

PROJECT/논문읽기  
[논문읽기] 10. CGAN

2019. 6. 30. 02:12

리틀코리아 아동도서대여

도서대여, 전집, 단행본, 영어책 대여물, 10~365일 대여가능, 전국 인터넷 대여.

<http://www.littlekorea.kr>



Let's play with 코드모스

코딩교육의 핵심은 스스로 문제를  
분석하고 해결하는 능력을 기르는 것  
코드모스를 만나보세요

코드모스

# < Conditional Generative Adversarial Net S >

*"GAN + Condition y"*

## 0. Abstract

이 논문에서는 conditional 버전의 GAN에 대해서 설명할 것이다. 이는 간단하게  $y$  데이터를 추가함으로써 generator와 discriminator에 condition을 줄 수 있다. 이 모델을 통해서 MNIST dataset을 클래스 라벨 조건에 맞춰서 생성할 수 있다.



Unconditioned generative model에서는 데이터가 생성되는데에 통제권이 없다. 하지만 추가적인 정보를 모델에 입력하여 conditioning하면 데이터 생성과정에 영향을 미칠 수 있다. 이러한 conditioning은 클래스 라벨, 인페인팅 데이터의 일부 부분 또는 심지어 다른 양식의 데이터에 기반할 수 있다.

클래스 라벨에 conditioned된 MNIST 데이터 세트와 다중모드 학습을 위한 MIR Flickr 데이터셋에 실험을 할 것이다.

## 2. Related Work

### 3.1 Multi-modal Learning For Image Labelling

여전히 매우 많은 예측 결과 카테고리를 수용하기 위해 모델을 확장하는 것은 어려운 일이다.

두 번째 문제는 지금까지 대부분의 작업이 입력에서 출력으로의 일대일 매핑을 학습하는데 초점을 맞췄다는 것이다.

하지만 많은 흥미로운 문제들은 자연적으로 확률론적 일대다 매핑으로 생각되어진다. 예를 들어, 이미지 라벨링의 경우 주어진 이미지에 적절하게 사용될 수 있는 서로 다른 다양한 태그들이 있으며, 다른 사람들이 동일한 이미지를 설명하기 위해 다른 용어를 사용할 수 있다.

첫 번째 문제를 해결하는데 도움이 되는 한 가지 방법은 다른 유형의 추가 정보를 활용하는 것이다. 예를 들어, 기하학적 관계가 의미론적으로 의미 있는 라벨에 대한 벡터 표현을 학습하기 위해 자연어 말뭉치를 사용하는 것과 같다. 그러한 공간에서 예측을 할 때, 예측 오류가 있을 때 종종 실제와 '가까워' 있다는 사실에서 이익을 얻는다.

Deep Visual-Semantic Embedding Model에서는 이미지 특징 공간으로부터 단어 표현 공간까지의 간단한 선형 매핑으로도 분류 성능을 향상시킬 수 있다고 보여졌다.

두 번째 문제를 해결하기 위한 방법으로는 conditional probability generative model을 사용하는 것이다. 입력은 conditioning 변수로 간주되고 일대다 매핑은 conditional predictive distribution으로 인스턴스화된다.

## 3. Conditional Adversarial Nets

### 3.1 Generative Adversarial Nets

GAN은 generative model을 학습하기 위한 방법으로 최근에 소개되어졌다. generator  $G$ 는 분포를 묘사하려고 하고 discriminator  $D$ 는 샘플이  $G$ 에서보다 실제 데이터로부터 왔을 확률을 추정한다.  $G$ 와  $D$ 는 비선형 매핑 함수가 될 수 있다.

데이터  $x$ 에 대한 generator 분포  $p_g$ 를 학습하기 위해, generator는 prior noise 분포  $p_z(z)$ 로부터 데이터 공간인  $G(z; \theta_g)$ 로의 매핑 함수를 생성한다. Discriminator는  $D(x; \theta_d)$ 로, 결과는 single scalar로  $x$ 가  $p_g$ 가 아닌 실제 학습 데이터로부터 왔을 확률을 나타낸다.

$G$ 와  $D$ 는 모두 동시에 학습된다:  $G$ 에 대한 파라미터를 조정하여  $\log(1 - D(G(z)))$ 를 최소화하고  $D$ 에 대한 파라미터를 조정하여  $\log D(x)$ 를 최소화하며, 다음의  $V(G, D)$ 와 같은 minimax-game을 한다.

## 3.2 Conditional Adversarial Nets

GAN은 generator & discriminator 모두 추가적인 정보  $y$ 에 의해 conditioned된다면 conditional model로 확장될 수 있다.  $y$ 는 어떠한 종류의 보조적인 정보일 수 있고, 클래스 라벨이나 다른 유형으로부터의 데이터도 가능하다. discriminator와 generator에 모두 추가적인 입력 layer로서  $y$ 를 추가하면 conditioning을 할 수 있다.

Generator에서 prior input noise인  $p_z(z)$ 와  $y$ 를 합동 은닉 표현(joint hidden representation)으로 결합되고, 적대적 학습 프레임워크는 이러한 은닉 표현이 구성되는 방식에 상당한 유연성을 허용한다.

Discriminator에서  $x$ 와  $y$ 는 입력과 구분 함수로 표시된다.

목적 함수는 다음과 같다.

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

Fig.1은 간단한 CGAN의 구조를 보여준다.

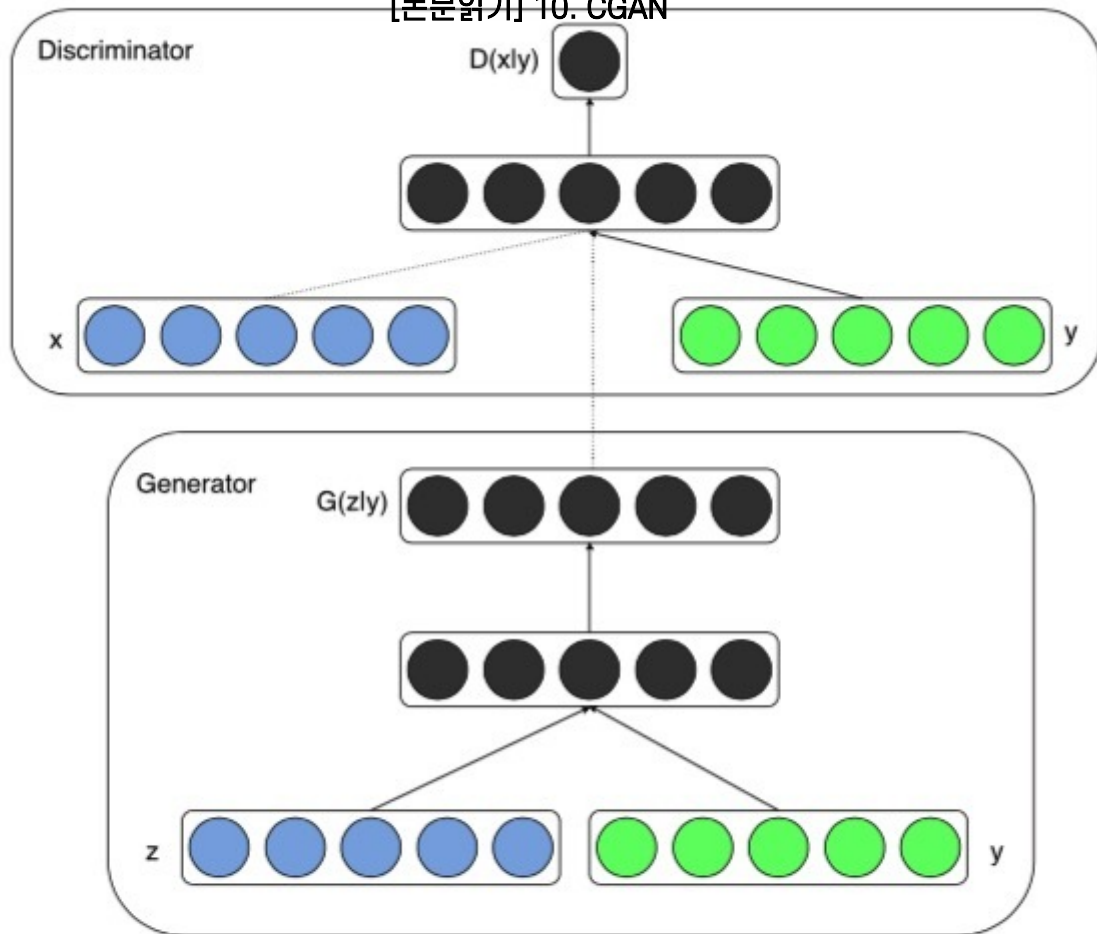


Figure 1: Conditional adversarial net

## 4. Experimental Results

### 4.1 Unimodal

논문에서는 CGAN을 one-hot 벡터들로 인코딩된 클래스 라벨에 conditioned된 MNIST 이미지들에 대해 학습했다. Generator net에서는 단위 하이퍼큐브 내의 균일한 분포에서 100 차원의 noise prior  $z$ 를 도출했다.  $z$ 와  $y$ 는 ReLU 활성화 함수와 함께 은닉층으로 매핑된 후, 각각 200, 1000의 레이어 크기를 가지고, 그 다음 1200 차원성의 은닉 ReLU 레이어에 매핑된다. 그리고 마지막으로 sigmoid 레이어로 784-차원의 MNIST 샘플을 생성한다.

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.

Discriminator는  $x$ 를 200 units, 5 pieces의 maxout layer로,  $y$ 는 50 units, 5 pieces의 maxout layer로 매핑한다. 두 은닉층은 sigmoid layer에 공급되기 전에, 240 units, 4pieces의 결합 maxout layer로 매핑한다.

SGD with mini-batch size = 100

초기 Learning\_rate = 0.1 ~ 0.000001 ( 1.00004배의 속도로 하강 )

Momentum = 초기 0.5에서 0.7로 증가

Dropout = 0.5 (G,D 모두)



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

## 4.2 Multimodal

Flickr와 같은 사진 사이트는 이미지 및 관련 사용자가 생성한 메타데이터(UGM : User Generated Metadata)의 풍부한 레이블링 데이터가 있다.

UGM의 경우 동의어가 널리 퍼져있다. 즉 사용자마다 동일한 개념을 설명하기 위한 용어를 서로 다르게 쓸 수 있다는 것이다. 결과적으로 이러한 라벨을 정상화하는 효율적인 방법을 갖는 것이 중요해진다. 개념적인 word embedding은 관련 개념들이 결국 유사한 벡터에 의해 표현되기 때문에 여기서 유용할 수 있다.



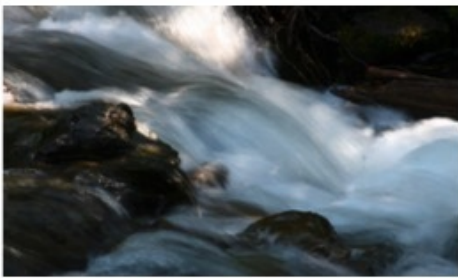
montanha, trem, inverno, frio, people, male, plant life, tree, structures, transport, car

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

모델 구조는 논문을 읽으면 알 수 있다.

## 5. Future Work

논문에서 보여진 결과는 극도로 초기 단계이지만, CGAN의 잠재력을 증명했고, 흥미롭고 유용한 적용에 대한 가능성을 보여준다.

CGAN은 condition variable  $y$ 를 추가함으로써 내가 원하는 조건을 걸어서 그에 대한 결과물을 얻는 것이다. word embedding처럼 의미있는 단어들을 벡터 표현에 맞게 생성해낼 수도 있다.

논문에서 중점은 GAN에 condition을 추가하는 것 만으로도 원하는 결과를 얻을 수 있다는 것을 보여주었다는 것이다.



## Let's play with 코드모스

코딩교육의 핵심은 스스로 문제를  
분석하고 해결하는 능력을 기르는 것  
코드모스를 만나보세요

코드모스



공감



### 'PROJECT/논문읽기' 카테고리의 다른 글



[논문읽기] 12. YOLO v1



[논문읽기] 11. WGAN-GP : Improved Training of Wasserstein GANs



[논문읽기] 09-1. LSGAN MNIST with Keras



[논문읽기] 09. LSGAN



### 개발차



이차민의 iOS/ML/DL 공부 블로그 iOS개발과 Computer Vision에 대한 전반적인 공부를 합니다. "Be The First Penguin"



이름

암호



< 1 ... 311 312 313 314 315 316 317 318 319 ... 532 >