(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0013438
(43) 공개일자 2024년01월30일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) *G06N 3/04* (2023.01)
(52) CPC특허분류
G06N 3/084 (2023.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2022-0090968
(22) 출원일자 2022년07월22일
심사청구일자 2022년07월22일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
박상우
서울특별시 서대문구 증가로 191, 107동 502호
서태원
경기도 남양주시 순화궁로 18, 4110동 902호
(74) 대리인
김등용

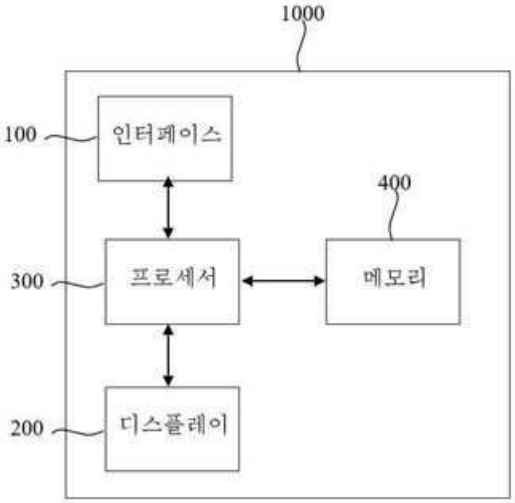
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **예측적 역전파 및 활성화 기록을 기반으로 하는 평생학습 장치 및 방법**

(57) 요 약

딥러닝 모델에 대한 평생학습 방법이 개시된다. 상기 딥러닝 모델의 평생학습 방법은 적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되고, 제1 태스크에 대하여 학습된 딥러닝 모델에 대한 제2 태스크 그리고 제n 태스크에 대한 평생학습을 위하여, 순전파 동작을 수행하는 순전파 단계, 역전파 동작을 수행하는 역전파 단계, 및 가중치 업데이트 동작을 수행하는 가중치 업데이트 단계를 포함하고, 상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 반복적으로 수행되고, 상기 업데이트 단계는 상기 딥러닝 모델에 포함된 뉴런들 각각의, 상기 제1 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향에 기초하여 수행된다.

대 표 도 - 도3



이 발명을 지원한 국가연구개발사업
과제고유번호 1711152774
과제번호 2019-0-00533-004
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보보호핵심원천기술개발(정진기금)
연구과제명 컴퓨터 프로세서의 구조적 보안 취약점 검증 및 공격 탐지대응
기 여 율 1/2
과제수행기관명 고려대학교산학협력단
연구기간 2022.01.01 ~ 2022.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711156705
과제번호 2022R1A2C1011469
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 Side-channel attack으로부터 안전한 IoT 디바이스 向 NPU (Neural Processing Unit) 연구 및 개발
기 여 율 1/2
과제수행기관명 고려대학교산학협력단
연구기간 2022.03.01 ~ 2023.02.28

명 세 서

청구범위

청구항 1

적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되는 딥러닝 모델의 평생학습 방법에 있어서,

제1 태스크에 대하여 학습된 딥러닝 모델에 대한 제2 태스크 그리고 n 태스크에 대한 평생학습을 위하여,

순전파 동작을 수행하는 순전파 단계;

역전파 동작을 수행하는 역전파 단계; 및

가중치 업데이트 동작을 수행하는 가중치 업데이트 단계를 포함하고,

상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 반복적으로 수행되고,

상기 업데이트 단계는 상기 딥러닝 모델에 포함된 뉴런들 각각의, 상기 제1 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향에 기초하여 수행되는,

딥러닝 모델의 평생학습 방법.

청구항 2

제1항에 있어서,

상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 상기 딥러닝 모델의 가중치들이 미리 정해진 범위로 수렴할 때까지 반복적으로 수행되는,

딥러닝 모델의 평생학습 방법.

청구항 3

제1항에 있어서,

상기 가중치 업데이트 단계는,

상기 제1 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향이 미리 정해진 값보다 큰 값을 갖는 뉴런에 대한 가중치 업데이트를 제한하는,

딥러닝 모델의 평생학습 방법.

청구항 4

제1항에 있어서,

상기 가중치 업데이트 단계는,

기울기(gradient)를 획득하는 단계;

상기 제1 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향이 미리 정해진 값보다 큰 값을 갖는 뉴런에 대하여,

획득된 기울기에 미리 정해진 상수(r)를 곱하여 가중치를 감소시키는 단계; 및

감소된 가중치를 이용하여 가중치를 업데이트하는 단계를 포함하는,

딥러닝 모델의 평생학습 방법.

청구항 5

제1항에 있어서,

상기 제2 태스크에 대한 학습 과정 중, 반복적으로 수행되는 순전파 단계와 역전파 단계는 적어도 일 회 이상

병렬적으로 진행되는,

딥러닝 모델의 평생학습 방법.

청구항 6

제5항에 있어서,

병렬적으로 진행되는 역전파 단계는, 적어도 이전 시점의 순전파 단계의 결과에 기초하여 현재 시점의 순전파 결과를 추론하고, 추론된 결과에 기초하여 역전파를 수행하는,

딥러닝 모델의 평생학습 방법.

청구항 7

적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되는 딥러닝 모델의 학습 방법에 있어서,

순전파 동작을 수행하는 순전파 단계;

역전파 동작을 수행하는 역전파 단계; 및

가중치 업데이트 동작을 수행하는 가중치 업데이트 단계를 포함하고,

상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 반복적으로 수행되고,

최초 실행 이후에 반복적으로 수행되는 순전파 단계와 역전파 단계는 적어도 일 회 이상 병렬적으로 진행되는,

딥러닝 모델의 학습 방법.

청구항 8

제7항에 있어서,

t=i 시점에서 병렬적으로 수행되는 역전파 단계는 t=(i-1) 시점에서의 순전파 결과와 t=(i-2) 시점에서의 순전파 결과를 이용하여 예측된 t=i 시점에서의 순전파 결과를 이용하여 수행되는,

딥러닝 모델의 학습 방법.

청구항 9

제8항에 있어서,

예측된 t=i 시점에서의 순전파 결과는 t=(i-1) 시점에서의 딥러닝 모델의 결과 보다 t=(i-2) 시점에서의 딥러닝 모델의 결과에 더 큰 가중치를 부여하여 예측되는,

딥러닝 모델의 학습 방법.

청구항 10

제8항에 있어서,

t=i 시점에서 각 뉴런들의 활성화 여부는 t=i-1 시점까지의 딥러닝 모델에 포함된 뉴런들 각각의 활성화 경향에 기초하여 예측되는,

딥러닝 모델의 학습 방법.

청구항 11

제7항에 있어서,

제1 태스크에 대한 학습이 완료된 후에 수행되는 제2 태스크에 대한 학습 과정에서,

상기 업데이트 단계는 상기 딥러닝 모델에 포함된 뉴런들 각각의, 상기 제1 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향에 기초하여 수행되는,

딥러닝 모델의 학습 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 인공지능에 관한 것으로, 보다 상세하게는 딥러닝(deep learning)의 평생학습(continual learning, life-long learning) 성능을 향상시키는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 딥러닝의 평생학습(continual learning, life-long learning)이란 하나의 딥러닝 모델이 점진적으로 다양한 도메인의 데이터를 지속적으로 학습하는 것을 목표로 한다. 하지만, 학습된 모델이 다른 도메인의 데이터를 학습할 때 기존에 학습했던 지식을 잊어버리게 되는 치명적 망각(catastrophic forgetting)이라는 문제가 존재한다.

[0003] 이와 같은 치명적 망각의 문제를 해결하기 위해 과거에 학습했던 도메인에 중요한 역할을 하는 모델 가중치를 찾아, 다른 도메인 학습시 해당 가중치가 업데이트되지 않도록 하는 EWC(Elastic Weight Consolidation) 기법이나 SI(Synaptic Intelligence) 기법 같은 방법론들이 제안되었다. EWC나 SI 기법은 효과적으로 치명적 망각을 완화하여 평생학습에서 state-of-the-art의 기준이 되었다.

[0004] EWC 기법은 Fisher information matrix를 기준으로 도메인에서 중요한 가중치를 찾고, SI 기법은 모델의 전체 gradients를 기준으로 도메인에서 중요한 가중치를 찾는다. 하지만, Fisher information matrix나 모델의 전체 gradients를 계산하는 데에는 많은 연산량이 요구되고, 이는 학습시간을 크게 증가시킨다.

[0005] 본 발명에서는 기존의 방법과는 다르게 활성화 기록(Activation History)으로 중요한 가중치를 찾아 추가적인 연산이 요구되지 않는 기법을 제안한다. 또한, 본 발명에서는 과거의 지식으로 학습하는 예측적 역전파를 평생학습에 최적화하여 적용하였다. 요컨대, 본 발명에서는 치명적 망각을 효과적으로 완화시키는 것뿐만 아니라 예측적 역전파를 이용한 병렬학습으로 학습 속도를 가속화하는 장치 및 방법을 제안한다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 대한민국 공개특허 제2022-0030108호 (2022.03.10. 공개)

(특허문헌 0002) 대한민국 등록특허 제2097724호 (2020.04.06. 공고)

(특허문헌 0003) 대한민국 공개특허 제2019-0050698호 (2019.05.13. 공개)

비특허문헌

[0007] (비특허문헌 0001) Park, Sangwoo, and Taeweon Suh, "Speculative Backpropagation for CNN Parallel Training." IEEE Access 8 (2020)

발명의 내용

해결하려는 과제

[0008] 본 발명이 이루고자 하는 기술적인 과제는 딥러닝 모델이 다양한 도메인의 데이터를 학습하는 평생학습 과정에서 발생하는 치명적 망각 문제를 완화하고 학습 속도를 가속화시킬 수 있는 장치 및 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 본 발명의 일 실시예에 따른 딥러닝 모델의 평생학습 방법은 적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되고, 제1 태스크에 대하여 학습된 딥러닝 모델에 대한 제2 태스크에 대한 학습을 위하여, 순전파 동작을 수행하는 순전파 단계, 역전파 동작을 수행하는 역전파 단계, 및 가중치 업데이트 동작을 수행하는 가중치 업데이트 단계를 포함하고, 상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 반복적으로 수행되고, 상기 업데이트 단계는 상기 딥러닝 모델에 포함된 뉴런들 각각의, 상기 제1 태스크에 대한 학

습이 진행되는 과정에서의 활성화 경향에 기초하여 수행된다.

[0010] 또한, 본 발명의 일 실시예에 따른 딥러닝 모델의 학습 방법은 적어도 프로세서(processor)를 포함하는 컴퓨팅 장치에 의해 수행되고, 순전파 동작을 수행하는 순전파 단계, 역전파 동작을 수행하는 역전파 단계, 및 가중치 업데이트 동작을 수행하는 가중치 업데이트 단계를 포함하고, 상기 순전파 단계, 상기 역전파 단계, 및 상기 가중치 업데이트 단계는 반복적으로 수행되고, 최초 실행 이후에 반복적으로 수행되는 순전파 단계와 역전파 단계는 적어도 일 회 이상 병렬적으로 진행된다.

발명의 효과

[0011] 본 발명의 실시예에 따른, 예측적 역전파 및 활성화 기록에 기반한 딥러닝 모델 학습 장치 및 방법에 의할 경우, 딥러닝 모델이 다양한 도메인의 데이터를 학습하는 평생학습 과정에서 발생하는 치명적 망각 문제를 완화시킬 수 있다.

[0012] 또한, 예측적 역전파를 이용한 병렬학습을 통해 딥러닝 모델에 대한 평생학습의 학습 속도를 가속화할 수 있는 효과가 있다.

도면의 간단한 설명

[0013] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 예측적 역전파(Speculative Backpropagation, SB)가 수행되는 방식을 설명하기 위한 예를 도시한다

도 2는 태스크 1과 태스크 2에서 활성화된 뉴런들의 예시를 도시한다.

도 3은 본 발명에 따른, 예측적 역전파 및/또는 활성화 기록을 기반으로 하는 딥러닝 모델 학습 장치의 구성을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0014] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0015] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0016] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0017] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0018] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0019] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이

속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0020] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다. 또한, 아하에서 상세하게 설명되는 방법들에 포함되는 각 단계들 중 적어도 일부는 컴퓨팅 장치(또는 컴퓨팅 장치에 포함되는 프로세서)에 의해 수행되는 것으로 이해될 수 있다. 또한, 본 명세서에서 설명되는 방법 및/또는 장치에 관한 구체적인 설명은 본 발명의 발명자에 의해 공개된 논문(Sangwoo PARK, and Taeweon SUH, "Continual Learning With Speculative Backpropagation and Activation History.", IEEE Access, April 11, 2022.)이 참조될 수 있다.

[0022] **신경망 학습(The Neural Network Training)**

[0023] 신경망(Neural Network)을 학습하기 위해서는, 순전파(forward propagation), 역전파(backward propagation), 및 가중치 업데이트(weight update)의 세 과정이 순차적으로 진행되어야 하며, 이 세 과정은 소정의 목표가 달성될 때까지(예컨대, 타겟 에러율에 도달할 때까지) 반복적으로 수행될 수 있다.

[0024] 순전파에서, 데이터는 입력 레이어(input layer)로부터 출력 레이어(output layer)로 전파된다. 각 뉴런(neuron)은 이전 레이어(prior layer)의 연결된 뉴런들로부의 입력에 대한 가중합을 계산한 후, 수학식 1과 같이 계산된 값을 자신의 바이어스(bias)에 더한다. 뉴런의 출력은 데이터를 다음 레이어(next layer)로 보낼지 여부를 결정하는 활성화 함수(activation function)를 거치게 된다. 흔히 이용되는 활성화 함수는 *ReLU(Rectified Linear Unit)*, *Tanh*, 및 *Sigmoid* 등이 있다. 특히, 수학식 2로 정의되는 활성화 함수인 *ReLU*는 뛰어난 성능과 간단한 연산을 수반하기 때문에 많은 심층 신경망(Deep Neural Network, DNN)에서 이용되고 있다. *ReLU*는, 음의 입력(즉, 해당 뉴런이 비활성화된 경우)에 대하여 0을 다음 레이어로 전파하고, 그렇지 않은 경우(즉, 해당 뉴런이 활성화된 경우)에는 입력값을 다음 레이어로 바이패스한다(즉, 입력값을 그대로 다음 레이어로 전파한다). 따라서, 활성화된 뉴런만이 딥러닝 모델의 추론 결과에 영향을 줄 수 있다. 출력 레이어(output layer)에서는, 수학식 3의 소프트맥스 함수(*Softmax* function)가 널리 이용된다. 소프트맥스 함수는 결과(outcomes) (y_z^o)의 확률 분포(probability distribution)를 연산한다.

[0025] [수학식 1]

$$u_j^l = \sum_{i=1}^{N} w_{ij}^l x_i^l + Bias^l$$

[0026]

[0027] [수학식 2]

$$y_j^l = max(0, u_j^l)$$

[0028]

[0029] [수학식 3]

$$y_z^o = \frac{e^{u_z^o}}{\sum_{i=1}^{n} e^{u_i^o}}$$

[0030]

[0031] 위 수학식에서, $1 \le i \le N$, $1 \le l \le O$ (i, j, z, k = 뉴런 인덱스, l = 레이어 인덱스)이다.

[0032] 역전파는 도함수(derivatives)를 계산함으로써 가중치(weights) (w_{ij}^l)를 조정(adjust)하기 위해 이용된다. 역전파는, 예컨대 소프트맥스 등에 기반한 출력 레이어로부터 시작한다. 출력 레이어에서의 도함수는 수학식 4로 표현된다. 즉, 수학식 4는 순전파 결과 (y_z^o)와 타겟 출력(target output) (t_z) 사이의 차이(differeence)를 계산한다. 가중치와 관련된 에러(error)의 도함수는 수학식 5로 계산된다. *ReLU* 활성화 함수의 도함수는 수학식 6으로

계산되고, 수학식 2의 결과가 0이면 0의 값을 갖고 그렇지 않은 경우 1의 값을 갖는다.

[0033] [수학식 4]

[0034] $$\frac{dE}{dy_z^o} = y_z^o - t_z$$

[0035] [수학식 5]

[0036] $$\frac{dE}{dy_k^l} = \sum_z w_{kz}^{l+1} \frac{dE}{dy_z^{l+1}}$$

[0037] [수학식 6]

[0038] $$\delta_k^l = \frac{dE}{du_k^l} = \frac{dE}{dy_k^l}\frac{dy_k^l}{du_k^l}$$

[0039] 심층 신경망(DNN) 학습에서, 가중치들은 역전파에서 계산된 에러에 기초하여 조정된다. 우선, 수학식 7과 같이, 역전파의 결과 δ_i^l를 순전파의 결과 y_j^{l-1}에 곱함으로써 Δw_{ij}^l가 계산된다. 그런 다음, 가중치들 (w_{ij}^l)은 수학식 8에 따라 업데이트된다. 이때, 학습율(learning rate) η는 학습의 정도(the degree of learning)를 결정한다. 이러한 프로세스는 모든 가중치들에 대하여 반복된다.

[0040] [수학식 7]

[0041] $$\Delta w_{ij}^l = \delta_i^l y_j^{l-1}$$

[0042] [수학식 8]

[0043] $$w_{ij}^l = w_{ij}^l - \Delta w_{ij}^l \eta$$

[0045] **치명적 망각의 완화(Mitigating Catastrophic Forgetting)**

[0046] 이하에서는, 치명적 망각을 완화하기 위한 기법을 상세히 설명한다. 구체적으로 예측적 역전파와 활성화 기록이 이용될 수 있다. 각각의 방법은 두 개의 은닉 레이어(hidden layers)(각 레이어에는 400개의 뉴런이 포함됨)을 갖는 멀티-레이어 퍼셉트론(multi-layer perceptron) 상에서 구현될 수 있으나, 본 발명의 권리범위가 이에 제한되는 것이 아님은 자명하다. 또한, 상술될 두 가지의 방법은 단독으로 사용되거나 결합하여(동시에) 사용될 수 있다.

[0048] **A. 소프트맥스 기록과 예측적 역전파를 갖는 바이어스된 *ReLU*(Softmax History and Biased *ReLU* with Speculative Backpropagation(SB))**

[0049] 인공 신경망(Artificial Neural Network, ANN) 학습에서, 역전파는 순전파 결과에 기초하여 수행된다. 이는, 역전파는 순전파가 완료된 이후에만 수행될 수 있음을 의미한다. 그러나, 본 발명에서 제안하는 예측적 역전파(Speculative Backpropagation, SB)는 순전파와 역전파의 연속적인 동작(또는 병렬적인 동작)을 가능케 한다. 즉, 예측적 역전파(SB)는 과거의 지식을 기반으로 순전파와 역전파를 동시에 수행시켜 학습 속도를 가속화할 수 있는 기술이다. 예측적 역전파(SB)는, 동일한 레이블들(same labels)에 대한 소프트맥스와 *ReLU* 결과는 인접한 순전파들(near-forward propagations), 예컨대 적어도 하나의 이전 순전파에서의 결과와 유사하다는 관찰 결과에 기초한다.

[0050] 도 1은 SB가 수행되는 방식을 설명하기 위한 예를 도시한다. 즉, 도 1은 과거 순전파의 활성화 함수 결과(활성화 함수의 활성화 여부)와 추론 결과를 기반으로 현재의 순전파를 추정하여 순전파와 역전파를 병렬로 수행하는 예시이다. 현재 시간 t(i)에서, 역전파는 각각 빨간색 뉴런들과 파란색 뉴런들로 묘사된 소프트맥스 기록과 이전 *ReLU* 결과에 기초하여 수행된다. 이는, 역전파 계산에서 SGD(Stochastic Gradient Descent)의 내재적으로 연속적인 천성을 파괴하고, 순전파와 역전파의 연속적인 동작(또는 병렬적인 동작)을 가능케 한다. 예측적 역전

파(SB)는 임계값(threshold)에 기초하여 추론 정확성(speculation correctness)을 체크한다. 현재 소프트맥스 출력과 추론된 결과의 차이가 임계값보다 크다면, 추론된 동작은 무효화되고, 역전파는 현재 소프트맥스 출력을 이용하여 다시 수행된다. 다시 말하면, t(i)에서 순전파의 결과와 역전파를 수행하기 위해 추론된 t(i)에서의 순전파의 결과의 차이가 임계값보다 크면, 병렬적으로 수행된 역전파 동작은 무효화될 수 있고, 역전파는 t(i)에서의 순전파의 결과에 기초하여 재차 수행될 수도 있다. 이에 대한 실험에 따르면, 데이터셋(dataset) 내의 클래스의 개수의 역수(reciprocal)는 임계값에 대한 우수한 후보가 될 수 있다. 본 발명에서 임계값은 0.1로 설정하였으나, 본 발명이 임계값의 구체적인 수치에 제한되는 것은 아니며, 관리자에 따라 또는 상황에 따라 임계값은 상이할 수 있다. 또한, 예측적 역전파에 관한 보다 구체적인 내용은 S. Park 등의 논문(S. Park and T. Suh, "Speculative backpropagation for CNN parallel training," *IEEE Access*, vol. 8, pp. 215365-215374, 2020.)이 참조될 수 있다.

[0051] 연속적인 동작(순전파와 역전파의 병렬적인 수행)에 더하여, SB는 또한 평생학습(continual learning)에서 이전에 학습된 지식을 보존하는 데 도움이 된다. 앞서 기재한 S. Park 등의 논문(2020년)에서, 소프트맥스 기록은 수학식 9를 이용하여 업데이트된다. 수학식 9에서, α와 β는 각각 현재와 누적된 소프트맥스 결과들에 대한 가중치이다. α와 β를 0.5로 설정하는 경우 학습 시간을 단축시킴은 물론 경쟁적인 또는 보다 나은 정확도를 나타냈다. 평생학습을 위하여, α와 β의 값을 변경하여 실험을 진행하였다. 표 1은 실험 결과를 보여준다. 구체적으로, 표 1은 Permuted MNISTs(tasts 1 ~4)에 대한 연속적인 학습 후에 태스크 1의 정확도를 보여준다. 표 1의 결과로부터, β에 보다 큰 가중치를 부여하면 치명적 망각을 보다 더 완화할 수 있음을 알 수 있다. 이는, 기록에 보다 큰 가중치를 부여함으로써 과거의 지식은 보존되기 때문이다. 태스크 1의 정확도는 α가 0.2이고 β가 0.8일 때, 가장 높은 값을 보였다.

[0052] 다시 말하면, 예측적 역전파(SB)는 과거의 순전파 추론 결과를 누적하여 현재의 추론 결과를 추정하는데, 과거와 현재의 결과 중 어디에 가중치를 더 두어 누적하는지에 따라 성능이 달라진다. 예측적 역전파를 평생학습에 최적화하기 위해 과거의 결과에 더 큰 가중치를 부여할 수 있고, 이를 통해 지식을 보존할 수 있다. 실험결과, 수학식 9에서 과거 및 현재 가중치의 비를 8:2로 설정했을 때 치명적 망각 문제를 크게 완화시킬 수 있었다. 다만, 본 발명의 권리범위가 가중치의 구체적인 비율에 제한되는 것은 아님은 자명하다. 또한, 예측적 역전파(SB)는 이전 순전파의 *ReLU* 활성화 결과로 현재의 활성화 결과를 추정하는데, 본 발명에서는 알고리즘 1, 2와 같이 활성화 기록의 누적으로 추정하여 평생학습에 최적화시킬 수 있다. 누적된 활성화 기록으로 추정한 결과 과거의 지식을 효과적으로 보존할 수 있다.

[0053] [수학식 9]

$$y^{o'}_{t(i)} = (y^{o}_{t(i)}\alpha) + (y^{o'}_{t(i-1)}\beta)\ where\ (\alpha+\beta=1)$$

[0054] $$y^{o}_{t(i)} = \text{current Softmax outcome},$$

$$y^{o'}_{t(i-1)} = \text{accumulated Softmax outcome until time t(i-1)}$$

[0055] [표 1]

[0056]

Hyperparameters value	$\alpha=0.1$ $\beta=0.9$	$\alpha=0.2$ $\beta=0.8$	$\alpha=0.3$ $\beta=0.7$	$\alpha=0.4$ $\beta=0.6$	$\alpha=0.5$ $\beta=0.5$	$\alpha=0.6$ $\beta=0.4$	$\alpha=0.7$ $\beta=0.3$	$\alpha=0.8$ $\beta=0.2$	$\alpha=0.9$ $\beta=0.1$
MNIST handwritten SB accuracies	90.78%	91.15%	91.14%	88.86%	88.19%	86.92%	83.57%	82.3%	81.99%
MNIST Fashion SB accuracies	52.02%	52.22%	52.2%	51.71%	50.1%	48.21%	43.77%	41.52%	40.08%

[0057] SB는 가장 최근의 *ReLU* 결과를 이용하여 역전파를 수행한다. 평생학습의 경우, 가장 최근의 *ReLU* 결과와 *ReLU*의 기록(히스토리)을 모두 고려하는 것이 지식을 보존하는 데 도움이 된다. 알고리즘 1은 활성화 결과를 반영하는 방법에 관한 알고리즘으로써, 바이어스된(편향된) *ReLU*(*biased ReLU*)로 명명되는 $f(u^l_j)'$를 정의한다. 실시예에 따라, 바이어스된 *ReLU*는 조정된 *ReLU*, 누적된 *ReLU* 등으로 명명될 수도 있다. 바이어스된 *ReLU*는 순전파가 완료될 때마다 현재 활성화 결과에 의해 조정된다. 뉴런이 비활성화되었을 때, 바이어스된 *ReLU*는 0에 가까워진다. 뉴런이 활성화되었을 때, 바이어스된 *ReLU*는 1에 가까워진다.

[0058] [알고리즘 1]

Algorithm 1 Biased ReLU According to Activation History

1: $f(u_j^l)$: current ReLU output, $f(u_j^l)'$: biased ReLU
2: **if** $f(u_j^l) == 0$ do (when ReLU output is zero)
3: $f(u_j^l)' \leftarrow f(u_j^l)'/2$; (ReLU is biased towards 0)
4: **Else**
5: $f(u_j^l)' \leftarrow (f(u_j^l)' + 1)/2$; (ReLU is biased towards 1)
6: Where biased ReLU $f(u_j^l)' : (0 <= f(u_j^l)' <= 1)$

[0059]

[0060] 알고리즘 2는 바이어스된 *ReLU*에 기초하여 뉴런의 활성화를 예측(또는 유추)하는 방법에 관한 알고리즘이다. 바이어스된 *ReLU*가 0.5보다 작은 경우, 뉴런은 비활성화될 것으로 예측된다. 그렇지 않은 경우(즉, 0.5보다 크거나 같은 경우), 뉴런은 활성화될 것으로 예측된다. 알고리즘 1과 알고리즘 2에 기초하여 SB가 동작할 때, 이전 태스크의 정확도는 Permuted Handwritten 및 Fashion에 대한 평균에 비해 각각 2.3%와 1.9% 높음을 보았다.

[0061] [알고리즘 2]

Algorithm 2 ReLU Outcome Speculation

1: **if** $f(u_j^l)' < 0.5$ **do** (when biased ReLU is smaller than 0.5)
2: $\delta_k^l \leftarrow 0$; (speculate as the deactivation of the neuron)
3: **else**
4: $\delta_k^l \leftarrow 1$; (speculate as the activation of the neuron)

[0062]

[0063] 상술한 바와 같이, t(i-2) 시점의 소프트맥스 결과(딥러닝 모델의 출력)와 활성화 함수의 활성화 결과 및 t(i-1) 시점의 소프트맥스 결과와 활성화 함수의 활성화 결과를 이용하여 t(i) 시점의 순전파 결과를 예측할 수 있다. 구체적으로, 소프트맥스를 이용한 t(t) 시점의 순전파 결과 예측은, t(i-1) 시점의 소프트맥스의 결과와 t(i-2) 시점의 소프트맥스의 결과에 각각의 가중치를 곱한 후 합산함으로써 t(i) 시점의 소프트맥스의 결과를 예측할 수 있다. 이때, 곱해지는 가중치에 따라 치명적 망각의 문제를 완화하고 학습 속도를 감소시킬 수 있다. 또한, 바이어스된 ReLU 결과를 이용하여 t(i) 시점에서 예측된 순전파 결과에서의 각 뉴런의 활성화 여부를 예측할 수 있다. 즉, 각 뉴런의 활성화 여부는 이전의 순전파 결과(즉, 각 뉴런의 활성화 경향)을 반영한 결과로 이해될 수도 있다.

[0065] **B. 활성화 기록을 통한 바이어스된 가중치 업데이트(Biased Weight Update with Activation History(AH)**

[0066] 평생학습을 위하여, 다른 태스크를 위한 학습을 진행할 때 이전 태스크를 위한 중요한 가중치들이 고립되고 더 적은 영향을 받는다면 보유한 지식을 보존하는 데 도움이 될 것이다. 다른 태스크들을 위하여 모델을 학습할 때 활성화된 뉴런들의 두 가지 패턴이 발견되었다. (a) 특정 태스크를 학습하는 동안, 과거에 활성화된 뉴런들은 미래에도 활성화되는 경향이 있고, 과거에 비활성화된 뉴런들은 미래에도 비활성화되는 경향이 있다. 즉, 활성화된 뉴런들은 계속 활성화될 확률이 높고, 화성화되지 않은 뉴런들은 계속 활성화되지 않을 확률이 높다. (b) 활성화된 뉴런들은 태스크마다 상이하다. 즉, 각 태스크에 따라 활성화되는 뉴런들이 상이하다. 활성화된 뉴런들만이 추론 결과(inference result)에 영향을 미치고 *ReLU*로 인하여 태스크 내에서 중요한 역할을 수행한다. *ReLU*는 활성화된 경우 입력 값을 다음 레이어(nest layer)로 전달하고, 비활성화된 경우 0을 다음 레이어로 전파한다. Permuted MNIST Handwritten 및 Fashion에 대하여 멀티-레이어 퍼셉트론으로 실험했을 때, 동일한 뉴런들의 활성화 확률(activation probabilities)은 평균적으로 각각 87.8%와 84.2%이었다. 도 2는 태스크 1과 태스크 2에서 활성화된 뉴런들의 예시를 도시한다. 파란색 뉴런들과 초록색 뉴런들은 각각 태스크 1과 태스크 2에서 활성화된 뉴런들이다. 주황색 원(orange circle)으로 표시된 뉴런은 태스크 1과 태스크 2에서 모두 활성화되었다. 태스크 1에서 파란색 뉴런들의 활성화 확률은 Handwritten과 Fashion에서 평균적으로 각각 81%와 77.2%이다. 태스크 2에서 초록색 뉴런들의 활성화 확률은 Handwritten과 Fashion에서 평균적으로 각각 88.6%와 88.1%이다. 즉, 태스크 1에서 파란색 뉴런들은 계속 활성화되는 경향이 있고, 태스크 2에서 초록색 뉴런들은 계속 활성화되는 경향이 있다. 따라서, 파란색 뉴런들과 초록색 뉴런들과 연결된 가중치들은 각각 태스크 1과 태스크 2

에서 중요하다고 볼 수 있다.

[0067] 알고리즘 3은 활성화 기록(activation history)에 따라 가중치들을 업데이트하는 방법에 관한 알고리즘이고, 표 2는 알고리즘 3에서 사용된 용어들에 대한 설명을 나타낸다. 활성화 기록으로 명명되는 변수 (ah_j^l)는 이전 태스크들을 위한 학습 동안에 뉴런의 활성화 경향(tendency)을 저장한다. 따라서, 활성화 기록은 활성화 경향 등으로 명명될 수도 있고, 활성화 경향을 미리 정해진 범위(예컨대, 0 보다 크거나 같고 1보다 작거나 같은 실수의 범위)로 나타낼 수 있다. 예시적으로, 활성화 경향이 0.5인 경우는 해당 뉴런이 활성화된 경우와 비활성화된 경우의 수가 같은 경우이다. 마찬가지로, 활성화 경향이 0.5 보다 작은 경우는 해당 뉴런이 비활성화된 경우가 활성화된 경우의 수보다 큰 경우이고, 활성화 경향이 0.5 보다 큰 경우는 해당 뉴런이 활성화된 경우가 비활성화된 경우의 수보다 큰 경우를 의미할 수 있다. 또한, 알고리즘 3의 line# 3에서의 0.5는 가변될 수 있다. 예컨대, 활성화 기록은 0.5 보다 작은 값이나 0.5 보다 큰 값과 비교될 수도 있다. 순전파와 역전파에 후행하는 가중치 업데이트는 딥러닝 모델에 포함된 가중치들 각각이 미리 정해진 범위내로 수렴할 때까지(또는 목표 에러율에 도달할 때까지) 반복적으로 수행될 수 있다.

[0068] 각 태스크를 위한 학습 이후에(즉, 하나의 태스크에 대한 학습이 완료된 후에), ah_j^l는 알고리즘 1에서 계산된 바이어스된 ReLU로 조정된다(line# 7). r은 가중치 업데이트의 정도를 나타내고 직접적으로 지식 보존 성능에 영향을 미친다. r에 비례하여, 새로운 태스크를 위한 학습시에, 과거에 활성화된 뉴런에 연결된 가중치들은 덜 업데이트된다(line# 3-4). 다시 말해, 새로운 태스크를 학습할 때, 과거 태스크에서 중요한 역할을 하는 가중치들에 대한 업데이트를 제한하거나 거의 업데이트하지 않게 만들어 이전 태스크에서 학습한 지식을 보존할 수 있다. 표 3은 순차적인 Permuted MNISTs 학습(태스크 1 ~4) 이후에 태스크 1의 r에 따른 정확도를 리포트하는 실험 결과를 보여준다. 개별적으로 선택된 r에 대하여, 정확도는 노멀 학습(baseline)에 비해 높고, 이는 보유 지식이 더 많이 보존되었음을 의미한다. 일반적으로, r이 작을수록 더 좋은 결과가 도출되는 경향이 있다. 정확도는, MNIST Handwritten의 경우 r이 0.3일때 가장 높고, Fashion의 경우 r이 0.4일 때 가장 높다. 그러나, r의 값이 너무 작아지면 새로운 태스크에 대한 정확도는 열화됨을 알 수 있다. 예컨대, r이 0.1일 때, 태스크 4의 정확도는 평균적으로 98.3%이지만, 이는 기준치(99%)에 비해 0.7% 낮은 수치이다. 이는, 몇몇 뉴런들(예컨대, 도 2에서 주황색 원으로 표시된 뉴런)이 이전 태스크와 현재 태스크에서 모두 활성화된기 때문이다. 이러한 경우에, 작은 r은, 이전에 활성화된 뉴런들과 연결된 가중치들들이 새로운 태스크에 대하여 최소한으로 학습되었음을 의미한다. 실험에 따르면, 47.2%의 뉴런들이 태스크 1과 태스크 2에서 공통적으로 활성화되었고, 49.8%의 뉴런들이 태스크 2와 태스크 3에서 공통적으로 활성화되었다. 이러한 뉴런들에 연결된 가중치들은 두가지의 태스크 모두의 추론에 영향을 미친다. 따라서, 지식 보유(knowledge retention)와 새로운 학습(new learning)을 밸런싱하기 위하여, 적절한 r을 찾는 것이 중요하다. 경험적으로, r=0.3이 지식을 보존하고 새로운 태스크를 학습하기 위한 올바른 선택일 수 있다. r이 0.3일 때, 이전 태스크들의 정확도는, 새로운 태스크에 대한 정확도 손실 없이 기준치(baseline)에 비해 평균적으로 대략 15.1% 높았다.

[0069] [알고리즘 3]

Algorithm 3 Biased Weight Update According to Activation History

1: **while** weight not converged **do**
2: $g \leftarrow \nabla_w f(w)$ (get gradients with objective function)
3: **if** $ah_j^l > 0.5$ **do** (if the neuron is activated with high probability for the previous task)
4: $g \leftarrow g \cdot r$ (reduce the gradients of the weights)
5: $w \leftarrow w - \eta \cdot g$ (update the weight)
6: **end while**
7: $ah_j^l \leftarrow (ah_j^l + f(u_j^l)')/2$; (accumulate biased ReLU to activation history after training each task)
8: **return** w

[0070]

[0071] [표 2]

Notation	Description
w	weight of neural network
$f(w)$	gradient descent objective function with weight
g	gradient
ah_j^l	ReLU activation history of the previous tasks
r	rate of biased weight update ($0 < r < 1$)
η	learning rate
$f(u_j^l)'$	biased ReLU

[0072]

[0073] [표 3]

weight update degree (r)	$r = 0.1$	$r = 0.2$	$r = 0.3$	$r = 0.4$	$r = 0.5$	$r = 0.6$	$r = 0.7$	$r = 0.8$	$r = 0.9$
MNIST Handwritten SB accuracies	91.10%	91.11%	92%	91.8%	91.83%	89.97%	85.12%	83.1%	82.69%
MNIST Fashion SB accuracies	60.70%	60.72%	61.35%	61.61%	61.42%	57.01%	52.83%	48.46%	45.19%

[0074]

[0076] **C. 예측적 역전파(SB)와 활성화 기록(AH)을 모두 사용한 치명적 망각의 완화(Mitigating Catastrophic Forgetting Using Both SB and AH)**

[0077] 예측적 역전파(SB)와 활성화 기록(AH)는 직교(orthogonal)하기 때문에, 두 가지 방법을 모두 평생학습에 함께 적용할 수 있다. 예측적 역전파(SB)는 순전파 과정에서 수학식 1 내지 3에 적용되고, 역전파 과정에서 수학식 4 내지 6에 적용된다. 활성화 기록(AH)은 알고리즘 3에 의한 가중치 업데이트 과정에서 수학식 8에 적용된다. 예측적 역전파(SB)와 활성화 기록(AH)을 모두 이용함으로써 평생학습의 성능을 보다 더 높일 수 있다.

[0079] 이상에서는 본 발명의 일 양상에 따른 예측적 역전파 및/또는 활성화 기록을 기반으로 하는 딥러닝 모델의 학습 방법에 대해 설명하였다. 이하에서는 본 발명의 다른 양상에 따른, 예측적 역전파 및/또는 활성화 기록을 기반으로 하는 딥러닝 모델 학습 장치에 대해 설명한다. 이와 관련하여, 딥러닝 모델의 학습 방법에 대한 모든 기술적 특징 및 각 과정의 동작과 이하 설명할 딥러닝 모델 학습 장치의 기술적 특징, 동작 및 구성이 상호 결합될 수 있다. 또한, 본 명세서에 기재된 딥러닝 모델은 ANN(Atrificial Neural Network), DNN(Deep Neural Network), CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), RBM(Restricted Boltzmann Machine), DBN(Deep Belief Network), GAN(Generativ Adversarial Network), DFN(Deep Feedforward Network), LSTM(Long Short-Term Memory), Autoencoder, VAE(Variational Autoencoder), DRN(Deep Residual Network), GCN(Graph Convolutional Network), SNN(Spiking Neural Network) 등과 같은 신경망 모델을 포함할 수 있다.

[0081] 이와 관련하여, 도 3은 본 발명에 따른, 예측적 역전파 및/또는 활성화 기록을 기반으로 하는 딥러닝 모델 학습 장치의 구성을 도시한다. 딥러닝 모델 학습 장치를 설명함에 있어, 앞선 기재와 중복되는 내용에 관하여는 그 구체적인 설명을 생략하기로 한다.

[0082] 도 3을 참조하면, 딥러닝 모델 평생학습 장치 등으로 명명될 수도 있는 딥러닝 모델 학습 장치(1000)는 인터페이스(100), 디스플레이(200), 프로세서(300), 및 메모리(400)를 포함하도록 구성될 수 있다. 인터페이스(100)는 복수의 태스크들 각각에 대응되는 학습 데이터 및/또는 테스트 데이터가 입력되도록 구성될 수 있다. 디스플레이(200)는 딥러닝 모델의 학습 과정에서 발생하는 데이터나 결과물들을 화면으로 구성하여 화면 상에 표시하도록 구성될 수 있다. 다만, 상술한 인터페이스(100)나 디스플레이(200)는 딥러닝 모델 학습 장치(100)의 필수 구성 요소로 고려되지는 않는다. 예컨대, 소정의 데이터가 딥러닝 모델 학습 장치(100)에 포함될 수 있는 저장 장치나 메모리(400)에 미리 저장되어 있거나 모든 동작이 자동으로 수행되어 별도의 디스플레이 동작이 필요 없는 경우에는 인터페이스(100) 및/또는 디스플레이(200)는 생략될 수도 있다.

[0083] 프로세서(300)는 인터페이스(100), 디스플레이(200), 및 메모리(400)와 동작 가능하게 결합될 수 있다. 프로세서(300)는 딥러닝 모델에 대한 학습을 수행할 수 있는 (어플리케이션) 프로그램이 실행되도록 구성될 수 있다. 프로세서(300)는 복수의 태스크들 각각에 대하여 수신된 데이터 및/또는 미리 저장된 데이터를 이용하여 딥러닝 모델에 대한 학습을 진행하고, 학습 결과물 등을 디스플레이(200) 상에 가시화되도록 제어할 수 있다. 결국, 앞서 설명한 딥러닝 모델의 학습 방법을 포함하는 단계들 중 적어도 일부는 딥러닝 모델 학습 장치(100)나 프로세서(300)의 동작으로 이해될 수 있다.

[0084] 프로세서(300)는 이전 태스크에 대한 학습의 결과로 인한 뉴런들의 활성화 기록에 기초하여 다음 태스크에 대한 딥러닝 모델의 학습을 수행하도록 구성될 수 있다. 다시 말하면, 프로세서(300)는 각각이 복수의 도메인들 각각에 대응하는 태스크들에 대한 학습(예컨대, 평생학습)을 실행하도록 구성될 수 있다. 구체적으로, 각 태스크에 대한 순전파 단계, 역전파 단계, 및 가중치 업데이트 단계를 수행할 수 있다. 순전파 단계, 역전파 단계, 및 가중치 업데이트 단계는 소정 조건이 만족될 때까지 반복적으로 수행될 수 있다. 또한, 프로세서(300)는, 딥러닝 모델에 포함된 뉴런들 각각의, 이전 태스크에 대한 학습이 진행되는 과정에서의 활성화 경향에 기초하여 업데이트 단계를 수행할 수 있다.

[0085] 또한, 프로세서(300)는 반복적으로 수행되는 순전파, 역전파, 및 가중치 업데이트 과정 중에서, 이전에 수행된 적어도 하나의 순전파 결과를 이용하여 현재 시점의 순전파 결과를 예측하고, 예측된 결과에 기초하여 현재 시점에서의 순전파와 역전파를 병렬적으로(또는 동시에) 수행할 수 있다.

[0087] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 콘트롤러, 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(Operation System, 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 기술 분야에서 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.

[0088] 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, 구성 요소(Component), 물리적 장치, 가상 장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 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.

[0089] 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), 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0090] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0091] 100 : 인터페이스

200 : 디스플레이

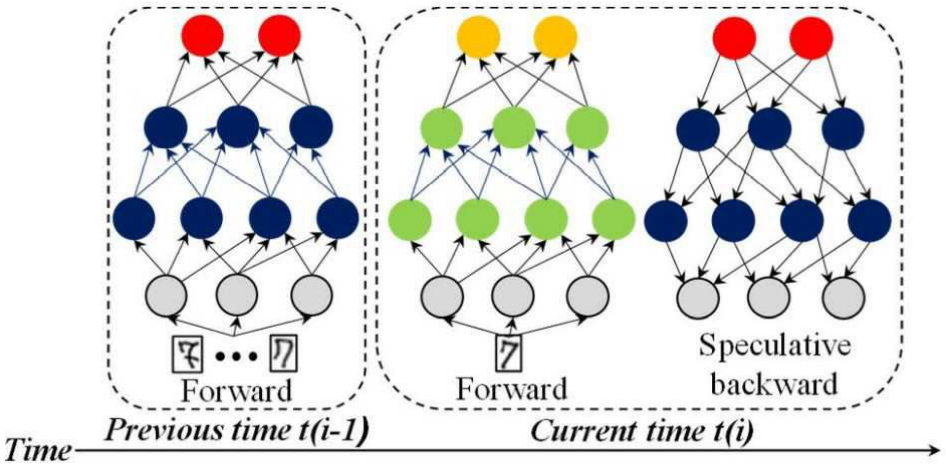
300 : 프로세서

400 : 메모리

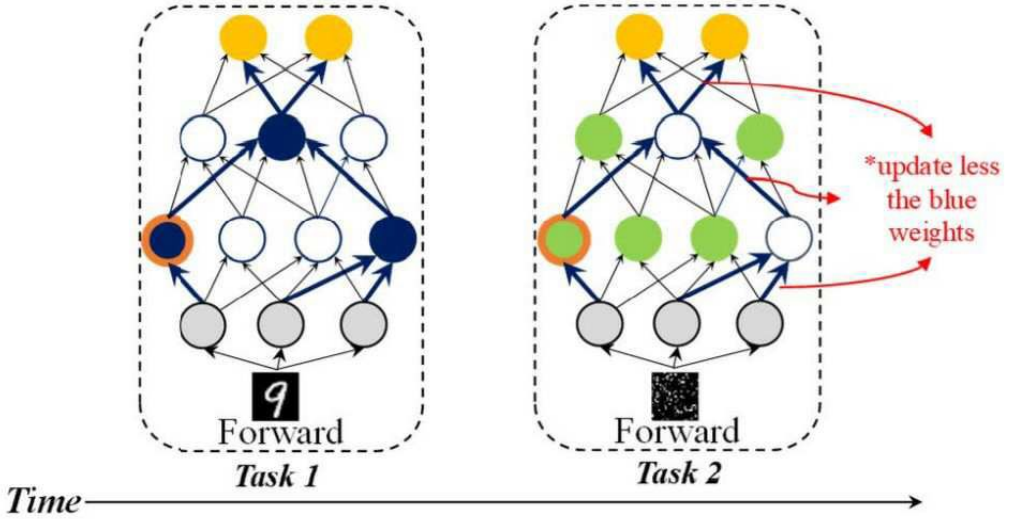
1000 : 딥러닝 모델 학습 장치

도면

도면1



도면2



도면3

