(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년04월06일
(11) 등록번호 10-2097724
(24) 등록일자 2020년03월31일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
(21) 출원번호 10-2018-0131572
(22) 출원일자 2018년10월31일
심사청구일자 2018년10월31일
(56) 선행기술조사문헌
JP2015052832 A*
Lee, Sang-Woo, et al. "Overcoming catastrophic forgetting by incremental moment matching." Advances in neural information processing systems. 2017. (2017.12.09. 공개)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
신병석
서울특별시 서초구
홍대용
경기도 남양주시 진접읍 해밀예당1로 272, 2307동 1101호(신안인스빌23단지아파트)
이연
인천광역시 남구 인하로 100, 1410호(용현동, 인하대학교 하이테크)
(74) 대리인
양성보

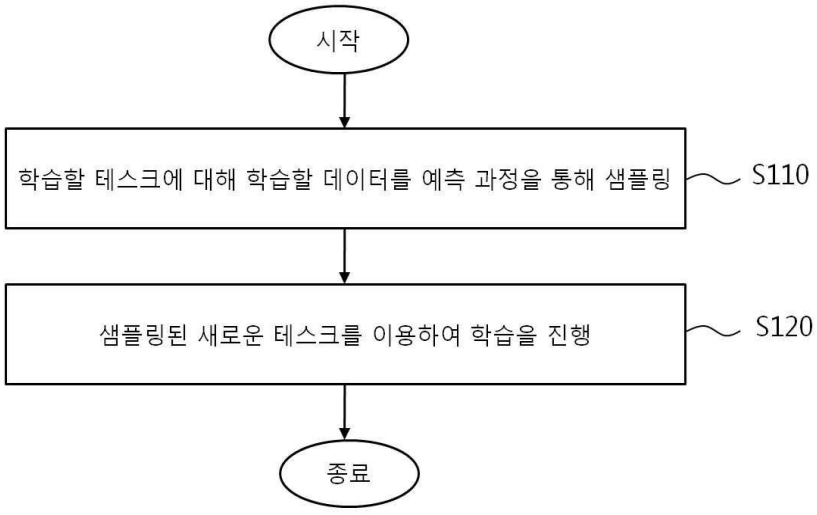
전체 청구항 수 : 총 5 항 심사관 : 박상현

(54) 발명의 명칭 **영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치**

(57) 요 약

영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치가 제시된다. 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법은, 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계; 및 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계를 포함하고, 상기 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계는, 상기 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트할 수 있다.

대 표 도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호	1711075658 (세부과제번호 2017-0-01815-002)
부처명	과학기술정보통신부
연구관리전문기관	정보통신기술진흥센터
연구사업명	국가전략프로젝트
연구과제명	(K-AR 1세부) AR기반 수술용 개발툴킷 및 응용개발
기 여 율	1/1
주관기관	서울대학교 산학협력단
연구기간	2018.05.01 ~ 2018.12.31

명 세 서

청구범위

청구항 1

영상 인식용 CNN에서의 최악망각현상의 완화 장치를 이용한 영상 인식용 CNN에서의 최악망각현상의 완화 방법에 있어서,

상기 영상 인식용 CNN에서의 최악망각현상의 완화 장치의 예측부에서 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계; 및

상기 영상 인식용 CNN에서의 최악망각현상의 완화 장치의 학습부에서 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계

를 포함하고,

상기 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계는,

상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬하는 단계; 및

정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하는 단계

를 포함하며,

상기 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계는,

상기 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트하는 것

을 특징으로 하는, 영상 인식용 CNN에서의 최악망각현상의 완화 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬하는 단계는,

현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정하는 단계; 및

상기 실수값에 따라 상기 이미지들을 내림차순으로 정렬하는 단계

를 포함하는, 영상 인식용 CNN에서의 최악망각현상의 완화 방법.

청구항 4

제1항에 있어서,

상기 정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하는 단계는,

현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차가 큰 이미지를 샘플링하여 상기 새로운 테스크를 생성하는 것

을 특징으로 하는, 영상 인식용 CNN에서의 최악망각현상의 완화 방법.

청구항 5

영상 인식용 CNN에서의 최악망각현상의 완화 장치에 있어서,

학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 예측부; 및

샘플링된 새로운 테스크를 이용하여 학습을 진행하는 학습부

를 포함하고,

상기 예측부는,

상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬한 후, 정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하며,

상기 학습부는,

상기 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트하는 것

을 특징으로 하는, 영상 인식용 CNN에서의 최악망각현상의 완화 장치.

청구항 6

삭제

청구항 7

제5항에 있어서,

상기 예측부는,

현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정하는 실수값 산정부;

상기 실수값에 따라 상기 이미지들을 내림차순으로 정렬하는 정렬부; 및

현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차가 큰 이미지를 샘플링하여 상기 새로운 테스크를 생성하는 샘플링부

를 포함하는, 영상 인식용 CNN에서의 최악망각현상의 완화 장치.

발명의 설명

기 술 분 야

[0001] 아래의 실시예들은 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치에 관한 것으로, 더욱 상세하게는 최악망각문제(CFP)를 완화하기 위해 새로운 이미지 중 기존에 학습된 것과 특성이 크게 차이가 나는 이미지만을 사용하여 학습을 진행하는 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 최근 딥러닝(Deep Learning) 관련 연구가 활발해짐에 따라 이미지 분류(image classification), 음성 인식(speech recognition), 얼굴 인식(face identification), 의미론적 세분화(semantic segmentation), VQA(visual question answering), Atari 게임(Atari games) 등 다양한 분야에서 CNN(Convolutional Neural Networks)이 활용되고 있다. 의료 영상 분야의 대표적인 사례로는 미세조정(finetuning)하여, 폐렴을 포함한 14개 병리를 감지할 수 있다.

[0003] 그러나 딥러닝에 주로 사용되는 인공신경망(Neural Networks)은 새로운 데이터 셋(data set)을 학습할 때 기존 데이터에 대한 정보를 상실하는 특성이 보이는 것으로 보고 되었으며, 이는 현재까지 해결되지 않았다. 이를 최악망각문제(Catastrophic Forgetting Problem: CFP)라 하며, CNN에서도 동일하게 발생한다. Abraham은 최악망각문제(CFP)의 원인을 안정성-가소성 딜레마(stability-plasticity dilemma)라 주장하였다. 안정성(stability)은 가중치의 안정성, 즉 뒤이어 들어오는 입력값에 따라 가중치 값이 크게 변하지 않고 천천히 수렴하는 정도

를 의미한다. 가소성(plasticity)은 가변성으로 입력값에 따라 가중치가 얼마나 빠르게 피팅(fitting)되는지를 의미한다. 네트워크는 안정성(stability)이 높을수록 이미 학습한 자료에 대해 보다 정확히 예측할 수 있는 능력을 가지며, 가소성(plasticity)이 높을수록 앞으로 학습할 새로운 자료에 대한 학습 능력이 증가한다. 두 가지 특성은 상충 관계(trade-off)에 있으므로 최악망각문제(CFP)가 발생하고 이를 해결하기 위한 연구가 꾸준히 진행되어 왔다.

[0004] 도 1은 일 실시예에 따른 최악망각문제의 발생을 설명하기 위한 도면이다.

[0005] 도 1의 (a)를 참조하면, 최악망각문제의 발생을 나타내는 것으로, 매 데이터 셋을 학습할 때마다 가중치가 조정되기 때문에 이전에 학습한 데이터 셋에 대한 분류 능력이 상실된다. 이 때 가중치를 업데이트하는 속도에 따라 가변성과 안정성이 결정되며, 가변성이 높을수록 최악망각문제(CFP)가 크게 발생한다.

[0006] 도 1의 (b)를 참조하면, 무작위 대입 공격(brute force)의 해결 방법을 나타내는 것으로, 최악망각문제(CFP)가 발생하지 않으려면 새로운 데이터 셋과 이미 학습한 모든 이미지 셋을 하나의 셋으로 합쳐 다시 학습을 진행해야 한다.

[0007] 이와 같이 최악망각문제(CFP)가 발생하지 않도록 하는 가장 단순한 방법은 새로운 이미지 셋(image set)이 입력되면 기존의 가중치를 버리고 새로운 가중치를 생성하는 것이다. 네트워크가 학습했던 이전의 가중치를 사용하지 않고, 새로운 이미지를 포함해 모든 이미지를 재학습한다.

[0008] 그러나, 이러한 방법은 이미지 셋이 추가됨에 따라 늘어난 테스크 수만큼 학습시간이 배가 되며 그 배수는 갈수록 늘게 된다. 또한 매번 독립적으로 새로운 학습이 진행되기 때문에, 시간과 메모리 측면에서 비효율적이다.

선행기술문헌

비특허문헌

[0009] Huszár (비특허문헌 0001) F(2017), "On quadratic penalties in elastic weight consolidation", arXiv:1712.03847.

발명의 내용

해결하려는 과제

[0010] 실시예들은 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치에 관하여 기술하며, 보다 구체적으로 최악망각문제(CFP)를 완화하기 위해 새로운 이미지 중 기존에 학습된 것과 특성이 크게 차이가 나는 이미지만을 사용하여 학습을 진행하는 기술을 제공한다.

[0011] 실시예들은 추가로 학습될 이미지 중 네트워크가 분류하기 어려운 이미지만을 샘플링하여 학습하는 예측 가능한 EWC(Predictive Elastic Weight Consolidation)를 제안함으로써, 학습할 테스크의 크기를 줄이는 한편, 모든 테스크에 이 과정을 반복함으로써 효율적인 학습을 수행할 수 있는 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0012] 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법은, 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계; 및 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계를 포함하고, 상기 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계는, 상기 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트할 수 있다.

[0013] 상기 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계는, 상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬하는 단계; 및 정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하는 단계를 포함할 수 있다.

[0014] 상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬하

는 단계는, 현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정하는 단계; 및 상기 실수값에 따라 상기 이미지들을 내림차순으로 정렬하는 단계를 포함할 수 있다.

[0015] 상기 정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하는 단계는, 현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차가 큰 이미지를 샘플링하여 상기 새로운 테스크를 생성할 수 있다.

[0016] 다른 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치는, 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 예측부; 및 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 학습부를 포함하고, 상기 학습부는, 상기 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트할 수 있다.

[0017] 상기 예측부는, 상기 학습할 테스크의 각 이미지에 대한 예측 결과와 상기 이미지의 주석을 통해 테스크 내 이미지들을 정렬한 후, 정렬된 상기 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성할 수 있다.

[0018] 상기 예측부는, 현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정하는 실수값 산정부; 상기 실수값에 따라 상기 이미지들을 내림차순으로 정렬하는 정렬부; 및 현재 네트워크의 각각의 상기 이미지에 대한 예측 결과인 예측값과 상기 이미지의 실제 주석값의 차가 큰 이미지를 샘플링하여 상기 새로운 테스크를 생성하는 샘플링부를 포함할 수 있다.

발명의 효과

[0019] 실시예들에 따르면 추가로 학습될 이미지 중 네트워크가 분류하기 어려운 이미지만을 샘플링하여 학습하는 예측 가능한 EWC(Predictive Elastic Weight Consolidation)를 제안함으로써, 학습할 테스크의 크기를 줄이는 한편, 모든 테스크에 이 과정을 반복함으로써 효율적인 학습을 수행할 수 있는 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치를 제공할 수 있다.

[0020] 실시예들에 따르면 최악망각문제(CFP)를 완화하기 위해 새로운 이미지 중 기존에 학습된 것과 특성이 크게 차이가 나는 이미지만을 사용하여 학습을 진행함으로써, 주요 가중치의 안정성은 유지한 채 그 외 가중치의 가변성을 높여 네트워크의 이미지 분류 능력을 향상시킬 수 있다.

도면의 간단한 설명

[0021] 도 1은 일 실시예에 따른 최악망각문제의 발생을 설명하기 위한 도면이다.

도 2는 일 실시예에 따른 EWC의 학습 구조를 설명하기 위한 도면이다.

도 3은 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법을 나타내는 흐름도이다.

도 4는 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치를 나타내는 블록도이다.

도 5는 일 실시예에 따른 예측 가능한 EWC의 학습 구조를 개략적으로 나타내는 도면이다.

도 6은 일 실시예에 따른 새로운 테스크를 생성하는 예측 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 첨부된 도면을 참조하여 실시예들을 설명한다. 그러나, 기술되는 실시예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 이하 설명되는 실시예들에 의하여 한정되는 것은 아니다. 또한, 여러 실시예들은 당해 기술분야에서 평균적인 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해서 제공되는 것이다. 도면에서 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0024] 인공신경망(Neural Networks)에서는 새로운 데이터 셋을 학습하는 과정을 반복함에 따라, 기존에 학습한 데이터 셋으로부터 얻어낸 특징을 분석하는 능력을 상실하는 현상이 나타난다. 이를 최악망각문제(Catastrophic Forgetting Problem: CFP)라 한다. CNN(Convolutional Neural Networks)을 이용하여 영상 분류를 할 때, 각 분류마다 나타나는 대표 특징(representative feature)은 네트워크가 예측 대상 이미지의 분류를 판단하는데 중요한 영향을 미치므로, 최악망각문제(CFP)는 매우 치명적이다.

[0025] 본 실시예에서 제안하는 예측 가능한 EWC(Predictive Elastic Weight Consolidation)는 추가로 학습될 이미지 중 네트워크가 분류하기 어려운 이미지만을 샘플링하여 학습할 수 있다. 이 때 샘플을 추출하는 기준은 네트워크의 예측값과 실제 주석값 차의 크기이다. 이를 통해 학습할 테스크의 크기를 줄이는 한편, 모든 테스크에 이 과정을 반복함으로써 효율적인 학습을 수행할 수 있다.

[0026] 아래에서 영상 인식용 CNN에서의 최악망각현상의 완화 방법 및 장치에 대해 상세히 설명한다.

[0028] CNN이 좋은 성능을 보이는 것은 저수준 특징(feature)이 아닌 중간수준 특징(feature)을 활용하는 데서 기인한다. 이를 위해 조정해야 할 가중치의 개수가 증가하기 때문에 많은 수의 주석 이미지가 필요하다. 따라서 대부분의 경우 매번 대규모 이미지 셋을 학습하는 대신 이미 학습을 마친 가중치로 네트워크를 초기화한 후 추가 학습을 진행한다. 이를 통해 생성될 새로운 가중치가 기존 가중치의 이미지 특징 인식 능력을 유지한 채 새로운 학습을 시작할 수 있다.

[0029] 이것을 전이 학습(transfer learning)]이라고 하며, 그 구현 방식으로는 특징 추출(feature extractor)과 미세조정(finetuning)이 있다. 특징 추출(feature extractor)은 가져온 가중치를 수정하지 않고 특징 인식 능력을 그대로 사용하며, 분석한 특징을 원하는 주석 별로 분류하는 마지막 단의 완전-연결(fully-connected)된 분류기만을 학습시킨다. 미세조정(finetuning)은 초기화는 같은 방식으로 하지만, 분류기를 포함한 모든 가중치를 새로 학습하는 이미지의 특성에 맞게 수정하며 학습을 진행한다. 특징 추출(feature extractor)은 수정할 파라미터의 양이 적어 메모리나 학습 시간 면에서는 효율성 높으나, 일반적으로 미세조정(finetuning)이 학습을 수행한 이미지에 더 잘 피팅(fitting) 되는 경향을 보인다.

[0030] 최악망각문제(CFP)를 해결하기 위한 연구는 크게 정규화(regularization), 앙상블(ensemble), 리허설(rehearsal), 듀얼-메모리(dual-memory), 스파스-코딩(sparse-coding) 의 다섯 가지로 분류할 수 있다. 정규화(regularization)은 가변성을 조절하는 조건을 학습 과정에서 최소화할 대상인 손실 함수(loss function)에 추가하였다. 앙상블(ensemble)은 각 학습 셋, 즉 각 테스크 마다 네트워크를 새로 생성하여 종합한 예측을 최종 결과로 한다. 리허설(Rehearsal) 기법은 네트워크에게 새로운 학습 셋을 줄 때, 이미 학습했던 셋의 일부를 함께 제공한다. 듀얼-메모리(dual-memory)는 안정성이 높은 네트워크와 가변성이 높은 네트워크를 따로 두어, 학습을 마친 후 둘의 결과를 종합한다. 스파스-코딩(sparse-coding)은 각각의 특징이 표현되는 수치의 범위를 겹치지 않게 함으로써, 서로 영향을 줄 가능성을 낮추고자 한 것이다. 각 기법은 서로 다른 아이디어를 사용했지만, 공통적으로 영상을 학습할 때 이미 학습한 테스크를 잘 표현하는 특징을 잘 보존하고자 하였다.

[0031] 앙상블(ensemble) 기법은 딥러닝의 초기부터 성능 향상을 위해 널리 이용되던 기법이다. 네트워크의 구조가 복잡해져 조정할 가중치 수가 많아지면 예측의 정확도는 올라가지만 계산량이 급증한다. 앙상블(ensemble)은 같은 이미지 셋에 대해, 가중치 수가 많은 하나의 복잡한 네트워크를 만드는 대신 여러 개의 단순한 네트워크를 독립적으로 학습시킨다. 그리고 각 네트워크의 결과를 종합하여 최종 예측을 수행함으로써 적은 컴퓨팅 파워로 정확한 예측할 수 있다. Ren은 앙상블(ensemble)의 아이디어를 이용해, 이미 학습한 테스크와 새로운 테스크를 독립적으로 학습한 뒤 각 네트워크의 예측을 종합함으로써 최악망각문제(CFP)을 완화하였다. 또한 단순히 새로운 테스크를 위한 학습을 진행하는데 그치지 않고, 생성된 새로운 가중치(weight)가 기존 가중치(weight)에 반영될지 여부를 판단하는 알고리즘을 추가로 제안하였다. 그러나 고해상도 이미지를 다루는 경우에는 컨볼루션(convolution)을 수행하는 부분에서 비용(cost)가 크다. 네트워크의 깊이와 너비를 감소시켜 얻는 계산량 절감 효과보다 입력 이미지의 해상도가 커짐에 따라 얻는 오버헤드가 크므로, 여러 개의 네트워크를 고해상도 이미지로 미세조정(finetuning)하는 것은 계산량 측면에서 적합하지 않다.

[0032] 리허설(Rehearsal) 기법은 네트워크의 구조를 변경하는 것이 아니라, 학습 자료(training dataset)를 재사용한다. 하나의 테스크를 학습한 뒤 다음 테스크를 학습하되, 이전에 학습한 테스크의 일부를 샘플링하여 새로운 테스크에 추가한 뒤 학습을 진행한다. 이처럼 복습하는 과정을 함께 함으로써 최악망각문제(CFP)를 방지한다. 그러나 이 기법은 샘플링을 위해서 지난 단계 학습에서 사용했던 이미지들을 모두 메모리에 가지고 있어야 한다는 단점이 있다. 이러한 비효율적인 자원(resource) 활용 문제를 개선하기 위해 Robins는 의사-리허설(pseudo-rehearsal) 기법을 제안했다. 이 기법은 이미지 자체를 재사용하지 않고, 기존 이미지를 분석하여 얻어낸 의사(pseudo) 패턴을 새로운 테스크에 추가한다. 패턴이 이미지 대신 들어가기 때문에, 메모리 상의 과부하를 줄였다.

[0034] 도 2는 일 실시예에 따른 EWC의 학습 구조를 설명하기 위한 도면이다.

[0035] 도 2를 참조하면, 기존의 EWC(Elastic Weight Consolidation)는 통계학에서 사용하는 피셔 정보(fisher

information) 개념을 도입하여, 새로운 테스크를 학습할 시 가중치 각 값이 어느 정도의 가변성을 가질지를 결정한다. 여기서, 피셔 정보는 어떤 사건이 일어났을 때, 사건의 관찰을 통해 얻어지는 정보의 양을 의미한다. 일어날 확률이 작은 사건일수록 얻을 수 있는 정보가 많기 때문에, 피셔 정보의 값은 사건이 일어날 확률이 작을수록 큰 값을 가진다. 하나의 테스크를 사건의 배경으로 보고, 그에 대해 학습을 마친 가중치 각 값을 독립사건으로 보아 피셔 정보를 적용할 수 있다. 이에 따라 해당 테스크로부터 현재 계산된 가중치가 나올 확률이 작을수록 피셔 정보는 큰 값을 가지게 된다.

[0036] EWC는 이처럼 피셔 정보를 반영한 손실(loss)을 손실 함수(loss function)에 추가함으로써 손실 일반화(loss regularization)를 수행했다. 이에 따라 가중치 값 중 피셔 정보 값이 큰 값일수록 업데이트가 많이 된다. 즉, 가변성이 높아진다.

[0037] 따라서 가중치가 이미 학습한 테스크로부터 영향을 받은 정도를 가중치의 가중치(Weight of Weight: WoW)로 생각할 수 있다. 주된 특성을 반영한 값일수록 가변성을 억제하는 과정을 통해 각 가중치의 가변성을 변경함으로써 하나의 네트워크가 안정성이 높은 네트워크와 가변성이 높은 네트워크를 앙상블(ensemble)한 것과 같은 결과를 만든다. 하나의 가중치 셋만을 저장하므로 메모리 측면에서 앙상블(ensemble)보다 효율적이며, 가중치마다 중요도가 다르다는 점에서 인공신경망이 추구하는 두뇌에 가까운 구조를 설계할 수 있다.

[0038] 도 2는 EWC의 학습 구조를 나타내고, 네트워크의 각 가중치는 피셔 정보를 가진다. 가중치의 가변성은 피셔 정보에 의존하기 때문에 피셔 정보를 가중치의 가중치로 생각할 수 있다. 새로운 테스크(D)가 들어옴에 따라 피셔 정보의 연산과 역전파(backpropagation)가 이루어져 가중치 셋(W)과 피셔 정보 셋(WoW)이 업데이트될 수 있다.

[0040] 도 3은 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법을 나타내는 흐름도이다.

[0041] 도 3을 참조하면, 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법은, 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계(S110), 및 샘플링된 새로운 테스크를 이용하여 학습을 진행하는 단계(S120)를 포함하여 이루어질 수 있다.

[0042] 여기서, 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링하는 단계(S110)는, 학습할 테스크의 각 이미지에 대한 예측 결과와 이미지의 주석을 통해 테스크 내 이미지들을 정렬하는 단계, 및 정렬된 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성하는 단계를 포함할 수 있다.

[0043] 또한, 학습할 테스크의 각 이미지에 대한 예측 결과와 이미지의 주석을 통해 테스크 내 이미지들을 정렬하는 단계(S120)는, 현재 네트워크의 각각의 이미지에 대한 예측 결과인 예측값과 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정하는 단계, 및 실수값에 따라 이미지들을 내림차순으로 정렬하는 단계를 포함할 수 있다.

[0044] 아래에서 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법을 하나의 예를 들어 설명하기로 한다.

[0046] 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 방법은 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치를 이용하여 보다 구체적으로 설명할 수 있다.

[0047] 도 4는 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치를 나타내는 블록도이다.

[0048] 도 4를 참조하면, 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치(100)는 예측부(110) 및 학습부(120)를 포함하여 이루어질 수 있다. 실시예에 따라 예측부(110)는 실수값 산정부(111), 정렬부(112) 및 샘플링부(113)를 더 포함할 수 있다. 또한 학습부(120)는 그래디언트 계산부(121), 피셔 정보 추출부(122) 및 업데이트부(123)를 더 포함할 수 있다.

[0049] 단계(S110)에서, 예측부(110)는 학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링할 수 있다.

[0050] 보다 구체적으로, 예측부(110)는 학습할 테스크의 각 이미지에 대한 예측 결과와 이미지의 주석을 통해 테스크 내 이미지들을 정렬한 후, 정렬된 이미지들 중 특정 이미지를 추출하여 새로운 테스크를 생성할 수 있다.

[0051] 여기서, 예측부(110)는 실수값 산정부(111), 정렬부(112) 및 샘플링부(113)를 더 포함할 수 있다.

[0052] 실수값 산정부(111)는 현재 네트워크의 각각의 이미지에 대한 예측 결과인 예측값과 이미지의 실제 주석값의 차를 통해 하나의 이미지당 하나의 실수값을 결정할 수 있다.

[0053] 정렬부(112)는 실수값에 따라 이미지들을 내림차순으로 정렬할 수 있다.

[0054] 그리고 샘플링부(113)는 현재 네트워크의 각각의 이미지에 대한 예측 결과인 예측값과 이미지의 실제 주석값의 차가 큰 이미지를 추출한 후 샘플링하여 새로운 테스크를 생성할 수 있다.

[0055] 단계(S120)에서, 학습부(120)는 샘플링된 새로운 테스크를 이용하여 학습을 진행할 수 있다.

[0056] 보다 구체적으로, 학습부(120)는 새로운 테스크가 전달됨에 따라 그래디언트(gradient) 및 피셔 정보(fisher information)를 계산하여 가중치 셋과 피셔 정보 셋을 업데이트할 수 있다.

[0057] 이러한 학습부(120)는 그래디언트 계산부(121), 피셔 정보 추출부(122) 및 업데이트부(123)를 더 포함할 수 있다.

[0058] 그래디언트 계산부(121)는 새로운 테스크가 전달됨에 따라 그래디언트를 계산할 수 있다.

[0059] 피셔 정보 추출부(122)는 새로운 테스크가 전달됨에 따라 피셔 정보를 계산할 수 있다.

[0060] 업데이트부(123)는 그래디언트 계산부(121) 및 피셔 정보 추출부(122)로부터 계산 결과를 전달받아 가중치 셋과 피셔 정보 셋을 업데이트할 수 있다.

[0061] 이와 같이, 실시예들은 최악망각문제(CFP)를 완화하기 위해 손실 일반화(loss regularization)를 통한 전이 학습(transfer learning) 기법인 EWC를 확장하여, 새로운 이미지 중 기존에 학습된 것과 특성이 크게 차이가 나는 이미지만을 사용하여 학습을 진행하는 기법을 제공할 수 있다.

[0062] EWC는 가중치 각각에 대해, 대표 특징을 인식하는데 관여한 정도가 클수록 가변성을 낮춘다. 이는 영상 분류에 있어 중요한 특징을 잘 잡아내는 가중치를 덜 조정되도록 하여 최악망각문제(CFP)를 완화하고 네트워크의 인식 능력을 보존할 수 있다. 이미 학습한 이미지 셋에 관여한 정도를 각 가중치마다 가진 채, 새로 입력된 이미지 셋을 모두 학습한다.

[0063] 일 실시예에 따른 영상 인식용 CNN에서의 최악망각현상의 완화 장치(100)(아래에서는, 간단히 "예측 가능한 EWC(100)"라고 하기로 한다.)는 가중치 별로 가변성을 조정하는 과정을 채택하나, EWC와 달리 새로운 이미지 셋을 입력 받는 즉시 학습하지 않는다. 대신 학습할 테스크의 이미지 중, 해당 시점에 네트워크가 인식해내기 어려운 특징을 많이 가진 이미지들을 추출하여 이 샘플링 셋을 학습할 테스크로 정한다. 이 때 EWC를 이용했기 때문에 이미지 분류에 크게 관여한 가중치는 안정성을 유지한다.

[0064] 추가적으로, 일 실시예에 따른 예측 가능한 EWC는 학습할 이미지를 샘플링하여 테스크를 대체함으로써 이미지 분류에 관여한 정도가 작아 가변성이 높은 가중치를 더 중점적으로 학습시킨다. 또한 기존의 EWC와 달리 모든 이미지를 다 사용하지 않으므로 학습할 이미지 수가 감소하여 학습 시간이 단축된다.

[0065] 따라서 일 실시예에 따른 예측 가능한 EWC는 어려운 이미지만을 학습함으로써 주요 가중치의 안정성은 유지한 채 그 외 가중치의 가변성을 높여 네트워크의 이미지 분류 능력을 향상시킬 수 있다.

[0066] 실험 결과, 샘플링하지 않은 EWC 학습의 1회 수행시간과 샘플링 비율을 0.5로 한 예측 가능한 EWC 학습이 3회 수행되는 시간이 비슷했다. 두 실험에서 AUROC 측정 결과 가장 낮은 AUROC 값이 EWC는 약 0.3인데 반해 예측 가능한 EWC는 약 0.6을 기록했다. 또한, 테스크를 나누지 않고 모든 이미지를 미세조정(finetuning)한 학습에서 가장 높은 AUROC가 약 0.6 것을 확인하였다. 이로써 샘플링을 통해 동일한 양의 메모리 리소스를 사용하면서도 인식 정확도가 2배 향상되었음을 확인하였다.

[0067] 아래에서는 제안하는 예측 가능한 EWC의 구조를 서술하고, 실험의 구체적인 요소들과 그 결과를 설명하기로 한다.

[0069] 일 실시예에 따른 예측 가능한 EWC은 리허설(rehearsal) 및 의사-리허설(pseudo-rehearsal)의 샘플링 개념과 EWC 기법의 피셔 정보를 이용할 수 있다. 기존에 샘플링을 이용하는 기법은, 학습에 이용되었던 이미지들 자체 또는 그 이미지들의 특징을 반영하여 생성된 의사-패턴(pseudo-pattern)을 메모리에 가지고 있어야 했다. 이러한 메모리 측면의 비효율성에도 불구하고 네트워크의 성능이 향상되지만 최악망각문제(CFP) 완화에 있어 타 기법들보다 뛰어난 개선을 보이지 못했다. 반면, 일 실시예에 따른 예측 가능한 EWC은 리허설(rehearsal) 기법과 같이 이미 학습에 사용되었던 테스크로부터 샘플링하지 않는다.

[0070] 도 5는 일 실시예에 따른 예측 가능한 EWC의 학습 구조를 개략적으로 나타내는 도면이다.

[0071] 도 5를 참조하면, 일 실시예에 따른 예측 가능한 EWC(100)에서 각 셋 D는 각 테스크를 의미하며, D'은 해당 테스크 D에서 네트워크의 예측과 샘플링을 통해 만들어진 새로운 테스크를 의미한다. 실시예에 따르면 예측부

(110)에서의 예측 과정을 통해 작아진 학습 셋으로, 학습부(120)에서 그래디언트 및 피셔 정보를 계산해 업데이트를 수행할 수 있다.

[0072] 일 실시예에 따른 예측 가능한 EWC(100)는 앞으로 학습할 테스크에 대해 학습할 데이터를 샘플링하는 과정을 수행하기 때문에 이미지를 메모리에 저장해둘 필요가 없다. 이를 위해 앞으로 학습할 테스크가 학습되기 이전에, 예측부(110)에서 해당 테스크를 검사용 데이터 셋으로 가정하고 예측 과정을 먼저 수행할 수 있다. 그리고 각 이미지에 대한 예측 결과와 이미지의 주석을 통해 테스크 내 이미지를 정렬한 뒤 새로운 테스크를 생성하고, 학습부(120)에서 이 새로운 테스크로 본래 테스크를 대체하여 학습을 진행할 수 있다. 이 때 정렬 기준은 각 이미지에 대한 네트워크의 예측값과 실제 주석값의 차의 L1-norm으로, 다음 식과 같이 표현될 수 있다.

[0073] [수학식 1]

[0074] $$\sum_{j=1}^{C} | f(\mathbf{x}_k)^j - \mathbf{y}_k^j |$$

[0075] [수학식 2]

[0076] $$\text{for } D_i = \{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{n_i} : \text{task}, \ i = 0..N-1$$

[0077] [수학식 3]

[0078] $$\mathbf{x}_k \in \mathbb{R}^d : an\ image$$

[0079] [수학식 4]

[0080] $$\mathbf{y}_k \in \mathbb{R}^C : label$$

[0081] [수학식 5]

[0082] $$f : \mathbb{R}^d \rightarrow \mathbb{R}^C \quad : network$$

[0083] 여기서, [수학식 2]는 하나의 테스크를 뜻하며, 그를 구성하는 한 원소는 한 장의 이미지[수학식 3]와 그에 대응하는 주석[수학식 4]의 쌍이다. n은 테스크 [수학식 2]의 원소의 개수, 즉 각 테스크의 크기를 뜻한다. C는 학습시키는 네트워크[수학식 5]에 주어진 분류(class)가 몇 개인지를 나타낸다. 분류기(classifier)의 마지막 단에서 시그모이드(sigmoid) 함수를 적용하기 때문에, 네트워크를 통과한 이미지의 예측값은 0에서 1사이 실수값들로 이루어진 벡터이다. 주석값은 각 분류에 대해 음성 또는 양성에 해당하는지에 따라 0 또는 1로 이루어진 길이 C인 벡터이다.

[0084] 그러므로 [수학식 1]에 따라 하나의 이미지에 대해 분류 별 네트워크의 예측값과 주석값의 차의 절대값들이 계산되고, 그를 합산하여 이미지 한 장당 하나의 양의 실수값이 계산된다. 이 때, 실수값이 클수록 네트워크의 예측값과 실제 주석값의 차이가 크므로 norm 값이 클수록 네트워크가 인식하기 어렵다고 해석할 수 있다.

[0085] 도 6은 일 실시예에 따른 새로운 테스크를 생성하는 예측 과정을 설명하기 위한 도면이다.

[0086] 도 6을 참조하면, 예측부(110)에서의 새로운 테스크 생성 과정을 나타내는 것으로, D에 있는 이미지를 네트워크로 예측한 결과와 그에 대응되는 주석에 1-norm을 취하여 정렬하고, 값이 큰 순으로 이미지를 추출할 수 있다.

[0087] 보다 구체적으로, 각 이미지 별로 L1-norm을 구한 후, 이를 이용해 테스크 내 이미지를 내림차순으로 정렬하고, 정해진 샘플링 비율에 따라 상위 이미지부터 샘플링하여 새로운 테스크를 생성할 수 있다. 결과로 생성되는 새로운 테스크는 본래 테스크보다 크기가 작으며, L1-norm 값이 큰 이미지만 추출하기 때문에 네트워크에게 분류하기 가장 어려운 이미지들의 셋에 해당된다.

[0088] 분류 별 네트워크의 예측값과 분류 별 주석값의 차의 절대값은 항상 0에서 1 사이이다. 따라서 L2-norm을 취하게 되면 제곱 연산을 통해 norm 값이 본래 차의 절대값보다 작아진다. 여기서 norm의 역할은 예측과 주석 사이의 거리(distance)를 측정하여 해당 이미지의 난이도를 판별하는 것이므로, 제곱의 합보다 절대값의 합이 보다 명확한 기준이 된다. 그러므로 L2-norm이 아닌 L1-norm을 기준으로 정렬 및 샘플링을 진행할 수 있다.

[0089] 이러한 과정을 수행하기 위해서는 새로운 테스크를 생성하는 단계에서 네트워크가 이미지를 예측하는 과정과 이

미지를 정렬하는 추가 비용이 든다. 그러나 예측 과정은 입력 이미지와 가중치 간의 선형적인 연산과 활성화 함수 통과 과정이며, 정렬과정은 양의 실수값들의 단순 내림차순으로, 복잡한 연산을 요구하지 않는다. 따라서 샘플링 비율을 1로 설정하여 새로운 테스크의 모든 이미지를 사용하더라도 기존 EWC와 총 학습시간에 있어 큰 차이가 나지 않는다.

[0090] 반면, 샘플링을 수행하면 이미지 수의 감소하고 이로 인해 가중치 업데이트를 위한 그래디언트(gradient)의 계산량이 대폭 줄기 때문에, 예측 및 정렬 과정으로 인해 추가되는 비용보다 큰 비용 절감 효과가 있다. 따라서 일 실시예에 따른 예측 가능한 EWC가 기존 EWC보다 총 학습시간이 적게 걸린다.

[0091] 학습을 위해 입력되는 새로운 테스크를 작은 크기의 테스크로 대체하면 네트워크가 학습하는 특징의 수가 줄어든다. 앞에서 설명한 도 1의 (a)에 도시된 바와 같이 최악망각문제(CFP)를 고려하지 않은 기본적인 형태에서는 학습하는 특징의 수가 줄면 새로운 이미지에 대한 학습도 부족해질 뿐 아니라, 최악망각문제(CFP)도 발생해 네트워크가 이미 가진 이미지 분류 능력도 잃는다.

[0092] 그러나 일 실시예에 따른 예측 가능한 EWC는 기존 EWC을 이용해 이미 가진 분류 능력을 보존하고, 테스크 생성을 통해 아직 학습이 부족한 특징만을 더 학습할 수 있다. EWC로부터 채택해온 피셔 정보는 이미 학습한 이미지의 특징 인식에 주요한 역할을 한 가중치의 안정성(stability)을 높여, 해당 가중치의 학습 속도를 늦추는 역할을 한다. 따라서 피셔 정보를 이용하는 네트워크가 새로운 테스크를 학습할 때, 어떤 가중치가 가변성이 높다면 이 가중치는 해당 시점에 잘 인식하는 이미지 특징과 연관성이 비교적 낮은 가중치이다. 일 실시예에 따른 예측 가능한 EWC는 학습시킬 새로운 테스크를 예측과 주석의 거리가 가장 먼 이미지로 구성한다.

[0093] 따라서 일 실시예에 따른 예측 가능한 EWC는 네트워크가 인식하는 능력이 부족한 특징이 풍부한 이미지를, 인식 능력에 기여하는 정도가 작은 가중치에 학습시킬 수 있다. 이로써 기존 EWC만큼 새로운 테스크 학습 이전의 분류능력을 보존하면서, 어려운 이미지만을 통해 효율적으로 학습할 수 있다.

[0095] 이상과 같이, 인공신경망은 네트워크가 새로운 데이터로 인해 기존 데이터의 특징(feature)을 인지하는 능력을 소실하는 문제인 최악망각(catastrophic forgetting)을 겪는다. 리허설(rehearsal)은 학습 자료의 재사용으로 최악망각문제(CFP)를 완화시켰고, EWC는 손실 제한(loss constraint)을 통해 가중치(weight) 내부 노드 별 업데이트 속도의 조절하여 최악망각문제(CFP)를 해결하려 했다.

[0096] 본 실시예들은 리허설(rehearsal)의 샘플링 방식과 EWC의 손실 제한(loss constraint) 기법에 기반하지만, 이미 학습을 한 테스크로부터 학습 자료를 샘플링하여 재사용하는 방식은 메모리상 과부하가 생기므로, 학습했던 테스크에서가 아닌 앞으로 학습할 테스크에서 진행하였다. 이 때 현재 네트워크의 이미지에 대한 예측값과 실제 주석의 차이가 큰 이미지를 샘플링 함으로써, 네트워크의 분류 능력이 저조한 특징이 풍부한 이미지를 중점적으로 학습하였다.

[0097] 실험은 전체 셋의 미세조정(finetuning), 테스크별 전이 학습(transfer learning), EWC, 예측 가능한 EWC의 4가지 케이스로 진행되었다. 미세조정(finetuning)을 제외한 다른 케이스의 worst AUROC(Area Under Receiver Operating Characteristic)는 약 0.3, 제안 기법의 worst AUROC는 약 0.6으로 향상된 결과를 확인하였다. 즉, 제안된 예측 가능한 EWC 기법은 worst AUROC가 약 0.6으로, 기존 EWC의 0.3보다 향상되었다.

[0099] 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, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable 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 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0100] 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.

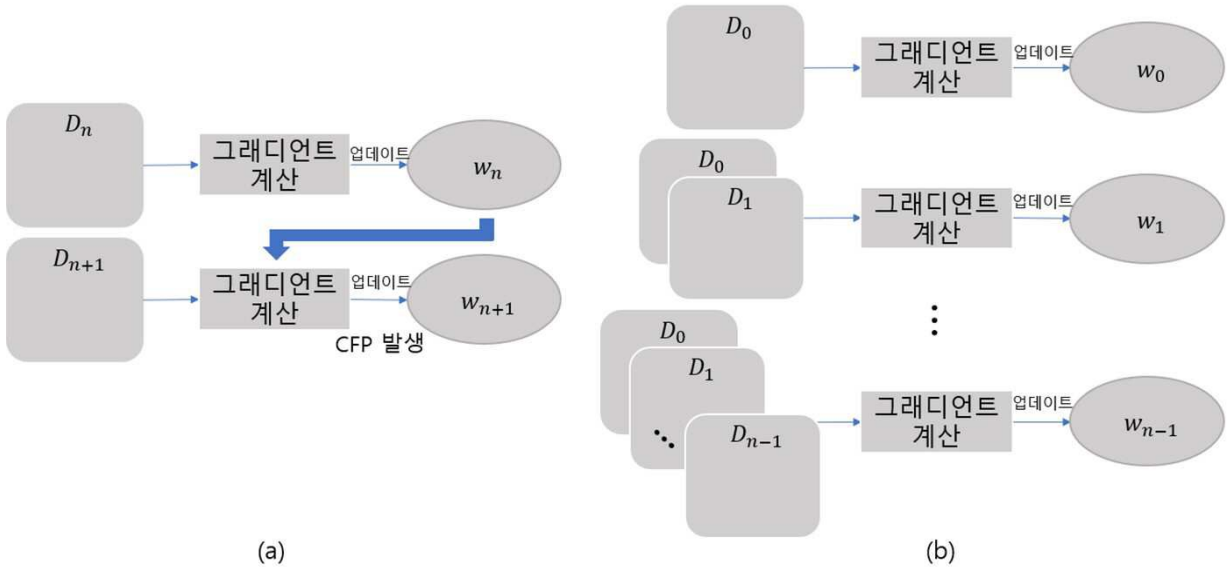
[0101] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.

[0102] 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.

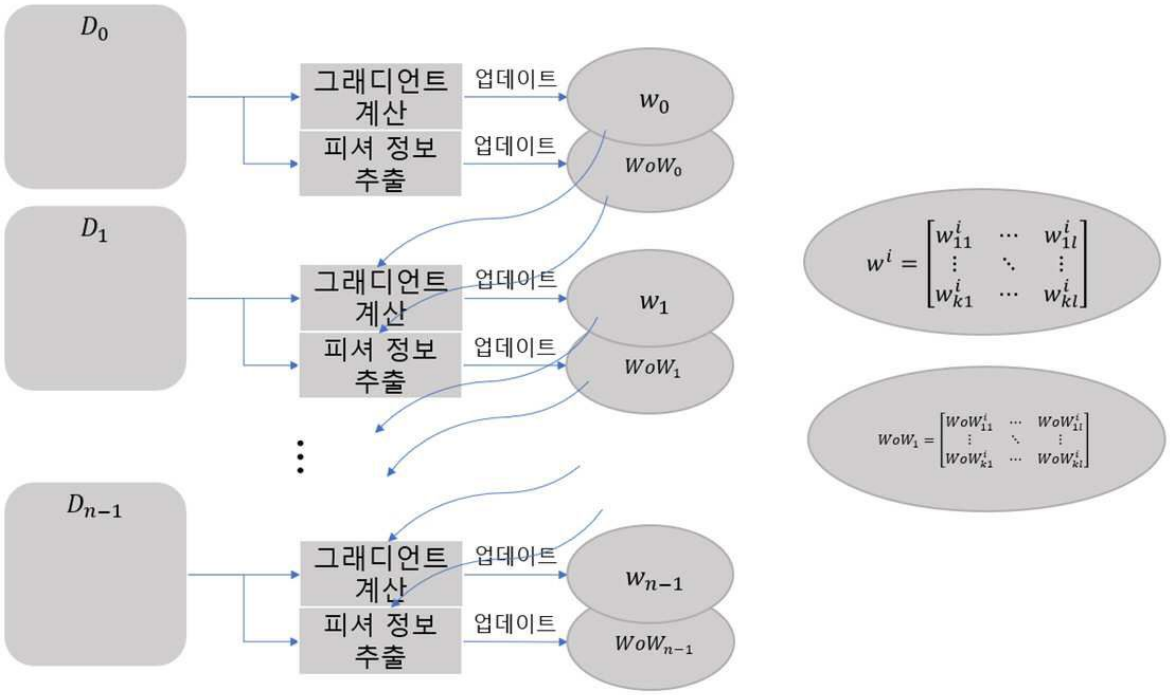
[0103] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

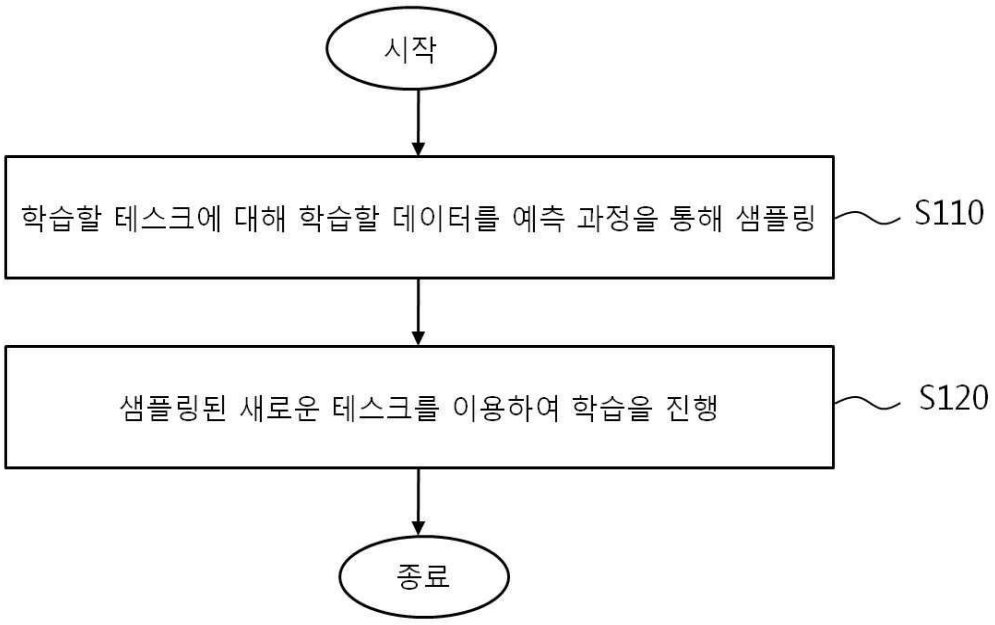
도면1



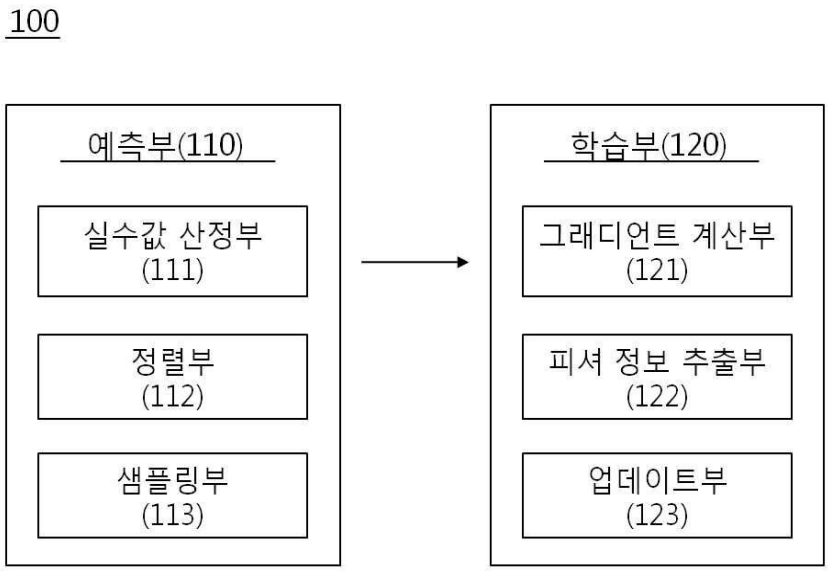
도면2


D_0
그래디언트 계산
업데이트
w_0
피셔 정보 추출
업데이트
WoW_0
D_1
그래디언트 계산
업데이트
w_1
피셔 정보 추출
업데이트
WoW_1
D_{n-1}
그래디언트 계산
업데이트
w_{n-1}
피셔 정보 추출
업데이트
WoW_{n-1}
$w^i = \begin{bmatrix} w_{11}^i & \cdots & w_{1l}^i \\ \vdots & \ddots & \vdots \\ w_{k1}^i & \cdots & w_{kl}^i \end{bmatrix}$
$WoW_1 = \begin{bmatrix} WoW_{11}^i & \cdots & WoW_{1l}^i \\ \vdots & \ddots & \vdots \\ WoW_{k1}^i & \cdots & WoW_{kl}^i \end{bmatrix}$

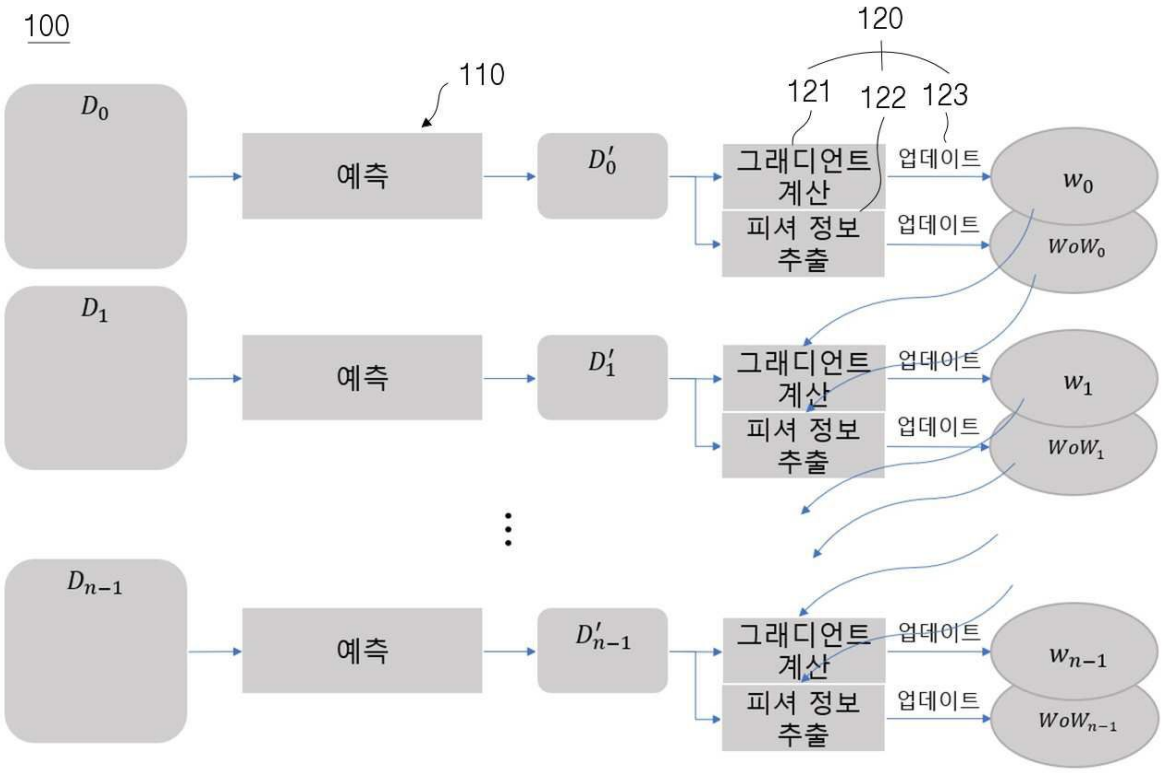
도면3


시작
학습할 테스크에 대해 학습할 데이터를 예측 과정을 통해 샘플링
S110
샘플링된 새로운 테스크를 이용하여 학습을 진행
S120
종료

도면4


100
예측부(110)
실수값 산정부
(111)
정렬부
(112)
샘플링부
(113)
학습부(120)
그래디언트 계산부
(121)
피셔 정보 추출부
(122)
업데이트부
(123)

도면5


100
120
110
121
122
123
D_0
예측
D_0'
그래디언트 계산
피셔 정보 추출
업데이트
업데이트
w_0
WoW_0
D_1
예측
D_1'
그래디언트 계산
피셔 정보 추출
업데이트
업데이트
w_1
WoW_1
D_{n-1}
예측
D_{n-1}'
그래디언트 계산
피셔 정보 추출
업데이트
업데이트
w_{n-1}
WoW_{n-1}

도면6

