(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0030063
(43) 공개일자 2021년03월17일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06F 16/55* (2019.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06F 16/55 (2019.01)
(21) 출원번호 10-2019-0111542
(22) 출원일자 2019년09월09일
심사청구일자 없음

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
양지훈
서울특별시 은평구 백련산로 2길 19, 백련산힐스테이트 1차 105동 603호
김상록
서울특별시 강남구 언주로 30길 56, 타워팰리스1차 C동 2201호
(74) 대리인
이지연

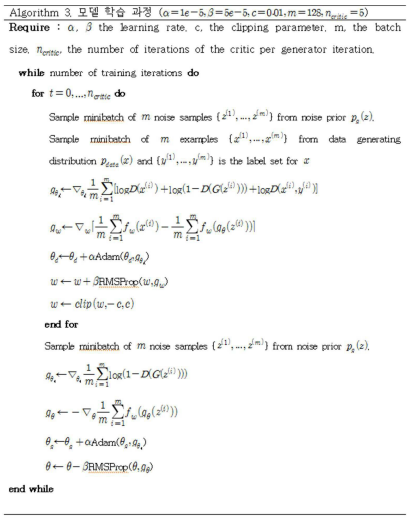
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **준지도 학습을 기반으로 한 이미지 분류를 위한 적대적 이미지 생성 모델 구축 시스템 및 방법**

(57) 요 약

본 발명은 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법에 관한 것이다. 상기 적대적 생성 모델 구축 방법은, (a) 판별자에 대하여 준지도 학습에 따른 제1 손실함수를 구하고, 상기 제1 손실함수가 최적화되도록 학습시키는 단계; (b) 판별자에 대하여 Earth Mover's 거리(EM 거리)를 최소화시키기 위한 제2 손실함수를 구하고, 상기 제2 손실함수가 최적화되도록 학습시키는 단계; (c) 생성자에 대하여 기존의 적대적 생성 모델에 따른 제3 손실 함수를 구하고, 상기 제3 손실함수가 최적화되도록 학습시키는 단계; (d) 생성자에 대하여 EM 거리를 최소화시키기 위한 제4 손실 함수를 구하고, 상기 제4 손실함수가 최적화되도록 학습시키는 단계; (e) 분류기를 이용하여 상기 판별자의 결과를 분류하여 출력값을 제공하는 단계;를 구비한다.

대 표 도 - 도7



이 발명을 지원한 국가연구개발사업

과제고유번호	2018R1D1A1B07048790
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구지원사업
연구과제명	GAN 기반 분산 딥러닝 프레임워크를 이용한 개인 정보 보호 데이터 마이닝 연구
기 여 율	1/1
과제수행기관명	서강대학교 산학협력단
연구기간	2018.06.01 ~ 2021.05.31

명 세 서

청구범위

청구항 1

입력된 무작위 잠재 변수로부터 이미지를 생성해 내는 생성자;

상기 생성자에 의해 생성된 이미지를 입력으로 받아서 진짜인지 가짜인지 여부를 판단하는 판별자; 및

상기 판별자의 결과를 이용하여 분류하여 출력값을 제공하는 분류기;

를 구비하고, 상기 판별자는 준지도 학습에 따른 판별자의 제1 손실 함수 및 Earth Mover's 거리(EM 거리)를 최소화시키기 위한 판별자의 제2 손실함수를 각각 학습 데이터를 이용하여 최적화되도록 학습시켜 구성된 것을 특징으로 하며,

상기 생성자는 기본적인 적대적 생성 모델의 생성자의 제3 손실 함수 및 EM 거리를 최소화시키기 위한 생성자의 제4 손실 함수를 각각 최적화되도록 학습시켜 구성된 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 시스템.

청구항 2

제1항에 있어서, 상기 판별자의 학습 데이터는 레이블된 진짜 데이터, 레이블되지 않은 진짜 데이터 및 생성자가 생성한 가짜 데이터인 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 시스템.

청구항 3

제1항에 있어서, 상기 판별자의 상기 제1 손실함수는,

지도 학습에 따른 레이블된 데이터에 의한 손실 함수와 비지도 학습에 따른 레이블되지 않은 데이터에 의한 손실 함수의 합으로 이루어진 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 시스템.

청구항 4

제1항에 있어서, 상기 판별자는 복수 층의 합성곱 신경망으로 구성되고,

상기 생성자는 복수 층의 합성곱 신경망으로 구성된 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 시스템.

청구항 5

제1항에 있어서, 상기 분류기는 소프트맥스(Softmax) 함수로 구성되어, 판별자의 결과에 대하여 클래스의 종류 및 가짜 유무를 분류하여 출력하도록 구성된 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 시스템.

청구항 6

무작위 잠재 변수로부터 이미지를 생성해 내는 생성자와 상기 생성자에 의해 생성된 이미지를 입력으로 받아서 진짜인지 가짜인지 여부를 판단하는 판별자를 구비하는 적대적 생성 모델 구축 방법에 있어서,

(a) 판별자에 대하여 준지도 학습에 따른 제1 손실함수를 구하고, 상기 제1 손실함수가 최적화되도록 학습시키는 단계;

(b) 판별자에 대하여 Earth Mover's 거리(EM 거리)를 최소화시키기 위한 제2 손실함수를 구하고, 상기 제2 손실함수가 최적화되도록 학습시키는 단계;

(c) 생성자에 대하여 기존의 적대적 생성 모델에 따른 제3 손실 함수를 구하고, 상기 제3 손실함수가 최적화되도록 학습시키는 단계;

(d) 생성자에 대하여 EM 거리를 최소화시키기 위한 제4 손실 함수를 구하고, 상기 제4 손실함수가 최적화되도록

학습시키는 단계;

(e) 분류기를 이용하여 상기 판별자의 결과를 분류하여 출력값을 제공하는 단계;

를 구비하는 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

청구항 7

제6항에 있어서, 상기 (a) 단계와 (b) 단계는

상기 판별자는 레이블된 진짜 데이터, 레이블되지 않은 진짜 데이터 및 생성자가 생성한 가짜 데이터가 입력되고, 판별자는 입력된 데이터들을 이용하여 학습하는 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

청구항 8

제6항에 있어서, 상기 (a) 단계는

상기 제1 손실함수는 레이블된 데이터에 의한 손실 함수와 레이블되지 않은 데이터에 의한 손실 함수의 합으로 이루어진 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

청구항 9

제6항에 있어서,

상기 판별자 및 상기 생성자는 복수 층의 합성곱 신경망으로 구성된 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

청구항 10

제6항에 있어서, 상기 (e) 단계는,

판별자의 결과를 분류하는 분류기는 소프트맥스 함수로 구성되어, 클래스의 종류 및 가짜 유무를 분류하여 출력하도록 구성된 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

청구항 11

제6항에 있어서, 판별자에 대한 상기 (a) 단계와 (b) 단계의 학습은 복수회 수행되고, 생성자에 대한 상기 (c) 단계와 (d) 단계의 학습은 1회 수행되는 과정이 반복되는 것을 특징으로 하는 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 적대적 이미지 생성 모델 구축 시스템 및 방법에 관한 것으로서, 더욱 구체적으로는 준지도 학습을 기반으로 한 적대적 이미지 생성 모델(Semi-supervised GAN; 'SGAN')과 Wassersteing GAN(WGAN) 모델을 결합하여 구성한 이미지 분류를 위한 향상된 적대적 이미지 생성 모델 구축 방법에 관한 것이다.

배 경 기 술

[0002] 일반적으로 딥러닝 기반의 분류 알고리즘들은 일정량의 훈련 데이터들을 사용하여 모델을 학습시키고, 이를 바탕으로 시험 데이터가 어떤 종류의 이미지인지 분류한다. 기계학습은 지도 학습, 비지도 학습, 준지도 학습의 3가지 유형으로 구분된다. 지도 학습(Supervised Learning)은 훈련 데이터로부터 하나의 함수를 유추해 내는 기계학습으로서, 훈련 데이터는 입력값과 이에 대한 목표값 또는 정답으로 이루어지게 된다. 비지도 학습(Unsupervised Learning)은 데이터가 어떻게 구성되었는지를 알아내는 문제의 범주에 속하는 것으로서, 훈련 데이터의 입력값에 대한 목표값이나 정답이 포함되어 있지 않는다. 한편, 준지도 학습(Semi-supervised Learning)은 기계 학습의 한 범주로 목표값이 표시된 데이터와 표시되지 않은 데이터를 모두 훈련에 사용하게 된다.

[0003] 이들 중 지도학습은 레이블된 데이터의 양이 충분히 많은 더 좋은 결과를 기대할 수 있다. 하지만, 최근 정보

기술의 급격한 발전에 따라 예전과는 비교할 수 없는 막대한 양의 데이터가 축적되기는 하나, 그 중에서 지도 학습을 위한 레이블된 데이터는 매우 작은 비율을 차지할 수 밖에 없다. 왜냐하면 레이블을 추가하기 위해서는 사람의 노동이나 추가적인 자원이 필요하기 때문이다.

[0004] 이러한 문제점을 해결하기 위한 방법 중 하나로 비지도 학습을 들 수 있다. 비지도 학습은 레이블되지 않은 데이터를 사용하기 때문에 무수히 많은 데이터들을 사용할 수 있다는 장점이 있다. 하지만, 학습 방법이 명확하게 제시되지 않고 종류가 다양하며 불안정성을 갖고 있는 단점이 있다. 현재 연구되고 있는 대부분의 생성 모델(Generative Model)은 비지도 학습을 사용하여 잠재 변수 z 가 주어졌을 때의 조건부 확률 p(x|z) 을 구함으로써 데이터를 생성한다.

[0005] 생성 모델은 주어진 확률 분포를 바탕으로 하여 데이터를 생성하는 모델로서, 입력 데이터를 생성하기 위하여 필요한 것은 수학식 1과 같은 결합 분포를 알아내는 것이다.

수학식 1

[0006] $$p(x,z)=p(x)p(z|x)$$

[0007] 이때, z를 잠재 변수(Latent Vector)라고 하며, 데이터의 확률 분포인 p(x)를 알아내는 것이 생성 모델의 목표이다. 따라서, 파라미터 값을 변화시키며 확률 분포의 최대 우도 추정을 반복함으로써 p(x)를 추정한다.

[0008] 전술한 생성 모델 중 대표적인 모델이 적대적 생성 신경망(Generative Adversarial Network; 'GAN')이며, GAN은 잠재 변수 z를 입력 데이터의 특징이라고 했을 때, 생성된 데이터가 원본과 가장 비슷하게 생성될 수 있도록 z를 변화시키게 된다. 전술한 GAN은 생성자(Generator)와 판별자(Discriminator)의 역할을 하는 두 개의 모델로 구성된다. 판별자는 실제 데이터만을 참으로 판단하도록 학습하고, 생성자는 판별자가 거짓으로 판별하지 못하도록 가짜 데이터를 생성하도록 학습한다. 이러한 적대적 경쟁을 통해 두 모델의 성능을 동시에 높임으로써 최적점에 도달하겠다는 것이 GAN의 핵심 아이디어이다. 그러나, 적대적 생성 신경망 역시 학습이 어렵고 불안정하며 생성된 이미지가 판별자가 구별하기 어려워하는 특정 이미지에 치우치게 되는 Mode Collapsing이 발생하기 쉽다는 문제점을 가지고 있다. 이러한 문제점들을 해결하기 위하여 다양한 모델들이 제안되고 있는 추세이다.

[0009] 한편, 지도 학습과 비지도 학습의 문제점들을 보완하기 위한 중간 단계의 학습 방법을 준지도 학습(Semi-Supervised Learning; 'SSL')이라고 한다. 앞서 설명한 바와 같이, 준지도 학습은 레이블된 데이터와 레이블되지 않은 데이터를 같이 사용하여 학습하도록 모델을 구성한다. 그리고, 이를 통해 분류(Classification)나 회귀분석(Regression) 등의 지도 학습 문제를 더 쉽고 빠르게 해결하고자 하는 추세이다.

[0010] 따라서, 본 발명에서는 적대적 생성 모델들을 결합하여 보다 향상된 성능을 갖는 준지도 학습을 기반으로 한 적대적 생성 모델을 구축할 수 있는 시스템 및 방법을 제안하고자 한다.

선행기술문헌

특허문헌

[0011] (특허문헌 0001) 한국등록특허공보 제 10-1925913호

(특허문헌 0002) 한국공개특허공보 제 10-2019-0078710호

발명의 내용

해결하려는 과제

[0012] 전술한 문제점을 해결하기 위한 본 발명은 준지도 학습을 기반으로 하여 적대적 생성 모델들을 결합함으로써, 성능이 향상된 새로운 형태의 적대적 생성 모델을 구축할 수 있는 시스템 및 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0013] 전술한 기술적 과제를 달성하기 위한 본 발명의 제1 특징에 따른 적대적 생성 모델 구축 시스템은, 입력된 무작위 잠재 변수로부터 이미지를 생성해 내는 생성자; 상기 생성자에 의해 생성된 이미지를 입력으로 받아서 진짜인지 가짜인지 여부를 판단하는 판별자; 및 상기 판별자의 결과를 이용하여 분류하여 출력값을 제공하는 분류기; 를 구비하고,

[0014] 상기 판별자는 준지도 학습에 따른 판별자의 제1 손실 함수 및 Earth Mover's 거리(EM 거리)를 최소화시키기 위한 판별자의 제2 손실함수를 각각 학습 데이터를 이용하여 최적화되도록 학습시켜 구성된 것을 특징으로 하며, 상기 생성자는 기본적인 적대적 생성 모델의 생성자의 제3 손실 함수 및 EM 거리를 최소화시키기 위한 생성자의 제4 손실 함수를 각각 최적화되도록 학습시켜 구성된다.

[0015] 전술한 제1 특징에 따른 적대적 생성 모델 구축 시스템에 있어서, 상기 판별자의 학습 데이터는 레이블된 진짜 데이터, 레이블되지 않은 진짜 데이터 및 생성자가 생성한 가짜 데이터인 것이 바람직하다.

[0016] 전술한 제1 특징에 따른 적대적 생성 모델 구축 시스템에 있어서, 상기 판별자의 상기 제1 손실함수는, 지도 학습에 따른 레이블된 데이터에 의한 손실 함수와 비지도 학습에 따른 레이블되지 않은 데이터에 의한 손실 함수의 합으로 이루어진 것이 바람직하다.

[0017] 전술한 제1 특징에 따른 적대적 생성 모델 구축 시스템에 있어서, 상기 판별자 및 상기 생성자는 복수 층의 합성곱 신경망으로 구성된 것이 바람직하다.

[0018] 전술한 제1 특징에 따른 적대적 생성 모델 구축 시스템에 있어서, 상기 분류기는 소프트맥스(Softmax) 함수로 구성되어, 판별자의 결과에 대하여 클래스의 종류 및 가짜 유무를 분류하여 출력하도록 구성된 것이 바람직하다.

[0019] 본 발명의 제2 특징에 따른 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법은, 무작위 잠재 변수로부터 이미지를 생성해 내는 생성자와 상기 생성자에 의해 생성된 이미지를 입력으로 받아서 진짜인지 가짜인지 여부를 판단하는 판별자를 구비하는 적대적 생성 모델 구축 방법에 관한 것으로서, (a) 판별자에 대하여 준지도 학습에 따른 제1 손실함수를 구하고, 상기 제1 손실함수가 최적화되도록 학습시키는 단계; (b) 판별자에 대하여 Earth Mover's 거리(EM 거리)를 최소화시키기 위한 제2 손실함수를 구하고, 상기 제2 손실함수가 최적화되도록 학습시키는 단계; (c) 생성자에 대하여 기존의 적대적 생성 모델에 따른 제3 손실 함수를 구하고, 상기 제3 손실함수가 최적화되도록 학습시키는 단계; (d) 생성자에 대하여 EM 거리를 최소화시키기 위한 제4 손실 함수를 구하고, 상기 제4 손실함수가 최적화되도록 학습시키는 단계; (e) 분류기를 이용하여 상기 판별자의 결과를 분류하여 출력값을 제공하는 단계; 를 구비한다.

[0020] 전술한 제2 특징에 따른 적대적 생성 모델 구축 방법에 있어서, 상기 (a) 단계와 (b) 단계는 상기 판별자는 레이블된 진짜 데이터, 레이블되지 않은 진짜 데이터 및 생성자가 생성한 가짜 데이터가 입력되고, 판별자는 입력된 데이터들을 이용하여 학습하는 것이 바람직하다.

[0021] 전술한 제2 특징에 따른 적대적 생성 모델 구축 방법에 있어서, 상기 (a) 단계는 상기 제1 손실함수는 레이블된 데이터에 의한 손실 함수와 레이블되지 않은 데이터에 의한 손실 함수의 합으로 이루어진 것이 바람직하다.

[0022] 전술한 제2 특징에 따른 적대적 생성 모델 구축 방법에 있어서, 상기 판별자 및 상기 생성자는 복수 층의 합성곱 신경망으로 구성된 것이 바람직하다.

[0023] 전술한 제2 특징에 따른 적대적 생성 모델 구축 방법에 있어서, 상기 (e) 단계는, 판별자의 결과를 분류하는 분류기는 소프트맥스 함수로 구성되어, 클래스의 종류 및 가짜 유무를 분류하여 출력하도록 구성된 것이 바람직하다.

[0024] 전술한 제2 특징에 따른 적대적 생성 모델 구축 방법에 있어서, 판별자에 대한 상기 (a) 단계와 (b) 단계의 학습은 복수회 수행되고, 생성자에 대한 상기 (c) 단계와 (d) 단계의 학습은 1회 수행되는 과정이 반복되는 것이 바람직하다.

발명의 효과

[0025] 전술한 구성을 갖는 본 발명에 따른 향상된 성능의 적대적 생성 모델 구축 시스템 및 방법은 SGAN과 WGAN을 결합하여 새로이 구축된 적대적 생성 모델을 제공하게 된다.

[0026] 본 발명에 따른 시스템은 준지도 학습 방식을 기반으로 함으로써, 레이블된 진짜 데이터를 사용하여 판별자의

학습을 돕기 때문에 비지도 학습 방식과 비교하여 학습 속도가 빨라지게 된다.

[0027] 또한, 본 발명에 따른 시스템은 Deep Convolution GAN의 합성곱 신경망 구조를 SGAN에 적용함으로써, SGAN에 따른 Mode Collapsing 문제를 해결할 수 있게 된다

[0028] 또한, 본 발명에 따른 시스템은 WGAN에 따른 EM 거리를 최소화하는 방식을 적용함으로써, 데이터의 크기가 증가되더라도 학습의 안정성을 향상시킬 수 있게 된다.

도면의 간단한 설명

[0029] 도 1은 기본적인 적대적 생성 모델(GAN)을 도시한 모식도이다.

도 2는 Deep Convolution GAN의 생성자의 구조를 도시한 모식도이다.

도 3은 Semi-Supervised GAN(SGAN)의 구조를 도시한 모식도이다.

도 4는 Semi-Supervised GAN(SGAN)의 학습 과정에 대한 알고리즘을 도시한 것이다.

도 5는 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 새로운 형태의 GAN 모델의 구조를 도시한 모식도이다.

도 6은 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법을 순차적으로 도시한 흐름도이다.

도 7은 본 발명의 바람직한 실시예에 따른 적대적 생성 모델 구축 방법에 있어서, GAN 결합 모델을 구축하는 알고리즘을 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 발명에 따른 새로운 형태의 적대성 생성 모델 구축 방법 및 시스템은 SGAN의 손실함수와 WGAN의 손실함수를 각각 최적화시킴으로써, 학습 속도를 향상시키고 Mode Collapsing의 문제를 해결함과 동시에 학습의 안정성도 향상시키도록 한 것을 특징으로 한다.

[0031] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 새로운 GAN 모델 시스템 및 상기 모델 구축 방법에 대하여 구체적으로 설명한다.

[0032] 도 1은 기본적인 적대적 생성 모델(GAN)을 도시한 모식도이다. 도 1을 참조하면, 기본적인 적대적 생성 모델인 Vanila GAN (VGAN)으로서, 기본적인 적대적 생성 신경망은 생성자(Generator;'G')와 판별자(Discriminator;'D')의 역할을 하는 두 개의 모델로 구성된다. 생성자는 무작위 잠재 변수 z로부터 이미지를 생성해 내는 것으로서, 생성자의 목적은 판별자를 통과한 출력이 진짜 이미지로 분류되는 것이 목적이다. 반면에, 판별자는 생성자가 만들어낸 이미지를 입력으로 받아서 이를 판별하는 역할을 하는 것으로서, 입력받은 이미지가 진짜인지 가짜인지 정확하게 구분하는 것이 목적이다. 수학식 2는 GAN의 목적함수이다.

수학식 2

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

[0033]

[0034] 여기서, V(G,D)는 변수 D와 G에 대한 minmax를 해결하는 가치 함수(value function)이며, E는 기대값으로서 $E_{x \sim p_{data}(x)}$는 $p_{data}(x)$에 속하는 x라는 데이터들의 기대값이 된다. D(x)는 x에 대한 판별기이며, G(z)는 잠재변수 z에 대한 생성기이다.

[0035] GAN의 목적 함수를 살펴 보면, D가 V(D,G)를 최대화하는 관점에서 보면 $x \sim p_{data}(x)$의 확률분포이고, x는 그 중 샘플링 데이터이다. 판별자 D는 출력이 실제 데이터가 들어오면 1에 가깝게 확률을 추정하고, 생성자 G가 만들어 낸 가짜 데이터가 들어오면 0에 가깝게 한다. 따라서, 로그를 사용했기 때문에 실제 데이터라면 최댓값인 0에 가까운 값이 나오고, 가짜 데이터라면 무한대로 발산하기 때문에 V(D,G)를 최대화하는 방향으로 학습하게

된다. 생성자 G 부분인 오른쪽 부분이 V(D,G)를 최소화하는 관점에서 보면, $z \sim p_z(z)$는 보통 정규분포로 사용하는 임의의 노이즈 분포이고 z는 노이즈 분포에서 샘플링한 것이다. 이 입력을 생성자 G에 넣어 만든 데이터를 판별자 D가 진짜로 판별하면 $\log(1-D(G(z)))$ 이기 때문에, $D(G(z))$는 1이며, D가 가짜로 판별하면 $D(G(z))$는 0 이기 때문에 $\log(1-D(G(z)))$는 0에 가까운 최댓값이 나오게 될 것이다. 따라서, G는 V(D,G)에 있어서 G는 이를 최소화하는 방향으로 D는 최대화하는 방향으로 간다.

[0036] 한편, DCGAN(Deep Convolution GAN)은 전술한 VGAN의 구조를 발전시킨 것으로서, 합성곱 신경망을 VGAN에 적용한 것으로서, GAN 모델의 기본이 된다. DCGAN은 VGAN에 아래의 5가지 방법을 적용하였는데, (1) 최대 풀링층을 없애고 합성곱을 사용하여 특징 맵의 크기를 조절하고, (2) 배치 정규화를 적용하였고, (3) 완전 연결 은닉 계층을 제거하고, (4) 생성자의 출력 활성 함수로 tanh를 사용하고 나머지 층은 relu를 사용하며, (5) 판별자의 활성함수로 leakyrelu를 사용한다.

[0037] 도 2는 Deep Convolution GAN의 생성자의 구조를 도시한 모식도이다. 도 2를 참조하면, DCGAN은 전술한 방법을 통해 DCGAN의 생성자 모델을 구성하였으며, 그 결과 VGAN에 비하여 훨씬 선명한 이미지를 생성할 수 있게 되고, 학습 과정에서 발생하는 Mode Collapsing 문제를 상당 수 해결하게 된다.

[0038] 본 발명에 따른 GAN 모델 구축 방법은, DCGAN의 판별자와 생성자 구조를 사용하는 Semi-Supervised GAN(SGAN)과 Wassertein(WGAN)의 손실 함수(Loss Function)를 결합하여 새로운 모델을 만들고, 레이블된 진짜 데이터와 생성자가 만들어낸 레이블되지 않은 가짜 데이터를 모두 사용하는 준지도 학습을 이용하여 판별자를 학습시키는 것을 특징으로 한다. 본 발명에 따른 적대적 생성 모델 구축 시스템은 사전에 학습 데이터 등이 데이터베이스에 저장되어 있는 고성능의 컴퓨터 등으로 구현될 수 있으며, 적대적 생성 모델 구축 방법은 적대적 생성 모델 구축을 위한 프로그램 등의 형태로 저장 장치등에 저장되고 상기 적대적 생성 모델 구축 시스템의 컴퓨터 또는 서버 등에 의해 실행되어 구현될 수 있다.

[0039] Semi-Supervised GAN(SGAN)은 DCGAN의 구조를 바탕으로 하는 발전된 GAN 모델이다. VGAN의 판별자가 단순히 생성자가 생성한 이미지가 진짜 혹은 가짜인지만을 구별하던 것과는 달리, SGAN은 판별자가 분류를 수행하여 입력 데이터의 클래스의 종류 및 가짜 여부를 구별할 수 있도록 한다. VGAN의 판별자는 진짜와 가짜만을 구별하므로 시그모이드 함수를 사용하여 출력하였으나, SGAN은 클래스의 종류 및 가짜 여부를 동시에 구별하여야 하므로 소프트맥스 함수를 사용하여 출력한다. 또한, VGAN의 최종 목적은 판별자가 구분할 수 없는 정교한 가짜 이미지를 생성하는 생성자를 얻는 것이므로 학습의 초점이 생성자에게 맞추어 있으나, SGAN은 클래스의 종류 및 가짜 여부를 동시에 구별할 수 있는 판별자를 학습시키는 것이 궁극적인 목표이다.

[0040] 도 3은 Semi-Supervised GAN(SGAN)의 구조를 도시한 모식도이다. 도 3을 참조하면, Semi-Supervised GAN(SGAN)는 기본적으로 준지도 학습을 통해 판별자를 학습시키는 것으로서, 잠재 변수 z를 바탕으로 하여, 생성자는 가짜 이미지를 생성하고 판별자는 이러한 가짜 이미지와 레이블된 진짜 데이터를 모두 사용하여 학습하게 된다. 즉, SGAN의 판별자는 레이블된 데이터와 레이블되지 않은 데이터를 모두 사용하는 준지도 학습을 통해 판별자를 효과적으로 학습시키고, 이를 통해 생성자 역시 경쟁을 통해 학습시킬 수 있게 된다. 생성자와 판별자는 기본적으로 DCGAN의 구조를 사용하게 된다.

[0041] SGAN에서는 판별자가 n개의 클래스를 구별함과 동시에 생성자가 만드는 이미지를 가짜로 구별해야 하므로, 총 n+1개의 클래스를 분류할 수 있어야 한다. x를 가짜 데이터라고 하면, x의 각 클래스의 확률은 수학식 3과 같이 표현될 수 있다.

수학식 3

[0042]
$$p_{model}(y=n+1|x)=\frac{\exp(l_{n+1})}{\sum_{j=1}^{n+1}\exp(l_j)}$$

[0043] 여기서, l은 logit의 약자이며, 분류기가 k+1 차원의 벡터를 다루기 때문에, 해당 vector에 포함된 원소들을

l1,l2, …, ln+1로 정의한다.

[0044] 만약 x가 레이블된 진짜 데이터라면, 클래스 중 하나로 분류되어야 하므로, 수학식 4와 같이 x의 확률을 표현할 수 있다.

수학식 4

$$p_{model}(y=i|x,i<n+1)=\frac{\exp(l_i)}{\sum_{j=1}^{n+1}\exp(l_j)}$$

[0045]

[0046] 따라서, 각각의 경우에 대한 판별자의 손실함수는 수학식 5 및 수학식 6과 같다.

수학식 5

$$L_{unsupervised}=-E_{x\sim p_{data}(x)}\log[1-p_{model}(y=n+1|x)]-E_{x\sim G}\log p_{model}(y=n+1|x)$$

[0047]

수학식 6

$$L_{supervised}=-E_{x,y\sim p_{data}(x,y)}\log p_{model}(y|x,y<n+1)$$

[0048]

[0049] 그러므로, SGAN의 판별자(D)의 최종 손실함수는 수학식 7과 같이 전술한 2개의 손실 함수를 더해줌으로써 구할 수 있다.

수학식 7

$$L_D=L_{unsupervised}+L_{supervised}$$

[0050]

[0051] 한편, SGAN의 생성자(G)의 손실 함수는 수학식 8과 같이 VGAN과 동일한 손실함수를 사용한다.

수학식 8

$$L_G=-E_{x\sim G}\log[1-p_{model}(y=n+1|x)]$$

[0052]

[0053] 위와 같이 손실함수를 설정하고 준지도 학습을 통해 판별자를 학습시킴으로서, 기존의 GAN보다 더 빠르게 판별자를 학습시키면서 좋을 결과를 얻어낼 수 있게 된다.

[0054] 도 4는 Semi-Supervised GAN(SGAN)의 학습 과정에 대한 알고리즘을 도시한 것이다.

[0055] 한편, Wasserstein GAN(WGAN)은 VGAN에서 손실함수를 재 정의하여 학습의 안정성을 추구한 모델로서, 손실함수를 설정할 때 Wassertein 거리를 사용하는 것을 특징으로 한다. 일반적으로 VGAN은 JS 발산(Jensen-Shannon Divergence)를 사용하는 것과 달리, Wasserstein GAN(WGAN)은 Wassertein 거리를 사용하는데, Wassertein 거리는 Earth Mover's 거리(EM 거리)라고도 한다. EM 거리는 거리의 개념을 어떤 확률 분포 모양을 띄는 흙더미를

다른 확률분포 모양을 가지도록 하는데 드는 최소비용이라고 해석하기 때문이다. 이때 비용은 흙의 양과 이동한 거리를 곱하여 정량화하게 된다. WGAN에서 JS 거리를 사용하지 않고 EM 거리를 사용하는 이유는, JS 거리로는 수렴하지 않는 확률 분포의 수열이 EM 거리에서는 수렴하는 경우가 있기 때문이다.

[0056] 도 5는 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 새로운 형태의 GAN 모델의 구조를 도시한 모식도이다. 도 5를 참조하면, 본 발명에 따른 새로운 형태의 적대적 생성 모델 시스템은 SGAN과 WGAN의 모델을 결합한 것으로서, 생성자와 판별자는 기본적으로 DCGAN의 구조를 사용하고, 각각 5층 및 4층의 합성곱 신경망으로 구성된다. 본 발명에 따른 GAN 모델(10)은 생성자(20) 및 판별자(30)를 구비하고, 생성자와 판별자는 합성곱 신경망으로 구성되고, 판별자는 분류기를 더 구비하여 판별자의 결과값을 클래스의 종류 및 가짜 유무를 판단하여 출력한다.

[0057] 이하, 도 6을 참조하여 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 GAN 모델의 구축 방법을 구체적으로 설명한다. 도 6은 본 발명의 바람직한 실시예에 따른 준지도 학습을 기반으로 한 적대적 생성 모델 구축 방법을 순차적으로 도시한 흐름도이다.

[0058] 본 발명에 따른 결합 모델의 판별자는 준지도 학습 방식을 적용하기 위하여, 실제 데이터에서 레이블 파라미터(γ)의 값을 조정하면서 일부분의 데이터만을 레이블되고 나머지 데이터는 레이블없이 사용되도록 입력데이터가 구성된다. 따라서, 본 발명에 따른 결합 모델의 판별자로 입력되는 입력데이터는 레이블된 진짜 데이터, 레이블되지 않은 진짜 데이터 및 생성자가 생성한 가짜 데이터이다.

[0059] 준지도 학습을 하게 되면, 레이블된 진짜 데이터를 사용하여 판별자의 학습을 돕기 때문에, 비지도 학습 방식보다 학습 속도가 향상된다. 그런데, 학습 초기에는 판별자가 입력 이미지를 구분하는 능력이 떨어지기 때문에 기울기 값이 큰 값을 갖게 되고 학습이 잘 진행되지만, 상대적으로 생성자보다 학습 속도가 빨라져서 분류 성능이 올라가게 되면 Mode Collapsing 문제가 일어날 수 있다. 이 경우, 학습의 안정성을 더욱 높이기 위하여, 본 발명에 따른 결합 모델은 WGAN의 EM 거리를 최소화하는 방식을 추가하는 것을 특징으로 한다.

[0060] 본 발명에 따른 GAN 결합 모델의 손실 함수는 각각 다른 최적화 방식을 통해 두 종류의 손실 함수를 최적화한다. 먼저, 준지도 학습에 따른 판별자의 손실 함수(L_D)는 수학식 9와 같은 레이블되지 않은 데이터를 이용한 비지도 학습에 따른 손실 함수와 수학식 10과 같은 레이블된 데이터를 이용한 지도 학습에 따른 손실 함수의 합으로 이루어진다. 한편, WGAN의 판별자의 손실 함수(L_{Dw})는 수학식 12과 같이 EM 거리를 최소화시키는 함수로 나타낼 수 있다. 따라서, 본 발명에 따른 GAN 결합 모델에 있어서, 판별자는 SGNA에 따른 손실함수(L_D)와 WGAN에 따른 EM 거리를 최소화하는 손실 함수(L_{Dw})를 각각 최소화시킨다.

수학식 9

[0061] $$L_{unsupervised} = -E_{x\sim p_{data}(x)}\log[1-p_{model}(y=n+1|x)]-E_{x\sim G}\log p_{model}(y=n+1|x)$$

수학식 10

[0062] $$L_{supervised} = -E_{x,y\sim p_{data}(x,y)}\log p_{model}(y|x,y<n+1)$$

수학식 11

[0063] $$L_D = L_{unsupervised} + L_{supervised}$$

수학식 12

[0064] $$L_{D_w} = -|E_{x \sim p_{data}(x)}(D) - E_{x \sim G}(1-D)|$$

[0065] 본 발명에 따른 GAN 결합 모델에 있어서, 생성자는 SGAN에 의한 손실 함수(L_G)와 WGAN에 의한 EM 거리를 최소화하는 손실함수(L_{Gw})를 각각 최소화시키게 된다. 수학식 13은 생성자는 SGAN에 의한 손실 함수(L_G)이며, 수학식 14는 WGAN에 의한 EM 거리를 최소화하는 손실함수(L_{Gw})이다.

수학식 13

[0066] $$L_G = -E_{x \sim G}\log[1 - p_{\text{model}}(y = n+1|x)]$$

수학식 14

[0067] $$L_{G_w} = -|E_{x \sim G}(1-D)|$$

[0068] 정리하면, 본 발명에 따른 GAN 모델 구축 방법은, 판별자에 대하여, 준지도 학습을 위해 레이블이 있는 경우의 손실함수와 레이블이 없는 경우의 손실함수를 각각 설정하고, 두 손실 함수의 값을 더해줌으로써 준지도 학습에 대한 손실함수를 구하고, 이에 대하여 최적화 알고리즘을 적용한다. 이때 최적화 알고리즘으로는 Adam Optimizer를 사용할 수 있다. 또한, 판별자에 대하여 WGAN에 따른 EM 거리를 최소화시키는 손실 함수를 설정하고, 이에 대하여 최적화 알고리즘을 적용하는데, 이때 최적화 알고리즘으로는 RMS Optimizer를 사용할 수 있다.

[0069] 한편, 생성자의 경우, 기본적인 적대적 생성 모델인 VGAN에 따른 손실함수를 설정하고 이에 대하여 최적화 알고리즘을 적용하는데, 이때 최적화 알고리즘으로는 Adam Optimizer를 사용할 수 있다. 또한 생성자에 대하여 WGAN에 따른 손실함수를 설정하고 이에 대하여 최적화 알고리즘을 적용하는데, 이때 최적화 알고리즘으로는 RMS Optimizer을 사용할 수 있다.

[0070] 본 발명에 따른 GAN 모델 구축 방법은 판별자를 복수 회 학습시킬 때 생성자를 1번 학습시키도록 함으로써, WGAN 손실 함수의 조건을 충족시키기 위해 가중치를 c로 클립해주는 과정에서 학습의 안정성을 위해 판별자를 더 학습시킬 수 있게 된다.

[0071] 도 7은 본 발명의 바람직한 실시예에 따른 적대적 생성 모델 구축 방법에 있어서, 본 발명에 따른 GAN 모델을 구축하는 알고리즘을 도시한 것이다. 도 7을 참조하면, 전술한 GAN 모델은, 먼저 각 파라미터들에 대하여 $\alpha = 1e-5, \beta = 5e-5, c = 0.01, m = 128, n_{critic} = 5$ 의 입력을 받아 동작한다. 먼저, 판별자 D를 5회 학습시킨 뒤 생성자 G를 1회 학습시켜야 하므로, t=0,…,5만큼 판별자 D를 우선적으로 학습시키게 된다. 각 수식들의 손실함수는 전술한 바와 동일하다.

[0072] 알고리즘의 $$g_{\theta_d} \leftarrow \nabla_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1 - D(G(z^{(i)}))) + \log D(x^{(i)}, y^{(i)})]$$ 는 SGAN의 손실함수를 의미하며 앞의 두분은 unsupervised loss 이며, 뒤의 두 부분은 supervised loss로서, 이들의 합으로 이루어진다. 한편,

알고리즘의 $g_w \leftarrow \nabla_w [\frac{1}{m}\sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z^{(i)}))]$ 는 WGAN의 손실함수를 의미한다.

[0073] 판별자의 학습 과정에서 SGAN의 손실함수와 WGAN의 손실함수를 각각 Adam optimizer 와 RMS optimizer를 이용하여 변수 θ_d, w 를 업데이트하게 된다.

[0074] 다음, 생성자 G를 한번 학습시키게 되며, 판별자의 학습 방법과 마찬가지로 진행된다.

[0075] 본 발명에 따른 GAM 결합 모델 구축 방법에 있어서, 판별자는 마지막에 분류기를 사용하게 되는데, 분류기로는 소프트맥스(Softmax) 함수를 사용하는 것이 바람직하다. 이와 같이, 판별자는 분류기로서 소프트맥스 함수를 사용함으로써, 클래스의 분류 및 가짜 여부를 판별하여 출력할 수 있게 된다. 따라서, 본 발명에 따른 GAN 결합 모델의 판별자는 준지도 학습을 통해 학습함으로써, 생성자가 생성한 데이터의 경우 진짜인지 가짜인지 여부를 구별하고 레이블된 진짜 데이터인 경우 어떤 레이블에 속하는 데이터인지 분류할 수 있게 된다.

[0076] 한편, 생성자도 마지막에 분류기를 사용하게 되는데, 분류기로는 시그모이드(Sigmoid)는 것이 바람직하다.

[0077] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0078] 10 : GAN 모델

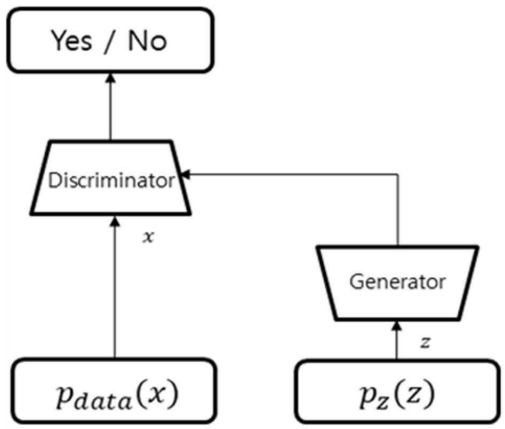
20 : 생성자

30 : 판별자

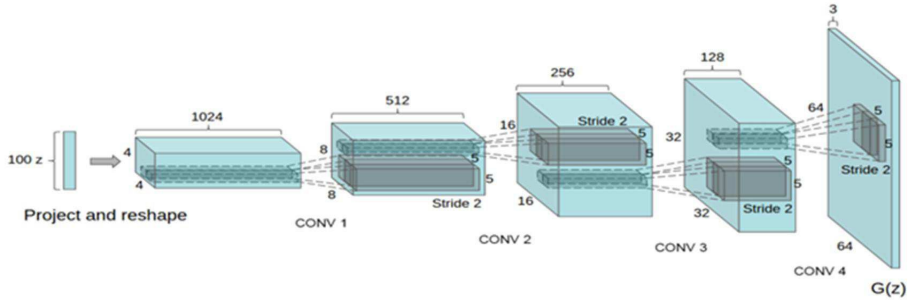
40 : 분류기

도면

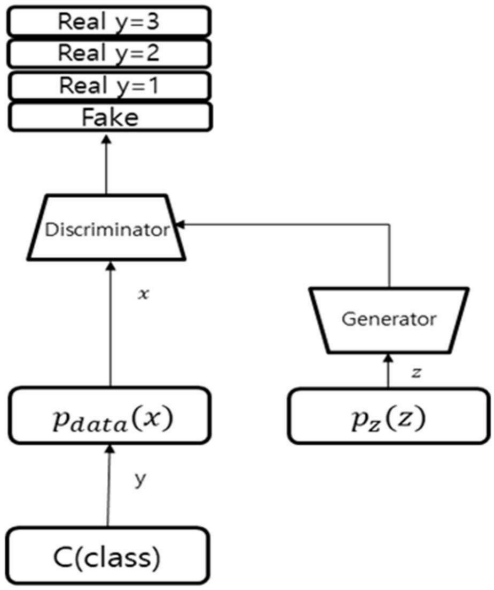
도면1



도면2



도면3



도면4

Algorithm 1. SGAN 학습 과정

Input : I : number of total iterations

for $i = 1$ **to** I **do**

Draw m noise samples $z^{(1)},\ldots,z^{(m)}$ from noise prior $p_g(z)$.

Draw m examples $(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})$ from data generation distribution $p_d(x)$.

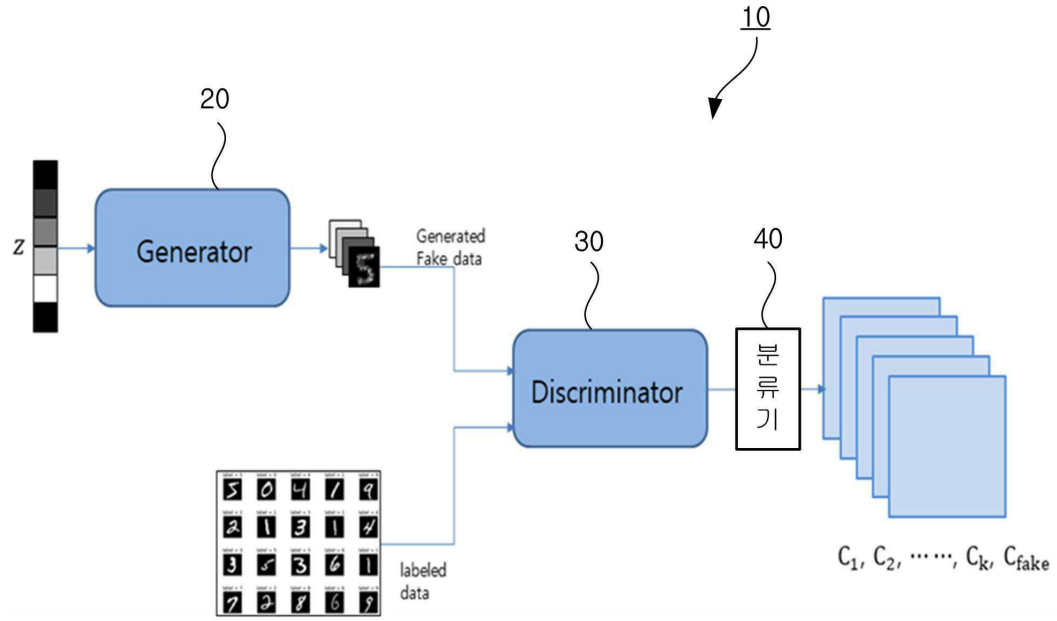
Perform gradient descent on the parameters of D w.r.t. the NLL of D/C's outputs on the combined minibatch of size 2m.

Draw m noise samples $\{z^{(1)},\ldots,z^{(m)}\}$ from noise prior $p_g(z)$

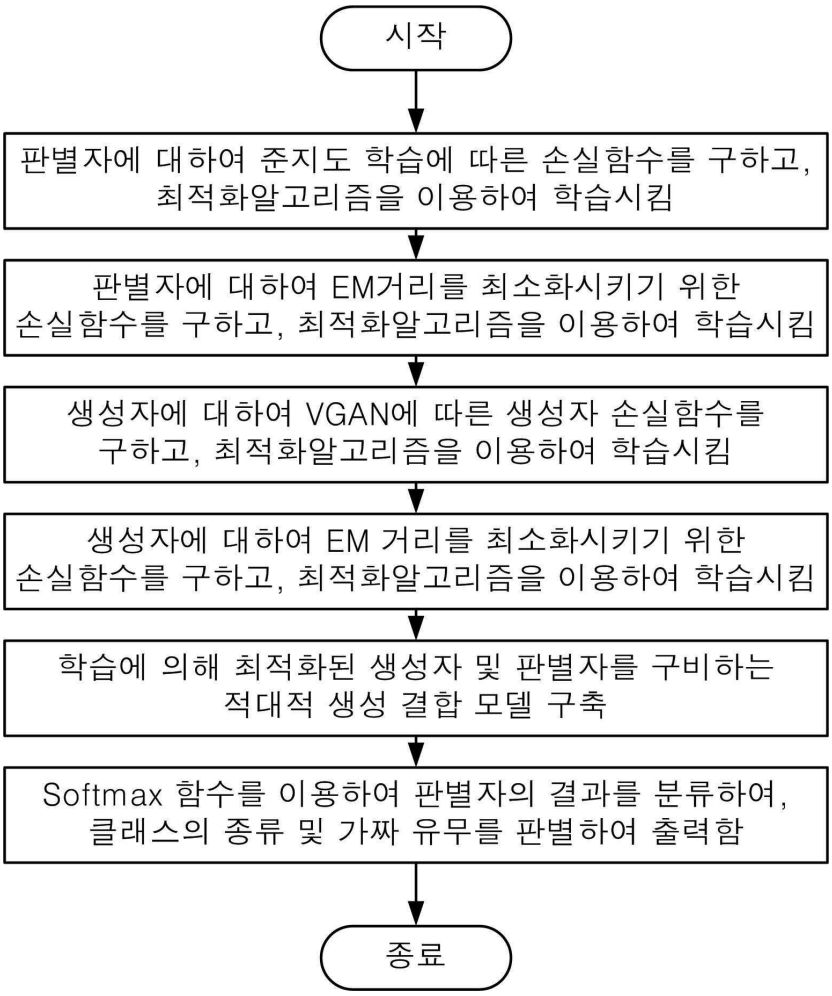
Perform gradient descent on the parameters of G w.r.t. the NLL of D/C's outputs on the combined minibatch of size m.

Output : z

도면5


10
20
z
Generator
Generated
Fake data
30
40
Discriminator
분
류
기
labeled
data
$C_1, C_2, \cdots\cdots, C_k, C_{fake}$

도면6


시작
판별자에 대하여 준지도 학습에 따른 손실함수를 구하고, 최적화알고리즘을 이용하여 학습시킴
판별자에 대하여 EM거리를 최소화시키기 위한 손실함수를 구하고, 최적화알고리즘을 이용하여 학습시킴
생성자에 대하여 VGAN에 따른 생성자 손실함수를 구하고, 최적화알고리즘을 이용하여 학습시킴
생성자에 대하여 EM 거리를 최소화시키기 위한 손실함수를 구하고, 최적화알고리즘을 이용하여 학습시킴
학습에 의해 최적화된 생성자 및 판별자를 구비하는 적대적 생성 결합 모델 구축
Softmax 함수를 이용하여 판별자의 결과를 분류하여, 클래스의 종류 및 가짜 유무를 판별하여 출력함
종료

도면7

Algorithm 3. 모델 학습 과정 ($\alpha=1e-5, \beta=5e-5, c=0.01, m=128, n_{critic}=5$)

Require : α, β the learning rate, c, the clipping parameter, m, the batch size, n_{critic}, the number of iterations of the critic per generator iteration.

while number of training iterations **do**

for $t=0,\ldots,n_{critic}$ **do**

Sample minibatch of m noise samples $\{z^{(1)},\ldots,z^{(m)}\}$ from noise prior $p_g(z)$.

Sample minibatch of m examples $\{x^{(1)},\ldots,x^{(m)}\}$ from data generating distribution $p_{data}(x)$ and $\{y^{(1)},\ldots,y^{(m)}\}$ is the label set for x

$$g_{\theta_d} \leftarrow \nabla_{\theta_d} \frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1-D(G(z^{(i)}))) + \log D(x^{(i)},y^{(i)})]$$

$$g_w \leftarrow \nabla_w [\frac{1}{m}\sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z^{(i)}))]$$

$$\theta_d \leftarrow \theta_d + \alpha \mathrm{Adam}(\theta_d, g_{\theta_d})$$

$$w \leftarrow w + \beta \mathrm{RMSProp}(w, g_w)$$

$$w \leftarrow clip(w, -c, c)$$

end for

Sample minibatch of m noise samples $\{z^{(1)},\ldots,z^{(m)}\}$ from noise prior $p_g(z)$.

$$g_{\theta_g} \leftarrow \nabla_{\theta_g} \frac{1}{m}\sum_{i=1}^{m}\log(1-D(G(z^{(i)})))$$

$$g_\theta \leftarrow -\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} f_w(g_\theta(z^{(i)}))$$

$$\theta_g \leftarrow \theta_g + \alpha \mathrm{Adam}(\theta_g, g_{\theta_g})$$

$$\theta \leftarrow \theta - \beta \mathrm{RMSProp}(\theta, g_\theta)$$

end while