



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0059207
(43) 공개일자 2024년05월07일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G06N 3/04 (2023.01)
G06N 7/00 (2023.01)
(52) CPC특허분류
G06N 3/088 (2023.01)
G06N 3/045 (2023.01)
(21) 출원번호 10-2022-0140197
(22) 출원일자 2022년10월27일
심사청구일자 2022년10월27일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
최동완
인천광역시 미추홀구 용정공원로 33, 108동 1303호(용현동, 인천 SK Sky VIEW)
강한얼
제주특별자치도 제주시 남광로 181, 301동 105호(일도이동, 대유대림아파트)
(74) 대리인
양성보

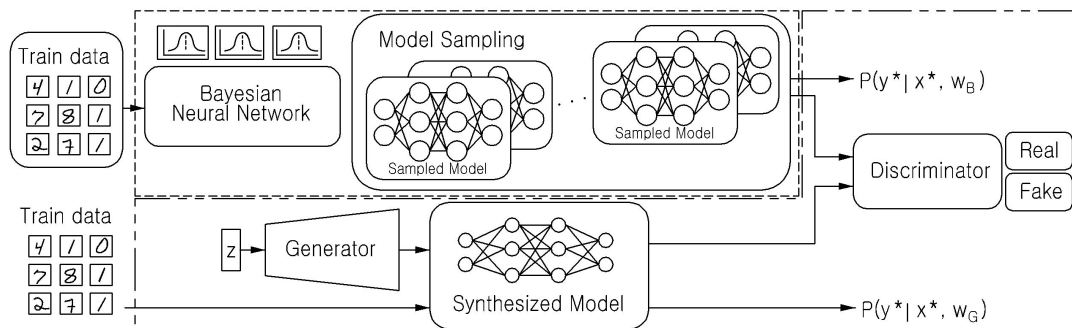
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 과거 모델 파라미터 재생성을 통한 태스크 점진 학습 방법 및 시스템

(57) 요약

과거 모델 파라미터 재생성을 통한 태스크 점진 학습 방법 및 시스템이 개시된다. 일 실시예에 따른 학습 시스템에 의해 수행되는 태스크 점진 학습 방법은, 적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득하는 단계; 및 상기 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성하는 단계를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/047 (2023.01)

G06N 7/01 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160586
과제번호	2022-0-00448-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발사업
연구과제명	[Ezbaro] 인간처럼 회상이 가능한 인공 신경망 지속학습 플랫폼 개발(1차년도)
기 여 율	50/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.04.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711163299
과제번호	2021R1F1A1060160
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기본연구
연구과제명	[Ezbaro] 질의 가능한 인공지능: 상호작용형 XAI 서비스를 실현하기 위한 질의 기반
딥러닝 모델 해석 기술	
기 여 율	30/100
과제수행기관명	인하대학교
연구기간	2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711174118
과제번호	RS-2022-00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(1차년도)
기 여 율	10/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.07.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711152759
과제번호	2020-0-01389-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합연구센터지원사업(국고)
연구과제명	[Ezbaro][정부] 인공지능융합연구센터지원(3차년도)
기 여 율	10/100
과제수행기관명	인하대학교 산학협력단
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

학습 시스템에 의해 수행되는 태스크 점진 학습 방법에 있어서,

적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득하는 단계;
및

상기 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 2

제1항에 있어서,

상기 사전 학습된 모델은, 베이지안 신경망을 기반으로 구성된 것이고,

상기 샘플링된 데이터를 획득하는 단계는,

상기 사전 학습된 모델로부터 파라미터를 샘플링하여 동일한 태스크에 대해 서로 다른 복수 개의 모델의 파라미터 집합을 추출하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 3

제2항에 있어서,

상기 샘플링된 데이터를 획득하는 단계는,

상기 사전 학습된 모델로부터 적대적 생성 신경망의 학습의 반복 시점(iteration)마다 미니 배치(mini-batch)의 개수만큼 파라미터를 샘플링하고, 상기 샘플링을 통해 파라미터 집합을 획득하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 4

제1항에 있어서,

상기 적대적 생성 모델은, 생성 신경망과 판별 신경망으로 구성되고,

상기 판별 신경망은, 상기 사전 학습된 모델로부터 샘플링된 파라미터와 상기 생성 신경망으로부터 획득된 합성 파라미터를 구분하도록 학습을 수행하고,

상기 생성 신경망은, 잠재 벡터를 이용하여 획득된 합성 파라미터를 상기 판별 신경망이 구분하지 못할 정도로 상기 샘플링된 파라미터와 유사한 합성 파라미터를 생성하도록 학습을 수행하는

것을 특징으로 하는 태스크 점진 학습 방법.

청구항 5

제1항에 있어서,

상기 재생성하는 단계는,

유사 리허설 방법을 사용하여 상기 사전 학습된 모델로부터 각각의 태스크에 해당하는 새로운 신경망을 재생성하도록 지속학습을 수행하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 6

제5항에 있어서,

상기 재생성하는 단계는,

적대적 생성 신경망에 구성된 생성 신경망을 통해 샘플링된 잠재 벡터와 태스크 정보(t)를 이용하여 상기 태스크 정보의 순서에서 태스크에 대한 재생성된 파라미터를 재생성하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 7

제6항에 있어서,

상기 재생성하는 단계는,

태스크를 학습하는 현재 시점에서, 현재 시점의 태스크에 해당하는 베이지안 신경망으로부터 샘플링된 파라미터와 현재 시점을 기준으로 이전 시점의 태스크에 대해 적대적 생성 신경망에 구성된 생성 신경망을 통해 재생성된 파라미터를 함께 학습하는 단계

를 포함하는 태스크 점진 학습 방법.

청구항 8

학습 시스템에 있어서,

적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득하는 샘플링부; 및

상기 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성하는 재생성부

를 포함하는 학습 시스템.

발명의 설명

기술 분야

[0001]

아래의 설명은 태스크 점진 학습 기술에 관한 것이다.

배경 기술

[0003]

딥러닝 모델은 최근 다양한 분야에 걸쳐 우수한 성능을 보이고 있다. 그러나 과거의 지식을 보존하며 새로 유입되는 정보를 지속적으로 학습할 수 있는 인간의 학습능력과 달리 딥러닝 알고리즘의 우수한 성능을 보장하기 위해서는 관측한 모든 데이터를 사용한 학습이 요구된다. 이러한 특징은 새로운 데이터를 학습할 시 기존의 모든 데이터를 보유하고 다시 학습해야 한다는 비효율을 야기한다. 딥러닝 모델은 현재 데이터만을 사용하여 학습할 경우(fine-tuning) 과거 데이터에 대한 극심한 성능 하락을 경험하며, 이는 지속학습에서 '치명적 망각 현상(catastrophic forgetting)'이라고 불린다. 지속학습(continual learning)은 딥러닝의 이러한 치명적 망각 현상을 완화하는 문제를 다루고 있다.

[0004]

한편, 적대적 생성 신경망(generative adversarial networks; GAN)의 등장 이후 이미지 생성 모델의 활용은 다양한 분야로 확장되고 있다. 지속학습(continual learning)에서도 과거 데이터에 대한 접근이 제한되는 상황에서 합성 이미지를 생성하여 부족한 과거 데이터를 대신하는 역할로 적대적 생성 신경망을 활용한다. 이러한 방식은 유사 리허설(pseudo-rehearsal)이라고 불린다. 하지만 이와 같이 적대적 생성 신경망을 활용하는 방식은 적대적 생성 신경망이 생성한 합성 이미지의 품질이 실제 이미지에 미치지 못하며, 지속학습에서는 이로 인한 적대적 생성 신경망의 성능 하락이 누적되고 치명적 망각 현상으로 이어진다는 한계점이 존재한다.

발명의 내용

해결하려는 과제

- [0006] 적대적 생성 모델을 통한 적대적 생성 학습을 활용하여 사전 학습한 베이지안 신경망의 파라미터와 유사한 분포를 가지는 새로운 신경망을 재생성하는 방법 및 시스템을 제공할 수 있다.
- [0007] 유사 리허설(pseudo-rehearsal) 방식을 통해 치명적 망각 없이 과거 모든 시점의 신경망을 재생성하여 지속학습을 가능하게 하는 방법 및 시스템을 제공할 수 있다.

과제의 해결 수단

- [0009] 학습 시스템에 의해 수행되는 태스크 점진 학습 방법은, 적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득하는 단계; 및 상기 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성하는 단계를 포함할 수 있다.
- [0010] 상기 사전 학습된 모델은, 베이지안 신경망을 기반으로 구성된 것이고, 상기 샘플링된 데이터를 획득하는 단계는, 상기 사전 학습된 모델로부터 파라미터를 샘플링하여 동일한 태스크에 대해 서로 다른 복수 개의 모델의 파라미터 집합을 추출하는 단계를 포함할 수 있다.
- [0011] 상기 샘플링된 데이터를 획득하는 단계는, 상기 사전 학습된 모델로부터 적대적 생성 신경망의 학습의 반복 시점(iteration)마다 미니 배치(mini-batch)의 개수만큼 파라미터를 샘플링하고, 상기 샘플링을 통해 파라미터 집합을 획득하는 단계를 포함할 수 있다.
- [0012] 상기 적대적 생성 모델은, 생성 신경망과 판별 신경망으로 구성되고, 상기 판별 신경망은, 상기 사전 학습된 모델로부터 샘플링된 파라미터와 상기 생성 신경망으로부터 획득된 합성 파라미터를 구분하도록 학습을 수행하고, 상기 생성 신경망은, 잠재 벡터를 이용하여 획득된 합성 파라미터를 상기 판별 신경망이 구분하지 못할 정도로 상기 샘플링된 파라미터와 유사한 합성 파라미터를 생성하도록 학습을 수행할 수 있다.
- [0013] 상기 재생성하는 단계는, 유사 리허설 방법을 사용하여 상기 사전 학습된 모델로부터 각각의 태스크에 해당하는 새로운 신경망을 재생성하도록 지속학습을 수행하는 단계를 포함할 수 있다.
- [0014] 상기 재생성하는 단계는, 적대적 생성 신경망에 구성된 생성 신경망을 통해 샘플링된 잠재 벡터와 태스크 정보(t)를 이용하여 상기 태스크 정보의 순서에서 태스크에 대한 재생성된 파라미터를 재생성하는 단계를 포함할 수 있다.
- [0015] 상기 재생성하는 단계는, 태스크를 학습하는 현재 시점에서, 현재 시점의 태스크에 해당하는 베이지안 신경망으로부터 샘플링된 파라미터와 현재 시점을 기준으로 이전 시점의 태스크에 대해 적대적 생성 신경망에 구성된 생성 신경망을 통해 재생성된 파라미터를 함께 학습하는 단계를 포함할 수 있다.
- [0016] 학습 시스템은, 적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득하는 샘플링부; 및 상기 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성하는 재생성부를 포함할 수 있다.

발명의 효과

- [0018] 과거 지식에 대한 망각없이 지속적인 추론을 가능하게 함으로써 치명적 망각현상을 해결할 수 있다.
- [0019] 사전 학습 모델로 베이지안 신경망을 사용하여 적대적 생성 신경망의 학습을 위해 다수의 모델을 학습시킬 필요 없이 학습된 하나의 베이지안 신경망에서 여러 개의 모델을 샘플링하는 것이 가능하다.

도면의 간단한 설명

- [0021] 도 1은 일 실시예에 있어서, 새로운 태스크 점진 학습의 개괄적인 동작을 설명하기 위한 도면이다.
- 도 2는 일 실시예에 있어서, 새로운 태스크 점진 학습의 상세 동작을 설명하기 위한 도면이다.
- 도 3은 일 실시예에 있어서, 학습 시스템의 구성을 설명하기 위한 블록도이다.
- 도 4는 일 실시예에 있어서, 태스크 점진 학습 방법을 설명하기 위한 흐름도이다.
- 도 5는 일 실시예에 있어서, 판별 신경망의 구분 결과를 비교한 그래프이다.
- 도 6은 일 실시예에 있어서, 기존 모델과 재생성 모델 파라미터 값의 L2-Norm을 비교한 그래프이다.
- 도 7은 일 실시예에 있어서, 각 태스크에서의 태스크 점진 학습 방법별 성능을 비교한 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0022] 이하, 실시예를 첨부한 도면을 참조하여 상세히 설명한다.

- [0024] 도 1은 일 실시예에 있어서, 새로운 태스크 점진 학습의 개괄적인 동작을 설명하기 위한 도면이다.

- [0025] 우선적으로, 실시예에서 제안하는 태스크 점진 학습 동작에 대한 설명의 이해를 돕기 위하여 적대적 생성 신경망, 베이지안 신경망 및 지속학습에 대하여 설명하기로 한다.

- [0026] 적대적 생성 신경망(GAN)은 딥러닝 기반의 이미지 생성 모델로서 적대적 생성 학습을 통해 실제와 유사한 합성 데이터를 생성한다. 적대적 생성 학습은 적대적 생성 신경망의 생성 신경망(generator)과 판별 신경망(discriminator) 사이의 술래잡기 게임을 통해 이루어진다. 판별 신경망은 실제 이미지와 합성 이미지를 정확하게 구분(binary classification)하도록 학습되고, 이와 반대로 생성 신경망은 실제 이미지처럼 보이는 합성 이미지를 생성하여 판별 신경망이 이를 구분하지 못하도록 학습된다. 수학적 1은 두 신경망 사이의 적대적인 학습 방식을 반영한 적대적 생성 신경망의 목적함수이다.

- [0027] 수학적 1:

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

- [0028]

- [0029] 수학적 1에서 D, G는 각각 판별 신경망과 생성 신경망을 의미하고, z는 생성 신경망의 입력(input)으로 주어지는 잠재 벡터(latent vector)로 가우시안 분포(Gaussian distribution) $N(0, 1)$ 로부터 랜덤하게 샘플링되어 주어진다($z \sim N(0, 1)$). 실시예에서는 사전 학습 모델의 파라미터 재생성을 위해 적대적 생성 신경망을 사용할 수 있다.

- [0030] 파라미터가 고정된 값을 가지는 기존의 인공 신경망과 달리, 베이지안 신경망은 신경망 파라미터(가중치)에 대한 확률 분포를 가정한다. 베이지안 신경망에서 신경망의 파라미터 w는 데이터를 관측한 후, 파라미터 w에 대한 확률 분포인 사후분포(posterior distribution) $P(w | D)$ 를 따른다. 하지만 파라미터 w의 실제 사후분포를 직접 계산하는 것은 계산적으로 다루기 힘든 적분 연산을 수반한다는 문제점을 가진다. 계산 불가능한 사후분포 사용의 대안으로 베이지안 인공 신경망 중 하나인 Bayes by Backprob(BBB)에서는 사후분포 $P(w | D)$ 를 다루기 쉬운 분포 $q(w | \theta)$ 로 근사하는 방법인 변분 추론법(variational Inference)을 사용한다. 참고로, 비특허문헌 1(Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra, "Weight uncertainty in neural network," Proc. of the International Conference on Machine Learning, pp.1613-1622, 2015)에 Bayes by Backprob(BBB)이 언급되어 있다. $q(w | \theta)$ 로는 가우시안 분포 $N(\mu, \sigma^2)$ 를 사용한다. 이때, θ 는 확률 분포 $q(w | \theta)$ 의 파라미터로 가우시안 분포의 평균과 분산에 해당한다($\theta = \{\mu, \sigma^2\}$). 따라서, BBB에서 가중치 w는 확률 분포 $q(w | \theta)$ 를 따르는 확률 변수이므로 가우시안 분포인 $q(w | \theta)$ 로부터 샘플링된다.

[0031] 베이지안 신경망에서는 가중치에 대한 확률분포를 가정한다는 특성으로 인해 가중치 업데이트 또한 기존 인공 신경망과 차이가 있다. 가중치 w 에 대한 그레디언트를 계산하고, 업데이트하는 기존의 인공 신경망과는 다르게, BBB에서는 w 가 따르는 확률 분포의 파라미터 $\theta = \{\mu, \sigma^2\}$ 에 대한 그레디언트가 계산되고 업데이트되는 값 역시 μ 와 σ^2 가 된다.

[0032] 변분 추론법은 사후분포 $P(w | D)$ 에 근사하는 $q(w | \theta)$ 를 구하는 방법이므로 BBB는 $P(w | D)$ 와 $q(w | \theta)$ 사이의 쿨백-라이블러 발산(KL-divergence)을 목적함수로 하여 이를 최소화한다. 따라서, 최적의 파라미터인 θ^* 은 수학식 2와 같이 구해질 수 있으며, 이는 $q(w | \theta)$ 와 실제 사후분포 $P(w | D)$ 사이의 쿨백-라이블러 발산(KL-divergence)을 최소화하여 실제 사후분포에 근사하는 $q(w | \theta^*)$, $\theta^* = \mu^*, \sigma^{*2}$ 를 학습하는 것을 나타낸다.

[0033] 수학식 2:

$$\begin{aligned} \theta^* &= \operatorname{argmin}_{\theta} KL[q(w | \theta) \| P(w | D)] \\ &= \operatorname{argmin}_{\theta} \int q(w | \theta) \log \frac{q(w | \theta)}{P(w)P(D | w)} dw \\ &= \operatorname{argmin}_{\theta} KL[q(w | \theta) \| P(w)] \\ &\quad - E_{q(w | \theta)}[\log P(D | w)]. \end{aligned}$$

[0034]

[0035] 지속학습(continual learning)은 딥러닝의 한 패러다임으로 데이터를 축적하지 못하는 상황에서 지속적으로 들어오는 데이터를 학습함과 동시에 이전에 학습한 데이터에 대한 성능을 유지하는 문제를 다룬다. 지속학습 시 발생하는 과거 데이터에 대한 성능 하락(치명적 망각 현상)을 완화하기 위해 현재까지 다양한 지속학습 방법이 제안되었다. 대표적인 방법으로 파라미터 갱신에 제약을 주는 정규화 기반(regularization-based) 방식이 있다. 정규화 기반의 지속학습은 새로운 데이터 학습 시에 이전 시점의 학습에서 중요했던 가중치가 변화하는 것에 페널티를 부과하여 과거 지식의 망각을 완화한다. 정규화 기반 방식 이외에도 이전에 학습한 데이터의 일부를 저장하여 다음 학습 시 새로운 데이터와 함께 학습시키는 리허설 메모리(rehearsal memory) 방식이 존재한다. 추가적으로 유사 리허설(pseudo rehearsal)방식은 리허설 메모리 방법의 일종으로 이전에 학습한 데이터를 그대로 저장하는 것이 아닌 이미지 생성 모델을 학습시켜 생성 모델이 생성한 과거 데이터를 현재 데이터와 같이 학습시키는 방식이다.

[0036] 실시예에서는 이미지가 아닌 모델의 파라미터를 재생성하는 방법인 태스크 점진 학습을 통해 과거 지식에 대한 망각없이 지속적인 추론을 가능하게 함으로써 치명적 망각현상을 해결하고자 한다. 실시예에서 제안하는 태스크 점진 학습은 사전 학습 모델로 베이지안 신경망을 사용하여 적대적 생성 신경망의 학습을 위해 다수의 모델을 학습시킬 필요 없이 학습된 하나의 베이지안 신경망에서 복수 개의 모델을 샘플링하는 것이 가능하다. 이는 기존의 지속학습 알고리즘에서는 시도된 적이 없는 새로운 방법이다

[0037] 도 2는 일 실시예에 있어서, 새로운 태스크 점진 학습의 상세 동작을 설명하기 위한 도면이다.

[0038] 적대적 생성 신경망이 적대적 생성 학습을 통해 데이터의 분포를 학습하기 위해서는 충분한 양의 학습 데이터가 필요하다. 실시예에서 제안되는 태스크 점진 학습 방법 역시 사전 학습된 신경망의 파라미터를 적대적 생성 모델을 통해 학습하기 때문에 다수의 사전 학습된 모델들의 집합이 필요하다. 그러나 이를 위해 사전에 다수의 모델을 학습시키는 것은 현실적으로 불가능하다. 이러한 문제를 해결하기 위해 실시예에서는 사전 학습되는 모델로 베이지안 신경망(bayesian neural networks)인 Bayes by Backprob(BBB) 모델이 사용될 수 있다.

[0039] 베이지안 신경망인 BBB는 파라미터가 고정된 값이 아닌 확률 분포를 따른다고 가정하며, 실제 사후분포에 근사하는 가우시안 분포 $q(w | \theta)$, $\theta = \{\mu, \sigma^2\}$ 를 학습한다. 가우시안 분포의 파라미터 θ 는 수학식 2의 목적함수를 통해 학습되며 학습이 종료된 이후에는 $q(w | \theta)$ 로부터 복수 개의 가중치 w 를 랜덤 추출하는 것이 가능해진다. 다시 말해서, 하나의 베이지안 신경망 BBB를 학습한 이후에는 필요할 때마다 신경망 파라미터

$w_B \sim N(\mu^*, \sigma^{*2})$ 를 샘플링하여 복수 개의 신경망 파라미터 집합을 확보하는 것이 가능해진다. 여기서, w_B 는 학습된 베이지안 신경망으로부터 샘플링된 파라미터를 의미한다. 이러한 파라미터 샘플링을 모델 내 모든 파라미터에 대해 수행하면 하나의 베이지안 신경망으로부터 동일한 태스크에 대해 복수 개의 서로 다른 모델을 추출할 수 있다. 도 1에서 볼 수 있듯이 실시예에서는 사전 학습 모델로 베이지안 신경망 BBB를 학습한 후 적대적 생성 신경망의 학습의 반복 시점(iteration)마다 미니 배치(mini-batch) 수만큼 신경망이 샘플링되고, 샘플링된 모델 집합이 적대적 생성 신경망의 적대적 생성 학습을 위한 데이터로 사용될 수 있다.

[0040] 실시예에서는 일반적인 적대적 생성 신경망과 유사하게 생성 신경망이 실제 데이터와 유사한 데이터를 생성하도록 적대적 생성 학습이 진행될 수 있다. 다만 도 1에서 볼 수 있듯이, 목적함수를 나타내는 수학적 식 3은 일반적인 적대적 생성 신경망과 마찬가지로 생성 신경망과 판별 신경망 사이의 술래잡기 게임을 반영한다.

[0041] 수학적 식 3:

$$\begin{aligned} \underset{G}{\text{MinMax}} \underset{D}{V}(D, G) &= E_{w_B \sim q(w \mid \theta)} [\log D(w_B)] \\ &+ E_{w_B \sim p_G} [\log(1 - D(w_G))]. \end{aligned}$$

[0042]

[0043] 수학적 식 3의 Min-Max 게임을 통해 판별 신경망(D)은 베이지안 신경망으로부터 샘플링된 파라미터 w_B 와 생성 신경망(G)이 재생성한 파라미터 w_G 를 구분(binary classification) 하도록 학습될 수 있다. 생성 신경망(G)은 잠재 벡터 $z \sim N(0, 1)$ 를 입력으로 받아 합성 파라미터 $w_G = G(z)$ 를 판별 신경망(D)가 구분하지 못할 정도로 실제와 유사한 합성 파라미터를 생성하도록 학습될 수 있다.

[0044] 예를 들면, 적대적 생성 신경망 모델로 MLP(multi-layer perceptron)로 이루어져 있는 Naïve GAN이 사용될 수 있다. 이후 파라미터 재생성을 지속학습으로 확장한 실험에서는 적대적 생성 신경망의 학습 수용력을 올리기 위해 WGAN-GP가 사용될 수 있다. WGAN-GP는 Wasserstein 거리와 그레디언트 페널티를 손실함수로 사용하여 학습이 불안정한 적대적 생성 신경망의 문제점을 개선한 적대적 생성 신경망의 일종이다.

[0045] 적대적 생성 신경망은 실제 데이터의 분포를 따르는 새로운 합성 데이터를 생성하는 모델이다. 따라서 생성 신경망(G)의 학습이 수렴한 이후, 합성 파라미터 w_G 로 구성된 새로운 신경망은 기존 사전 학습된 신경망과 동일한 분포를 따른다. 그러므로 새로운 데이터(x^*, y^*)를 관측하였을 때, 합성 파라미터로 구성된 새로운 신경망을 사용한 추론 결과는 사전 학습된 신경망의 추론 결과와 유사하게 나타난다($P(y^* \mid x^*, w_G) \approx P(y^* \mid x^*, w_B)$).

[0046] 또한, 실시예에서는 신경망 재생성 동작에 유사 리허설 방법을 적용하여 과거 여러 시점에서 학습한 신경망을 재생성해내는 지속학습 동작에 대하여 설명하기로 한다. 도 2는 지속학습 패러다임에서 학습되는 과정을 보여준다. 도 2에서 볼 수 있듯이 지속학습에서는 전체 데이터셋이 한번에 주어지는 것이 아닌, 데이터셋의 일부가 각각의 태스크(task)에 걸쳐서 관측된다. 따라서 각각의 태스크에 해당하는 신경망을 재생성하도록 학습이 수행될 수 있다.

[0047] 첫번째 태스크(Task1)에서는 Task1에 해당하는 단일한 신경망 w_G^1 만을 생성하도록 학습된다. w_G^1 은 Task 1에 해당하는 데이터를 추론할 수 있도록 재생성된 파라미터(합성 파라미터)를 의미한다. 이때, 기존의 적대적 생성 신경망의 학습 방법과 다른 한 가지 차이점은 적대적 생성 신경망이 몇 번째 태스크에 대한 신경망을 생성해야 하는지 지시해주는 추가 정보가 있어야 한다는 점이다. 이를 위해 합성 데이터에 대한 추가적인 라벨을 부여하여 주어진 라벨에 해당하는 데이터의 생성을 가능하게 하는 조건부 GAN(conditional GAN)을 사용한다. 이때, 적대적 생성 신경망이 각각의 태스크에 대한 신경망 파라미터를 생성할 수 있어야 하므로 생성해야 하는 파라미터가 몇 번째 태스크에 해당하는 파라미터인가를 지시해주는 태스크 정보 t를 적대적 생성 신경망의 학습 및 추론 시에 제공한다. 따라서 생성 신경망(G)는 가우시안 분포에서 샘플링된 잠재 벡터 z와 태스크 지시 정보 t를 입력 데이터로 받아 t번째 태스크에 대해 재생성된 신경망 w_G^t 을 재생성한다.

[0048] 수학적식 4:

$$w_G^t = G(z, t),$$

$$z \sim N(0, 1), \quad t = 1, 2, 3, \dots$$

[0049]

[0050] 두번째 태스크(Task2)부터는 현재 태스크에 대한 신경망 파라미터뿐만 아니라 과거 태스크에 대한 파라미터 또한 생성해야 한다. 이를 위해 과거 지식을 망각하지 않으면서 현재 지식을 학습하는 지속학습이 적대적 생성 신경망에서 이루어져야 한다. 지속학습을 위해 유사 리허설(Pseudo-Rehearsal) 방법을 사용하여 과거의 지식을 보존한다. 두 번째 태스크(Task2)를 학습하는 시점에서 Task2에 해당하는 베이지안 신경망으로부터 샘플링된 파라미터 w_B^2 을 학습한다. 아울러 이전 태스크의 파라미터 w_B^1 을 학습하기 위해 Task 1에 대해 재생성된 파라미터 w_G^1 을 현재 태스크의 파라미터 w_B^2 와 함께 학습하게 된다. 즉, T번째 태스크를 학습하는 시점에서 현재 태스크에 해당하는(베이지안 신경망으로부터 샘플링된) 파라미터 w_B^T 와 바로 이전 시점(Task T-1)에 재생성된 과거 모든 시점에 해당하는 파라미터 $w_G^{1:T-1}$ 를 합친 파라미터 집합 W 를 학습할 수 있다.

[0051] 수학적식 5:

$$W = w_B^T \cup w_G^{1:T-1},$$

$$w_G^{1:T-1} = w_G^1 \cup w_G^2 \cup \dots \cup w_G^{T-1}.$$

[0052]

[0053] 이때, $w_G^{1:T-1}$ 을 구성하는 각 시점의 파라미터 $w_G^i (i = 1, \dots, T-1)$ 은 과거에 재생성된 파라미터가 저장된 것이 아니며, T시점의 학습에서 바로 이전 시점(T-1)의 파라미터만을 저장해두고 태스크 정보 $t = 1, \dots, T-1$ 만을 달리하여 재생성한 파라미터이다. 이러한 과거 시점의 파라미터 재생성은 적대적 생성 신경망 학습의 반복 시점(iteration)마다 미니 배치(mini-batch) 수만큼만 발생하여 기존의 미니 배치 이상의 메모리를 필요로 하지 않는다.

[0054] 도 2의 파란색 점선 화살표는 이전 시점에 생성된 과거 모든 시점의 파라미터($w_G^{1:T-1}$)가 태스크 라벨(t)과 함께 현재 시점의 학습 데이터로 주어지는 유사 리허설 방식을 나타낸다.

[0055] 도 3은 일 실시예에 있어서, 학습 시스템의 구성을 설명하기 위한 블록도이고, 도 4는 일 실시예에 있어서, 태스크 점진 학습 방법을 설명하기 위한 흐름도이다.

[0056] 학습 시스템(100)의 프로세서는 샘플링부(310) 및 재생성부(320)를 포함할 수 있다. 이러한 프로세서의 구성요소들은 학습 시스템에 저장된 프로그램 코드가 제공하는 제어 명령에 따라 프로세서에 의해 수행되는 서로 다른 기능들(different functions)의 표현들일 수 있다. 프로세서 및 프로세서의 구성요소들은 도 4의 태스크 점진 학습 방법이 포함하는 단계들(410 내지 420)을 수행하도록 학습 시스템을 제어할 수 있다. 이때, 프로세서 및 프로세서의 구성요소들은 메모리가 포함하는 운영체제의 코드와 적어도 하나의 프로그램의 코드에 따른 명령(instruction)을 실행하도록 구현될 수 있다.

[0057] 프로세서는 태스크 점진 학습 방법을 위한 프로그램의 파일에 저장된 프로그램 코드를 메모리에 로딩할 수 있다. 예를 들면, 학습 시스템에서 프로그램이 실행되면, 프로세서는 운영체제의 제어에 따라 프로그램의 파일로부터 프로그램 코드를 메모리에 로딩하도록 학습 시스템을 제어할 수 있다. 이때, 샘플링부(310) 및 재생성부(320) 각각은 메모리에 로딩된 프로그램 코드 중 대응하는 부분의 명령을 실행하여 이후 단계들(410 내지 420)을 실행하기 위한 프로세서의 서로 다른 기능적 표현들일 수 있다.

[0058] 단계(410)에서 샘플링부(310)는 적대적 생성 모델의 입력 데이터로 사용하기 위해 사전 학습된 모델로부터 샘플링된 데이터를 획득할 수 있다. 이때, 사전 학습된 모델은, 베이지안 신경망을 기반으로 구성된 것일 수 있다. 샘플링부(310)는 사전 학습된 모델로부터 파라미터를 샘플링하여 동일한 태스크에 대해 서로 다른 복수 개의 모델의 파라미터 집합을 추출할 수 있다. 샘플링부(310)는 사전 학습된 모델로부터 적대적 생성 신경망의 학습의 반복 시점(iteration)마다 미니 배치(mini-batch)의 개수만큼 파라미터를 샘플링하고, 샘플링을 통해 파라미터

집합을 획득할 수 있다.

- [0059] 단계(420)에서 재생성부(320)는 획득된 샘플링된 파라미터를 이용하여 적대적 생성 모델을 통한 적대적 생성 학습을 수행함에 따라 구성된 합성 파라미터로 새로운 신경망을 재생성할 수 있다. 이때, 적대적 생성 모델은, 생성 신경망과 판별 신경망으로 구성된 것일 수 있다. 판별 신경망은, 사전 학습된 모델로부터 샘플링된 파라미터와 생성 신경망으로부터 획득된 합성 파라미터를 구분하도록 학습을 수행하고, 생성 신경망은, 잠재 벡터를 이용하여 획득된 합성 파라미터를 판별 신경망이 구분하지 못할 정도로 샘플링된 파라미터와 유사한 합성 파라미터를 생성하도록 학습을 수행할 수 있다. 재생성부(320)는 유사 리히설 방법을 사용하여 사전 학습된 모델로부터 각각의 태스크에 해당하는 새로운 신경망을 재생성하도록 지속학습을 수행할 수 있다. 재생성부(320)는 적대적 생성 신경망에 구성된 생성 신경망을 통해 샘플링된 잠재 벡터와 태스크 정보(t)를 이용하여 태스크 정보의 순서에서 태스크에 대한 재생성된 파라미터를 재생성할 수 있다. 재생성부(320)는 태스크를 학습하는 현재 시점에서, 현재 시점의 태스크에 해당하는 베이지안 신경망으로부터 샘플링된 파라미터와 현재 시점을 기준으로 이전 시점의 태스크에 대해 적대적 생성 신경망에 구성된 생성 신경망을 통해 재생성된 파라미터를 함께 학습할 수 있다.
- [0060] 실시예에서 제안된 태스크 잠재 학습 방법의 기술적 효과에 대하여 설명하기로 한다. 지속학습 성능을 보이기에 앞서 적대적 생성 신경망이 적대적 생성 학습을 통해 이미지가 아닌 신경망 파라미터를 합성해낼 수 있는가를 다양한 실험을 통해 입증될 수 있다. 이후 지속학습으로 확장됐을 때에 과거 태스크에 대한 치명적 망각 현상이 발생하지 않고 신경망 재생성 능력이 유지됨을 확인하기 위해 태스크 점진 학습 시나리오에서 보편적으로 사용되는 Split-MNIST, Permuted-MNIST 벤치마크(benchmark) 실험이 진행될 수 있다. 마지막으로 기존의 태스크 점진 학습 방법과의 비교를 통해 실시예에서 제안된 태스크 점진 학습 방법이 더 나은 결과를 보임을 확인함으로써 지속학습에서 효과적임을 검증할 수 있다.
- [0061] 실시예에서 제안된 태스크 점진 학습 방법의 신경망 재생성 결과를 확인하기 위해 Iris와 MNIST 데이터셋을 학습한 각각의 사전 학습 신경망의 성능과 실시예에서 제안된 태스크 점진 학습 방법을 통해 재생성한 신경망의 성능을 비교하는 실험이 진행될 수 있다.
- [0062] Iris 데이터셋은 붓꽃 식물의 종류를 나타내는 3개의 클래스 및 4개의 특성들로 이루어진 150개의 정형 데이터를 가진 데이터셋이며 실시예에서는 7:3의 비율로 학습 데이터와 평가 데이터로 구분하여 사용하기로 한다. 이는 분류 실험을 위한 가장 단순하고 적은 수의 학습 데이터로 이루어진 데이터셋에 해당한다. MNIST 데이터셋은 이미지 분류 실험에 보편적으로 사용되는 손 글씨 숫자 데이터셋으로 0부터 9까지 10종류의 숫자 이미지로 구성되며 60,000장의 학습 데이터와 10,000장의 테스트 데이터로 이루어져 있다.
- [0063] 이러한 각 데이터가 분류(classification) 실험에 사용될 수 있다. 각각의 데이터셋으로 베이지안 신경망을 학습시키고, 베이지안 신경망과 실시예에서 제안된 태스크 잠재 학습 방법으로 재생성한 신경망의 추론 정확도를 비교하는 실험이 진행될 수 있다. 학습에 사용한 베이지안 신경망은 각각 20개의 노드로 구성된 두 개의 은닉층을 가진 Bayes by Backprop(BBB)모델이 사용될 수 있다. 적대적 생성 신경망의 구조로는 MLP(multi-layer perceptron)으로 구성된 Vanila GAN가 사용될 수 있다.
- [0064] 두 신경망 간의 정확도를 비교하기에 앞서 실시예에서 제안된 태스크 점진 학습의 판별 신경망(discriminator)의 구분 결과를 확인해볼 수 있다. 도 5는 MNIST를 사용하여 실시예에서 제안된 태스크 점진 학습의 동작 시에 Epoch에 따른 판별 신경망의 추론 결과를 나타낸다. 적대적 생성 신경망의 판별 신경망은 입력으로 받은 데이터가 합성 데이터인지 실제 데이터인지 구분한 결과를 확률로 출력한다. 도 5에서 볼 수 있듯이 학습을 거듭할수록 판별 신경망의 구분 결과가 0.5에 수렴하는 것을 실시예에서 제안된 태스크 점진 학습 방법에서도 적대적 생성 학습이 이루어지며, 학습된 판별 신경망이 구분하지 못할 정도로 재생성된 신경망이 기존의 신경망과 유사함을 확인할 수 있다.

[0065] 표 1: Iris, MNIST 데이터에 대한 사전 학습 모델과 재생성 모델의 정확도

	Iris		MNIST	
	BNN	Synthetic	BNN	Synthetic
Acc	95.67	98.00	96.12	95.49
(std)	(6.1)	(2.21)	(0.1)	(0.07)

[0066]

[0067]

표 1에서는 실시예에서 제안된 태스크 점진 학습 방법을 통해 재생성된 신경망(Synthetic)과 사전 학습된 베이 지안 신경망(BNN)의 성능 비교를 확인할 수 있다. Iris 데이터셋의 경우 재생성된 신경망(Synthetic)의 정확도는 98.00%로 사전 학습 모델(BNN)의 성능 95.67%를 능가한다. MNIST 데이터셋의 경우 재생성된 모델 (Synthetic)의 정확도는 95.49%로 사전 학습 모델(BNN)의 성능 96.12%에 준하는 성능을 보인다. 이를 통해 실시예에서 제안된 신경망 재생성 기법(실시예에서 제안된 태스크 점진 학습 방법)을 사용하여 사전 학습된 신경망의 재생성이 가능함을 확인할 수 있다.

[0068]

이때 재생성된 신경망이 기존의 신경망보다 높은 성능을 보일 수 있던 이유는 적대적 생성 신경망을 통한 신경망 합성 시에 인풋 벡터 $z \sim N(0, 1)$ 가 가우시안 분포로부터 샘플링됨으로써 랜덤성(randomness)이 반영되어 사전 학습된 신경망보다 일반화 성능이 올라간 것으로 보인다. 사전 학습된 신경망과 재생성된 신경망을 비교하는 추가 실험으로 각각의 신경망의 파라미터 Norm을 비교하는 실험이 진행될 수 있다. 도 6에서는 사전 학습 모델의 파라미터와 재생성된 모델의 파라미터 간의 차이를 파악하기 위해 L2-Norm을 서로 비교한 결과를 보여준다.

[0069]

표 2: Split-MNIST에 대한 태스크 별 사전 학습 신경망(BNN)과 재생성된 신경망의 정확도

Task	1	2	3	4	5	Avg
BNN	99.9	99.9	99.2	99.2	99.7	99.6
1	99.9					99.9
2	99.9	99.9				99.9
3	99.9	99.9	99.3			99.7
4	99.9	99.9	99.1	99.5		99.6
5	99.9	99.9	99.3	99.4	99.7	99.6

[0070]

[0071]

실험 결과 사전 학습 모델과 재생성 모델의 L2-Norm이 서로 유사한 값을 가지는 것을 확인할 수 있다. 이를 통해 두 모델 간의 추론 결과가 유사한 것뿐만 아니라 실제 파라미터의 Norm 또한 유사하게 나타나는 것을 확인할 수 있다.

[0072]

다중 태스크에서의 파라미터 재생성을 확인하기 위해 태스크 점진 학습 벤치마크 실험에서 실시예에서 제안된 태스크 점진 학습 방법의 지속학습 성능이 확인될 수 있다. 첫 번째 벤치마크 실험 데이터로는 MNIST를 2개의 클래스씩 총 5개의 태스크로 분할한 Split-MNIST 데이터셋이 사용될 수 있다. 각 태스크마다 2개의 클래스가 주어지며, 검증 시에는 학습한 모든 클래스에 대한 추론 정확도가 측정될 수 있다. 이때, 태스크 점진적 학습(task-Incremental learning)을 가정하기로 한다. 태스크 점진 학습에서는 학습과 추론 단계에서 데이터가 몇 번 태스크에 해당하는지 알려주는 태스크 번호가 주어진다.

[0073]

태스크 점진 학습에 태스크 정보를 조건으로 받는 조건부 GAN(conditional GAN)가 사용될 수 있다. 부가적으로 태스크 점진 학습 시에 태스크를 거둬들수록 학습해야 하는 파라미터가 다양해지기 때문에 WGAN-GP 손실함수가 적용될 수 있다.

[0074]

표 2는 실시예에서 제안된 태스크 점진 학습 방법을 활용한 Split-MNIST 태스크 점진 학습 결과를 보여준다. 표 2의 행과 열은 각각 현재 학습하는 태스크, 추론하는 태스크를 의미한다. 예를 들어 (Task3, Task2)는

Task3까지 학습한 이후 Task2에 대한 추론 정확도를 의미한다. "BNN"이라고 표기된 행은 각 태스크에 대한 사전 학습 모델인 베이지안 신경망의 정확도이며, 이는 치명적 망각 현상 없이 실시예에서 제안된 태스크 점진 학습 방법을 통해 지속학습이 성공했을 경우 모방해야 할 각 태스크에 대한 정확도 수치를 의미한다.

[0075] 표 2에 따르면 태스크를 거듭하여 학습함에도 불구하고 정확도가 떨어지지 않는 것을 확인할 수 있다. Task1에 대한 정확도는 Task5를 학습한 뒤에도 99.9%를 유지 하고 있으며, 태스크를 거듭함에도 모든 태스크에 대한 정확도가 99%로 유지되고 있다. 아울러 (Task5, Task4) 과 같이 다음 태스크를 학습한 뒤에도 과거 태스크에 대해서 실시예에서 제안된 태스크 점진 학습 방법을 통해 재생성한 신경망(99.4%)이 사전 학습된 신경망(BNN)의 정확도(99.2%) 보다 높은 정확도를 보이는 경우 또한 존재한다. 이는 적대적 생성 신경망을 통한 파라미터 재생성 시에 랜덤성이 반영 되어 일반화 성능이 향상된 것으로 판단된다. 이를 통해 실시예에서 제안된 태스크 점진 학습 방법이 단일 신경망을 재생성하는 것뿐만 아니라 과거 여러 시점의 태스크에 해당하는 신경망을 재생성하는데 효과적이고, 치명적 망각 현상 없이 지속학습이 이루어 짐을 확인할 수 있다.

[0076] 두 번째 벤치마크 실험으로 Permuted-MNIST에 대한 태스크 점진 학습 실험이 진행될 수 있다. Permuted-MNIST는 MNIST 데이터셋의 변형으로 MNIST 이미지의 픽셀 순서를 다르게 하여 인위적으로 데이터를 증강시킨 형태의 데이터셋이다. 태스크가 클래스로 구분되는 Split-MNIST와 달리 Permuted-MNIST에서는 픽셀 순서를 다르게 한 규칙(permutation)에 의하여 태스크가 구분된다. 따라서 변경 규칙(permutation)에 따라서 태스크의 수가 늘어날 수 있다. 또한 Split-MNIST에 비하여 한 태스크 내 더 많은 클래스(0~9)를 가지고 있으며 이에 따라 학습하는 데이터 수 역시 증가한 더욱 복잡한 형태의 지속학습 벤치마크 실험에 해당한다. 이때, 실험에서는 태스크 수를 10개로 지정될 수 있다. Permuted-MNIST에 대한 베이지안 신경망으로는 각 200개의 노드로 구성된 두 개의 은닉층을 가진 Bayes by Backprop(BBB) 모델이 사용될 수 있다. 적대적 생성 신경망의 구조로는 Split-MNIST와 동일하게 WGAN-GP 손실함수를 사용하는 조건부 GAN이 사용될 수 있다.

[0077] 표 3: Permuted-MNIST에 대한 태스크 별 사전 학습 신경망(BNN)과 재생성된 신경망의 정확도

Task	1	2	3	4	5	6	7	8	9	10	Avg
BNN	97.73	97.73	97.60	97.67	97.57	97.62	97.94	97.80	97.72	97.62	97.70
1	98.03										98.03
2	98.08	98.04									98.06
3	98.05	97.96	97.76								97.92
4	98.00	97.92	97.70	97.86							97.87
5	98.12	97.95	97.84	97.83	97.60						97.87
6	97.99	98.03	97.75	97.88	97.72	97.76					97.86
7	97.98	97.90	97.76	97.86	97.72	97.70	98.09				97.86
8	97.92	97.81	97.72	97.88	97.73	97.69	98.11	97.99			97.86
9	97.88	97.85	97.76	97.79	97.68	97.64	98.05	97.89	97.93		97.83
10	97.86	97.77	97.73	97.86	97.67	97.63	98.06	97.90	97.91	97.81	97.82

[0078]

[0079] 표 3을 통해 알 수 있듯이 Permuted-MNIST에서도 Split-MNIST의 지속학습 양상이 일관되게 나타난다. 실시예에서 제안된 태스크 점진 학습 방법을 사용한 지속학습 결과에서는 태스크를 거듭함에도 치명적 망각 현상이 발견되지 않았고, 사전 학습 모델인 베이지안 신경망(BNN)의 성능에 준하거나 빈번하게 이를 상회하는 정확도를 보였다. 이를 통해 실시예에서 제안된 태스크 점진 학습 방법의 지속학습 효과는 더욱 복잡한 형태의 태스크 점진 학습에서도 일관되게 유지되었음을 확인할 수 있다.

[0080] Permuted-MNIST 데이터셋에 대하여 기존의 다른 태스크 점진 학습 방법과 실시예에서 제안된 태스크 점진 학습 방법의 태스크 점진 학습 성능을 비교하는 실험이 진행될 수 있다. 성능 비교 지표로는 모든 태스크에 걸친 정확도의 평균인 평균 정확도(Average accuracy)가 사용될 수 있다.

[0081] 표 4의 Fine-tuning은 별도의 지속학습 방법을 적용하 지 않고 다음 태스크의 데이터로 모델을 미세 조정시킨 방법에 해당한다. 각 방법은 태스크 점진 학습에서 치적 망각 현상을 완화하는 방법이며 그 중에서도 HAT 는 태스크 점진 학습에서 높은 성능을 보이는 것으로 알려져 있다.

[0082] 표 4: 10개 태스크 Permuted-MNIST에 대한 태스크 점진 학습 방법별 평균 정확도.

Approach	Avg Acc
Fine-tuning	0.87
EWC [6]	0.96
IMM-mean [16]	0.83
IMM-mode [16]	0.94
HAT [8]	0.97
ParameterGAN(ours)	0.98

[0083]

[0084]

표 4에서 알 수 있듯이 실시예에서 제안된 태스크 점진 학습 방법(ParameterGAN)이 다른 방법들과 비교했을 때 가장 높은 평균 정확도를 보이는 것을 확인할 수 있다. 도 7은 태스크가 거듭될수록 변 화하는 정확도 양상을 그래프로 나타낸 것이다. 미세 조정(fine-tuning) 방법을 포함한 여러 방법에서 태스크가 진행됨에 따라 정확도가 감소하는 치명적 망각 현상이 발생함을 확인할 수 있다. 이와 대조적으로 실시예에서 제안된 태스크 점진 학습 방법은 태스크를 거듭함에도 망각 없이 가장 높은 정확도를 유지하고 있다. 이를 통해 실시예에서 제안된 태스크 점진 학습 방법이 다른 방법과 비교하였을 때 더욱 높은 성능을 보임을 검증할 수 있다.

[0085]

이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0086]

소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0087]

실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0088]

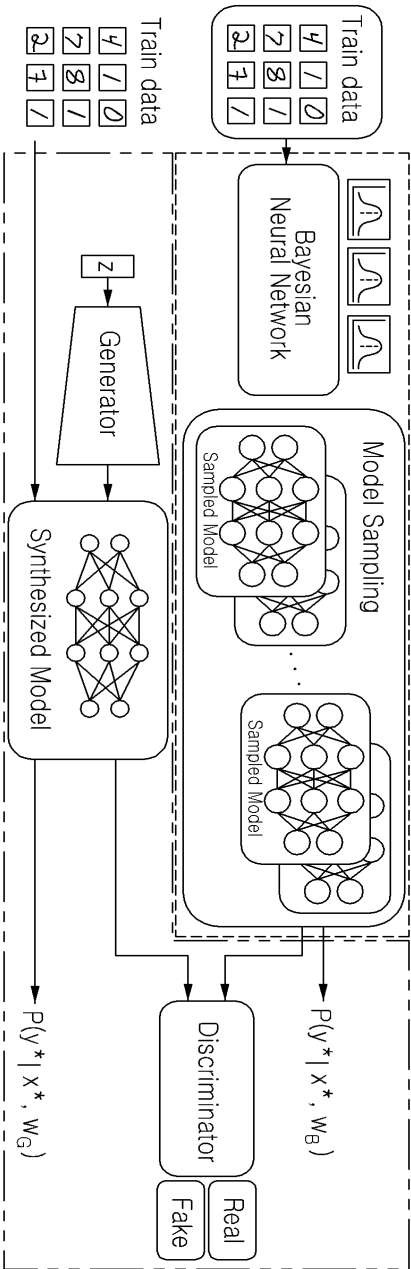
이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될

수 있다.

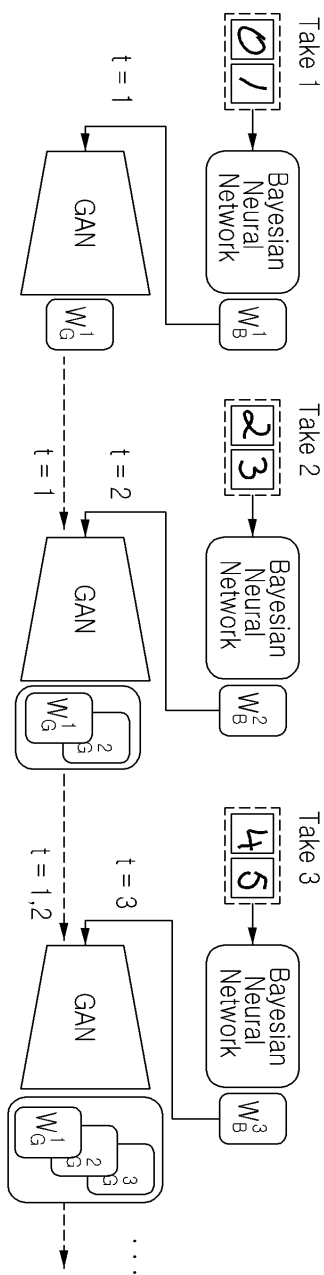
[0089] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

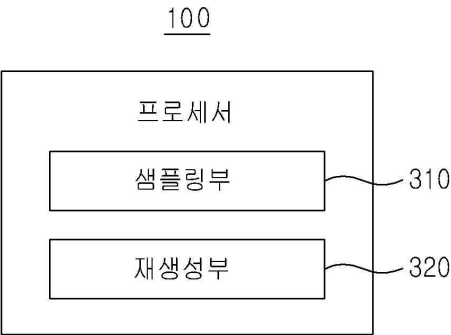
도면1



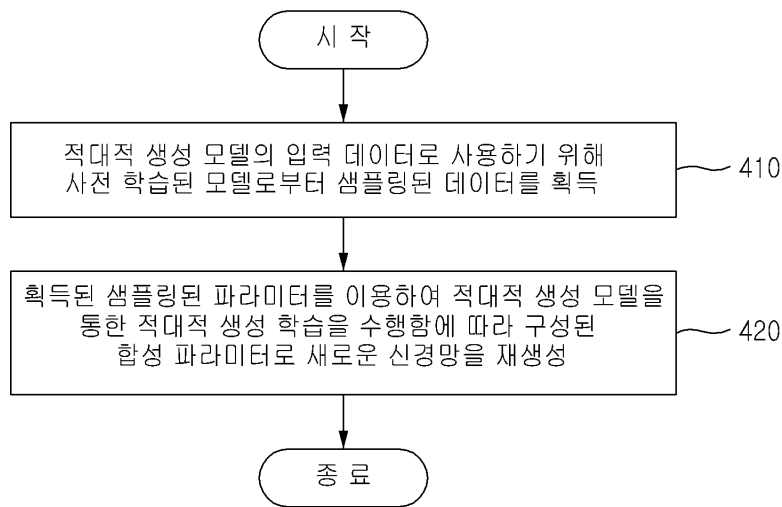
도면2



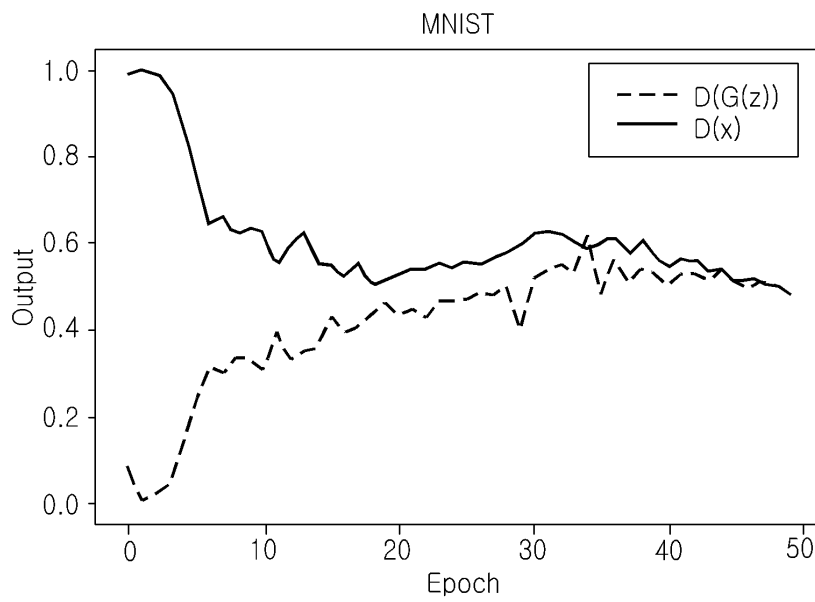
도면3



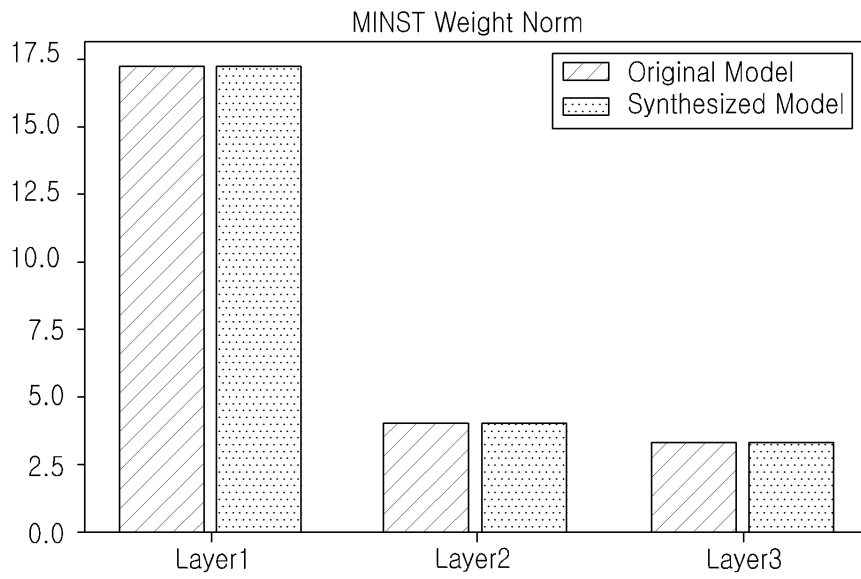
도면4



도면5



도면6



도면7

