제할 수 있다는 점에서 의미있는 연구였다는 생각이 듬. 근데 논문에 오타 2개정도 있는거 같은데 나만느끼나..? 논문에 오타가 있을수도 있나요..??

후속 연구들로 흥미있는 논문들이 굉장히 많이나와 뭐부터 읽어봐야할지 행복한 고민중에 있음 ㅎㅎ...



오창대

말보다는 행동



다음 포스트

InfoGAN (2016) 논문리뷰



이전 포스트

Auto-Encoding Variational Bayes (VAE) 논문리뷰 (2)

관심 있을 만한 포스트

생성모델 세미나  
Tobig's 13기 이재지

Disentanglement



Disentanglement

생성모델 세미나  
Tobig's 14기 김상현

생성모델 세미나  
Tobig's 12기 김태한



GAN의 평가와 편향  
evaluation and bias

VAE  
[Variational AutoEncoder]

GAN의 평가와 편향

Variational Autoencoder

kaggle™

딥러닝 개념 1. VAE(Variational Auto Encoder)

0개의 댓글

댓글을 작성하세요

댓글 작성