



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0093648
(43) 공개일자 2021년07월28일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/082 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2020-0007509
(22) 출원일자 2020년01월20일
심사청구일자 2020년01월20일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
배성호
경기도 용인시 기흥구 예현로 15 서그내마을에스케이아파트 102동 501호
박진배
경기도 남양주시 도농로 34 부영그린타운 104동 1301호
(74) 대리인
특허법인지원

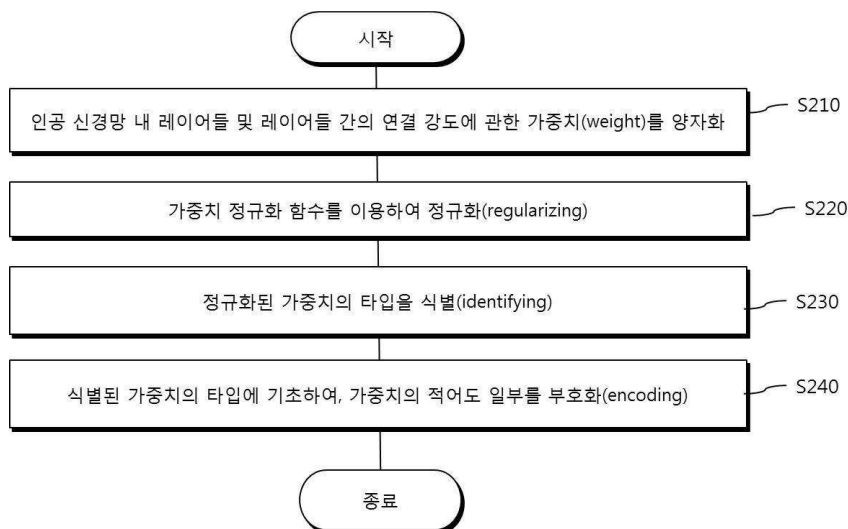
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **인공 신경망의 가중치를 처리하는 방법 및 장치**

(57) 요약

본 개시는 인공 신경망의 가중치를 처리하는 방법 및 전자 장치에 관한 것이다. 인공 신경망의 가중치를 처리하는 방법은 상기 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하는 단계; 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하는 단계; 상기 정규화된 가중치의 타입을 식별하는 단계; 및 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 단계; 를 포함할 수 있다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호	10085646
부처명	과학기술정보통신부
과제관리(전문)기관명	한국산업기술평가관리원
연구사업명	전자정보디바이스산업원천기술개발사업(반도체)
연구과제명	3D 메모리스터 어레이 결점 극복을 위한 뉴로모픽 시스템
기 여 율	1/1
과제수행기관명	경희대학교산학협력단
연구기간	2020.01.01 ~ 2020.06.30

명세서

청구범위

청구항 1

인공 신경망(Neural Network)의 가중치를 처리하는 방법에 있어서,
 상기 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하는 단계;
 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하는 단계;
 상기 정규화된 가중치의 타입을 식별하는 단계; 및
 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 단계; 를 포함하는, 방법.

청구항 2

제1항에 있어서, 상기 방법은
 상기 인공 신경망 내 가중치들이 양자화 되기 전, 사전 학습된(pre-trained) 가중치에 기초하여 상기 인공 신경망 내 가중치들을 초기화 하는 단계; 를 더 포함하고,
 상기 양자화 하는 단계는 상기 초기화된 인공 신경망 내 가중치들을 양자화하는 것인, 방법.

청구항 3

제2항에 있어서, 상기 양자화 하는 단계는
 상기 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 제1 양자화 함수를 결정하는 단계; 및
 상기 결정된 제1 양자화 함수에 따라 양자화된 가중치들을 포함하는 상기 인공 신경망의 제1 손실 함수를 최소화하도록 상기 가중치를 양자화 하는 단계; 를 포함하는, 방법.

청구항 4

제3항에 있어서, 상기 정규화 하는 단계는
 부분 L1 제약 조건 및 L2 제약 조건에 기초하여 제1 가중치 정규화 함수를 결정하는 단계; 및
 상기 결정된 제1 가중치 정규화 함수에 따라 정규화된 가중치들을 포함하는 상기 인공 신경망의 제2 손실 함수를 최소화하도록 상기 가중치를 정규화 하는 단계; 를 포함하는, 방법.

청구항 5

제1항에 있어서, 상기 식별하는 단계는
 상기 가중치의 절대 값 및 기 설정된 임계치를 비교하는 단계; 및
 상기 비교 결과에 기초하여, 상기 정규화된 가중치를 -1, 0 또는 1 중 하나를 나타내는 제1 타입의 가중치 또는 상기 가중치 중, 상기 제1 타입의 가중치가 아닌 가중치를 제2 타입의 가중치로 식별하는 단계; 를 포함하는, 방법.

청구항 6

제5항에 있어서, 상기 부호화 하는 단계는
 상기 제1 타입의 가중치를, 2비트로 표현 가능한 양자화 상태에 각각 대응되는 이진 값으로 변환하는 단계; 및
 상기 제2 타입의 가중치를, 상기 이진 값으로 변환된 제1 타입의 가중치에 대응되지 않는 양자화 상태를 이용하여 상기 제2 타입의 가중치를 부호화 하는 단계; 를 포함하는, 방법.

청구항 7

제4항에 있어서, 상기 방법은

상기 식별된 가중치의 타입에 기초하여 상기 제1 양자화 함수와 다른 제2 양자화 함수를 결정하는 단계;

상기 식별된 가중치의 타입에 따라 상기 제1 양자화 함수 및 상기 제2 양자화 함수를 선택적으로 적용함으로써, 상기 가중치를 재 양자화 하는 단계; 및

상기 재 양자화된 가중치를 상기 L2 제약 조건에 대응되는 제2 가중치 정규화 함수를 이용하여 재 정규화 하는 단계; 를 더 포함하는, 방법.

청구항 8

제7항에 있어서, 상기 재 양자화 하는 단계는

상기 제1 양자화 함수 또는 상기 제2 양자화 함수에 따라 양자화된 가중치를 포함하는 상기 인공 신경망의 제3 손실 함수를 최소화하도록 상기 가중치를 재 양자화 하는 단계; 를 포함하는, 방법.

청구항 9

제7항에 있어서, 상기 재 정규화 하는 단계는

상기 결정된 제2 가중치 정규화 함수에 따라 재 정규화된 가중치들을 포함하는 상기 인공 신경망의 제4 손실 함수를 최소화하도록 상기 재 양자화된 가중치를 재 정규화하는 단계; 를 포함하는, 방법.

청구항 10

제7항에 있어서, 상기 방법은

상기 재 정규화된 가중치의 타입을 식별하는 단계; 를 더 포함하고,

상기 부호화 하는 단계는 상기 재 정규화된 가중치의 타입에 기초하여 상기 재 정규화된 가중치의 적어도 일부를 부호화 하는 것을 특징으로 하는, 방법.

청구항 11

제1항에 있어서, 상기 방법은

상기 부호화된 가중치 및 상기 가중치 중, 부호화되지 않은 가중치를 저장하는 단계; 를 더 포함하는, 방법.

청구항 12

제8항에 있어서, 상기 제1 양자화 함수는

상기 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 상기 초기화된 가중치들을 클리핑하는 클리핑 함수의 출력 값 및 상기 가중치의 비트 수를 입력으로 하는 라운드 함수인 것을 특징으로 하는 방법.

청구항 13

인공 신경망(Neural Network)의 가중치를 처리하는 전자 장치에 있어서,

하나 이상의 인스트럭션을 저장하는 메모리; 및

상기 하나 이상의 인스트럭션들을 실행하는 적어도 하나의 프로세서; 를 포함하고,

상기 적어도 하나의 프로세서는 상기 하나 이상의 인스트럭션을 실행함으로써,

상기 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하고,

상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하고,

상기 정규화된 가중치의 타입을 식별하고,

상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 전자 장치.

청구항 14

제13항에 있어서, 상기 적어도 하나의 프로세서는

상기 인공 신경망 내 가중치들이 양자화 되기 전, 사전 학습된(pre-trained) 가중치에 기초하여 상기 인공 신경망 내 가중치들을 초기화 하고,

상기 초기화된 인공 신경망 내 가중치들을 양자화하는 것을 특징으로 하는, 전자 장치.

청구항 15

제14항에 있어서, 상기 적어도 하나의 프로세서는

상기 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 제1 양자화 함수를 결정하고,

상기 결정된 제1 양자화 함수에 따라 양자화된 가중치들을 포함하는 상기 인공 신경망의 제1 손실 함수를 최소화하도록 상기 가중치를 양자화 하는, 전자 장치.

청구항 16

제15항에 있어서, 상기 적어도 하나의 프로세서는

부분 L1 제약 조건 및 L2 제약 조건에 기초하여 제1 가중치 정규화 함수를 결정하고,

상기 결정된 제1 가중치 정규화 함수에 따라 정규화된 가중치들을 포함하는 상기 인공 신경망의 제2 손실 함수를 최소화하도록 상기 가중치를 정규화 하는, 전자 장치.

청구항 17

제13항에 있어서, 상기 적어도 하나의 프로세서는

상기 가중치의 절대 값 및 기 설정된 임계치를 비교하고,

상기 비교 결과에 기초하여, 상기 정규화된 가중치를 -1, 0 또는 1 중 하나를 나타내는 제1 타입의 가중치 또는 상기 가중치 중, 상기 제1 타입의 가중치가 아닌 가중치를 제2 타입의 가중치로 식별하는, 전자 장치.

청구항 18

제17항에 있어서, 상기 적어도 하나의 프로세서는

상기 제1 타입의 가중치를, 2비트로 표현 가능한 양자화 상태에 각각 대응되는 이진 값으로 변환하고,

상기 제2 타입의 가중치를, 상기 이진 값으로 변환된 제1 타입의 가중치에 대응되지 않는 양자화 상태를 이용하여 상기 제2 타입의 가중치를 부호화 하는, 전자 장치.

청구항 19

제16항에 있어서, 상기 적어도 하나의 프로세서는

상기 식별된 가중치의 타입에 기초하여 상기 제1 양자화 함수와 다른 제2 양자화 함수를 결정하고,

상기 식별된 가중치의 타입에 따라 상기 제1 양자화 함수 및 상기 제2 양자화 함수를 선택적으로 적용함으로써, 상기 가중치를 재 양자화 하고,

상기 재 양자화된 가중치를 상기 L2 제약 조건에 대응되는 제2 가중치 정규화 함수를 이용하여 재 정규화 하는, 전자 장치.

청구항 20

인공 신경망(Neural Network) 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하는 단계;

상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하는 단계;

상기 정규화된 가중치의 타입을 식별하는 단계; 및

상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 단계; 를 포함하는, 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

발명의 설명

기술 분야

[0001] 본 개시는 인공 신경망의 가중치를 처리하는 방법 및 장치에 관한 것이다. 보다 상세하게는, 인공 신경망의 가중치를 양자화함으로써 가중치를 표현할 수 있는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 인공 신경망(Artificial Neural Network)은 인공 뉴런들의 상호 연결된 집합들을 구현하기 위하여 컴퓨팅 기기 또는 컴퓨팅 기기에 의해서 수행되는 방법을 지칭할 수 있다. 인공 신경망의 일 실시 예로, 심층 신경망(Deep Neural Network) 또는 딥 러닝(Deep Learning)은 멀티 레이어 구조를 가질 수 있고, 레이어들 각각이 다수의 데이터에 따라 학습될 수 있다.

[0003] 최근 인공 신경망 기술의 개발이 활성화 됨에 따라, 인공 지능 분야에서 연산량을 줄이면서도 동시에 전력 효율성을 향상시키기 위한 방법으로 양자화(quantization) 기술이 활발히 연구되고 있다. 인공 신경망의 양자화 기술은 고성능 및 저전력 소모라는 장점을 제공할 수 있지만, 양자화 기술에 따른 저장밀도 기반의 저비트 연산은 인공 지능 신경망의 정확도(accuracy)가 저하되는 문제점이 있다.

[0004] 따라서, 인공 신경망에서 양자화 기술을 통하여 저비트 연산을 수행함으로써 효과적으로 인공 신경망을 압축함과 함께 인공 신경망의 정확도를 향상시키기 위한 기술개발이 요구되고 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국공개특허 제2019-0034985호

발명의 내용

해결하려는 과제

[0006] 일 실시 예에 따르면, 인공 신경망의 가중치를 처리하는 방법 및 전자 장치가 제공될 수 있다.

[0007] 또한, 일 실시 예에 의하면, 양자화된 인공 신경망의 가중치의 적어도 일부를 부호화하는 방법 및 전자 장치가 제공될 수 있다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 달성하기 위한 본 개시의 일 실시 예에 따라, 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하는 단계; 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하는 단계; 상기 정규화된 가중치의 타입을 식별하는 단계; 및 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 단계; 를 포함하는 인공 신경망(Neural Network)의 가중치를 처리하는 방법이 제공될 수 있다.

[0009] 또한, 상기 기술적 과제를 해결하기 위한 본 개시의 또 다른 실시 예에 따라, 하나 이상의 인스트럭션을 저장하는 메모리; 및 상기 하나 이상의 인스트럭션들을 실행하는 적어도 하나의 프로세서; 를 포함하고, 상기 적어도 하나의 프로세서는 상기 하나 이상의 인스트럭션을 실행함으로써, 상기 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하고, 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하고, 상기 정규화된 가중치의 타입을 식별하고, 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는, 인공 신경망(Neural Network)의 가중

치를 처리하는 전자 장치가 제공될 수 있다.

[0010] 또한, 상기 기술적 과제를 해결하기 위한 본 개시의 또 다른 실시 예에 따라, 인공 신경망(Neural Network) 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하는 단계; 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하는 단계; 상기 정규화된 가중치의 타입을 식별하는 단계; 및 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화하는 단계;를 포함하는, 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체가 제공될 수 있다.

도면의 간단한 설명

[0011] 도 1은 일 실시 예에 따른 전자 장치가 인공 신경망의 가중치를 처리하는 과정을 나타내는 도면이다.
 도 2는 일 실시 예에 따른 인공 신경망의 가중치를 처리하는 방법의 흐름도이다.
 도 3은 일 실시 예에 따른 인공 신경망의 가중치를 양자화하는 구체적인 방법을 설명하기 위한 도면이다.
 도 4는 일 실시 예에 따른 인공 신경망의 가중치를 정규화하는 구체적인 방법을 설명하기 위한 도면이다.
 도 5는 일 실시 예에 따른 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 나타내는 도면이다.
 도 6은 일 실시 예에 따른 가중치의 타입을 식별하는 방법을 구체적으로 설명하기 위한 도면이다.
 도 7은 일 실시 예에 따른 가중치의 적어도 일부를 부호화하는 방법을 구체적으로 설명하기 위한 도면이다.
 도 8은 일 실시 예에 따른 전자 장치가 가중치의 적어도 일부를 부호화하는 방법을 설명하기 위한 도면이다.
 도 9는 또 다른 실시 예에 따른 전자 장치가 인공 신경망의 가중치를 처리하는 방법의 흐름도이다.
 도 10은 일 실시 예에 따른 전자 장치가 식별된 가중치의 타입에 따라 서로 다르게 양자화된 가중치들을 설명하기 위한 도면이다.
 도 11은 일 실시 예에 따른 인공 신경망의 가중치를 처리하는 전자 장치의 블록도이다.
 도 12는 일 실시 예에 따른 전자 장치와 통신 가능한 서버의 블록도이다.
 도 13은 일 실시 예에 따라 전자 장치가 이용하는 인공 신경망의 성능을 설명하기 위한 도면이다.
 도 14는 또 다른 실시 예에 따라 전자 장치가 이용하는 인공 신경망의 성능을 설명하기 위한 도면이다.
 도 15는 일 실시 예에 따른 전자 장치 내 인공 신경망의 동작을 설명하기 위한 도면이다.
 도 16은 일 실시 예에 따른 전자 장치 내 인공 신경망의 동작을 설명하기 위한 도면이다.
 도 17은 또 다른 실시 예에 따른 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 개시에 대해 구체적으로 설명하기로 한다.

[0013] 본 개시에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0014] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0015] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 개시를 명확하게 설명하기 위해서 설

명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

- [0016] 도 1은 일 실시 예에 따른 전자 장치가 인공 신경망의 가중치를 처리하는 과정을 나타내는 도면이다.
- [0017] 일 실시 예에 의하면, 전자 장치(1000)는 인공 신경망(Artificial Neural Network)(102)을 포함할 수 있고, 인공 신경망의 가중치를 처리할 수 있다. 예를 들어, 전자 장치(1000)는 인공 신경망의 가중치를 양자화(quantizing)하고, 양자화된 가중치의 적어도 일부를 부호화(encoding)할 수 있다.
- [0018] 일 실시 예에 의하면, 전자 장치(1000)는 인공 신경망의 가중치를 처리하기 위한, AI 프로그램이 탑재되고 음성 인식 기능을 포함하는 스마트폰, 태블릿 PC, PC, 스마트 TV, 휴대폰, 미디어 플레이어, 서버, 마이크로 서버, 기타 모바일 또는 비모바일 컴퓨팅 장치일 수 있으나, 이에 제한되지 않는다.
- [0019] 일 실시 예에 의하면, 전자 장치(1000)가 이용하는 인공 신경망(Artificial Neural Network)은 생물학적 신경망에 착안된 컴퓨팅 시스템을 지칭할 수 있다. 인공 신경망은 미리 정의된 조건에 따라 작업을 수행하는 고전적인 알고리즘과 달리, 다수의 샘플들을 고려함으로써 작업을 수행하는 것을 학습할 수 있다. 인공 신경망은 인공 뉴런(neuron)들이 연결된 구조를 가질 수 있고, 뉴런들 간의 연결은 시냅스(synapse)로 지칭될 수 있다. 뉴런은 수신된 신호를 처리할 수 있고, 처리된 신호를 시냅스를 통해서 다른 뉴런에 전송할 수 있다. 뉴런의 출력은 액티베이션(activation)으로 지칭될 수 있고, 뉴런 및/또는 시냅스는 변동될 수 있는 가중치(weight)를 가질 수 있고, 가중치에 따라 뉴런에 의해 처리된 신호의 영향력이 증가하거나 감소할 수 있다.
- [0020] 예를 들어, 인공 신경망은 복수의 신경망 레이어들로 구성될 수 있다. 복수의 신경망 레이어들 각각은 복수의 가중치들(weight values, weights, 104)을 갖고 있으며, 이전(previous) 레이어의 연산 결과와 복수의 가중치들 간의 연산을 통해 신경망 연산을 수행한다. 복수의 신경망 레이어들이 갖고 있는 복수의 가중치들은 인공 신경망의 학습 결과에 의해 최적화될 수 있다.
- [0021] 예를 들어, 학습 과정 동안 인공지능 모델에서 획득한 손실(loss) 값 또는 코스트(cost) 값이 감소 또는 최소화 되도록 복수의 가중치들이 수정 및 갱신될 수 있다. 본 개시에 따른 인공 신경망은 심층 신경망(DNN:Deep Neural Network)를 포함할 수 있으며, 예를 들어, CNN (Convolutional Neural Network), DNN (Deep Neural Network), RNN (Recurrent Neural Network), RBM (Restricted Boltzmann Machine), DBN (Deep Belief Network), BRDNN(Bidirectional Recurrent Deep Neural Network) 또는 심층 Q-네트워크 (Deep Q-Networks) 등이 있으나, 전술한 예에 한정되지 않는다. 이하에서는 편의상 본 개시에 따른 인공 신경망은 심층 신경망인 경우를 예로 설명하기로 한다.
- [0022] 일 실시 예에 의하면, 전자 장치(1000)는 인공 신경망의 가중치를 양자화(S112)하고, 양자화된 가중치의 분포를 정규화(S114)할 수 있다.
- [0023] 일 실시 예에 의하면, 본 개시에 따른 전자 장치(1000)가 수행하는 양자화는 예컨대 반올림을 통해서 실수(real number)를 정수(integer)로 맵핑하는 것과 같이, 입력 값들을 입력 값들의 개수 보다 작은 개수의 값들로 맵핑하는 과정을 지칭할 수 있다. 또 다른 실시 예에 의하면, 전자 장치(1000)가 수행하는 양자화는 부동 소수점(floating point) 연산 기반의 신경망을 고정 소수점(fixed point) 연산 기반의 신경망으로 변환하는 과정을 의미할 수 있다. 본 개시에 따른 양자화는 입력값을 단순화된 단위의 값으로 변환하기 위한 기타 알고리즘을 포함할 수 있다.
- [0024] 예를 들어, 전자 장치(1000)는 인공 신경망의 가중치를 양자화(S112)고, 양자화 과정에서 인공 신경망의 가중치를 정규화(S114)함으로써, 가중치의 크기가 큰 일부 가중치들을 가중치의 크기가 작은 가중치로 정규화함으로써 인공 신경망의 가중치 계산이 필요한 높은 계산 복잡도(Computational complexity)를 줄일 수 있다. 예를 들어, 도 1의 S114를 참조하면, 전자 장치(1000)가 가중치들을 양자화 하는 과정에서 양자화된 가중치들을 정규화함으로써, 가중치들에 대하여 중앙 집중화된 양자화(Centralized Quantization)를 수행할 수 있다.
- [0025] 또한, 전자 장치(1000)는 S114에서 양자화된 가중치들을 미리 설정된 타입으로 분류할 수 있다. 예를 들어, 전자 장치(1000)는 양자화된 가중치들의 절대 값 크기를 미리 설정된 임계치와 비교함으로써, 가중치들의 타입을 제1 타입의 가중치(122) 또는 제2 타입의 가중치(124)로 식별할 수 있다. 전자 장치(1000)는 식별된 가중치의 타입에 기초하여, 양자화된 가중치들의 적어도 일부를 부호화(encoding)할 수 있다.
- [0026] 본 개시에 따른 전자 장치(1000)는 양자화 과정에서 가중치의 정규화를 통하여 분포된 가중치의 타입을 식별하고, 식별된 가중치의 타입에 따라 가중치를 값(value) 및 지시자(indicies) 모두를 이용하여 나타낼 수 있기 때

문에, 효과적으로 인공 신경망을 압축함과 함께 인공 신경망의 정확도를 향상시킬 수 있다. 전자 장치(1000)가 가중치의 타입에 기초하여 가중치를 부호화하는 구체적인 방법은 후술하는 도 6 내지 도 8을 참조하여 구체적으로 설명하기로 한다.

[0027] 도 2는 일 실시 예에 따른 인공 신경망의 가중치를 처리하는 방법의 흐름도이다.

[0028] S210에서, 전자 장치(1000)는 인공 신경망 내 레이어들 및 레이어들 간의 연결 강도에 관한 가중치를 양자화할 수 있다. 예를 들어, 전자 장치(1000)는 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 제1 양자화 함수를 결정하고, 결정된 양자화 함수에 따라 양자화된 가중치들을 포함하는 인공 신경망의 제1 손실 함수를 최소화하도록 가중치를 양자화할 수 있다.

[0029] 보다 상세하게는, 전자 장치(1000)는 양자화된 가중치들을 포함하는 인공 신경망에 입력 데이터를 입력 시키고, 입력 데이터에 대응되는 인공 신경망의 출력 값과 정답 값의 차이에 관한 제1 손실 함수를 결정할 수 있다. 전자 장치(1000)는 양자화된 가중치를 포함하는 인공 신경망의 제1 손실 함수가 최소화되도록 가중치를 양자화 할 수 있다. 전자 장치(1000)가 가중치를 양자화 하는 구체적인 방법은 도 3을 참조하여 더 구체적으로 설명하기로 한다.

[0030] S220에서, 전자 장치(1000)는 가중치 정규화 함수를 이용하여 양자화된 가중치를 정규화할 수 있다. 예를 들어, 전자 장치(1000)는 과적합을 피하기 위한 회귀 방법론들 중, 미리 설정된 제약 조건에 따라 결정되는 가중치 정규화 함수를 이용하여 양자화되는 가중치들을 정규화할 수 있다. 도 2에서는 전자 장치(1000)가 가중치를 양자화 한 후, 양자화된 가중치들을 정규화(regularizing)하는 것으로 도시되었으나, 이에 한정되는 것은 아니며, 전자 장치(1000)가 가중치들을 양자화 하는 동작 및 가중치들을 미리 설정된 제약 조건에 따라 결정되는 정규화 함수를 이용하여 정규화 하는 동작은 함께 수행될 수도 있다. 전자 장치(1000)가 가중치를 정규화 하는 구체적인 방법은 후술하는 도 4 내지 5를 참조하여 구체적으로 설명하기로 한다.

[0031] S230에서, 전자 장치(1000)는 정규화된 가중치의 타입을 식별(identifying)할 수 있다. 예를 들어, 전자 장치(1000)는 양자화된 가중치들을 포함하는 인공 신경망의 제1 손실 함수 및 양자화된 가중치들을 정규화함에 따라 발생하는 인공 신경망의 제2 손실 함수가 최소화되도록 인공 신경망의 가중치들을 수정(modify) 및 갱신(refine)할 수 있다. 전자 장치(1000)는 제1 손실 함수 및 제2 손실 함수가 최소화되는 인공 신경망의 가중치와 미리 설정된 임계치를 비교함하고, 비교 결과에 기초하여 가중치들을 제1 타입의 가중치 또는 제2 타입의 가중치로 분류(classify)할 수 있다. 전자 장치(1000)가 정규화된 가중치들의 타입을 식별하는 구체적인 방법은 도 6을 참조하여 구체적으로 설명하기로 한다.

[0032] S240에서, 전자 장치(1000)는 식별된 가중치의 타입에 기초하여 가중치의 적어도 일부를 부호화할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 가중치의 전부 또는 일부를 부호화(encoding)할 수 있다. 전자 장치(1000)가 가중치의 타입에 기초하여 가중치를 부호화하는 구체적인 방법은 도 7 내지 도 8을 참조하여 구체적으로 설명하기로 한다.

[0033] 또한, 일 실시 예에 의하면, 도 2에는 도시되지 않았지만, 전자 장치(1000)는 S210단계에 앞서, 인공 신경망 내 가중치들이 양자화 되기 전, 사전 학습된(pre-trained) 가중치에 기초하여 인공 신경망 내 가중치들을 초기화하고, 초기화된 인공 신경망 내 가중치들을 양자화할 수 있다.

[0034] 예를 들어, 전자 장치(1000)는 활성화 함수로써 ReLU1를 사용하여 활성화된 값들의 범위를 제한함으로써, 인공 신경망 내 활성화 함수의 출력을 양자화할 수 있다. 예를 들어, 양자화된 활성화 함수의 출력 값은 하기 수학식 1과 같이 표현될 수 있다.

수학식 1

$$a = ReLU1(x) = clip(x, 0, 1) = min(max(0, x), 1)$$

[0035]
$$a_q = Q_a(a, k) = round(a \cdot (2^k - 1)) / (2^k - 1)$$

[0036] 여기에서 a는 활성화 함수 ReLU1의 출력 값이고, 활성화 함수 ReLU1은 입력 변수 x가 0보다 작을 경우 0을 출력하고, 입력 변수가 0에서 1사이에 위치할 경우 입력 변수를 그대로 출력하며, 입력 변수가 1보다 클 경우 1을 출력하는 클리핑 함수에 대응될 수 있다. 또한, 활성화 함수 ReLU1은 max(0, x)의 출력 값을 입력으로 하는 min

함수에 대응될 수도 있다. 또한 a_q 는 양자화된 활성화 함수의 출력 값을 나타내고, Q_a 활성화 함수의 출력 값을 양자화하기 위한 양자화 함수로써, 0과 1사이의 단계의 수를 나타내는 변수 k 및 활성화 함수의 출력 a 를 입력으로 하는 round 함수(예컨대 반올림 함수)를 이용하여 정의될 수 있다. 본 개시에 따른 전자 장치(1000)는 인공 신경망의 가중치들이 양자화 되기 전, Relu1 활성화 함수를 통하여 사전 학습된 FP(Full Precision) 모델의 가중치를 이용하여 인공 신경망을 초기화함으로써 인공 신경망의 정확도를 더 향상시킬 수 있다.

[0037] 도 3은 일 실시 예에 따른 인공 신경망의 가중치를 양자화하는 구체적인 방법을 설명하기 위한 도면이다.

[0038] S320에서, 전자 장치(1000)는 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 제1 양자화 함수를 결정할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 하기의 수학적 식 2에 따라 결정되는 제1 양자화 함수를 이용하여 가중치들을 양자화할 수 있다.

수학적 식 2

$$w_q = Q_w(w_c, rng, s) = round(w_c \cdot \frac{s-1}{2}) \cdot \frac{rng}{2}, \quad w_c = clip(w, -rng, rng)$$

[0039]

[0040] 여기에서, w_q 는 양자화된 가중치이고, Q_w 는 제1 양자화 함수, w_c 클리핑된 가중치, rng 는 가중치들의 고정 범위로써, 사전 학습된 가중치들의 최대 절대 값, s 는 제2 타입의 가중치들의 비트 수로부터 유도된 비트수, round 함수는 반올림 함수, clip 함수는 clip 함수로 입력된 가중치가 $-rng$ 보다 작으면, $-rng$ 를 출력하고, 입력된 가중치가 rng 보다 크고, rng 보다 작으면 입력된 가중치의 값을 그대로 출력하며, 입력된 가중치가 rng 보다 크면 rng 를 출력하는 함수이다. 일 실시 예에 의하면, rng 값은 인공 신경망의 사전 학습(pre-training) 단계에서 결정되고, 양자화 손실을 최소화하도록 하는 재 학습(re-training) 단계에서는 고정될 수 있다.

[0041] 즉, 수학적 식 2에 도시된 바와 같이 일 실시 예에 의하면, 제1 양자화 함수는 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 상기 초기화된 가중치들을 클리핑하는 클리핑 함수의 출력 값 및 상기 가중치의 비트 수를 입력으로 하는 라운드 함수일 수 있다. 상기 수학적 식에 따른 양자화 함수는 클립 함수를 양자화 함수에 입력함으로써 하기 수학적 식 3과 같이 다시 표현될 수도 있다.

수학적 식 3

$$w_q = Q_w(clip(w, -M_{wp}, M_{wp}), M_{wp}, s_t + s_{sl})$$

[0042]

[0043] 여기에서, w_q 는 양자화된 가중치이고, Q_w 는 제1 양자화 함수, M_{wp} 는 사전 학습된 가중치들의 최대 절대 값, clip은 clip 함수로써 입력된 가중치가 $-M_{wp}$ 보다 작으면, $-M_{wp}$ 를 출력하고, 입력된 가중치가 M_{wp} 보다 크고, M_{wp} 는 보다 작으면 입력된 가중치의 값을 그대로 출력하며, 입력된 가중치가 M_{wp} 보다 크면 M_{wp} 를 출력하는 함수이고, s_t 는 제1 타입의 가중치를 나타내는데 필요한 양자화 상태수, s_{sl} 은 제2 타입의 가중치를 나타내는데 필요한 양자화 상태수를 나타낸다.

[0044] 본 개시에 따른 전자 장치(1000)는 가중치를 양자화 하기 전, 클립함수 w_c 를 이용하여 클리핑함으로써 라운드 함수에 따른 양자화 오류(misquantization)를 방지할 수 있다.

[0045] S340에서, 전자 장치(1000)는 제1 양자화 함수에 따라 양자화된 가중치들을 포함하는 인공 신경망의 제1 손실 함수를 최소화하도록 가중치를 양자화할 수 있다. 예를 들어, 전자 장치(1000)는 상기 수학적 식 2에 기초하여 결정되는 제1 양자화 함수를 이용하여 가중치를 양자화한 후, 양자화된 가중치들을 이용하여 인공 신경망의 가중치를 수정 및 갱신할 수 있다. 전자 장치(1000)는 양자화된 가중치를 포함하는 인공 신경망에 입력 데이터를 인가한 후, 입력 데이터에 대한 인공 신경망의 출력 값 및 정답 값의 차이에 관한 제1 손실 함수를 결정하고, 결정된 제1 손실 함수가 최소화되도록 가중치를 양자화할 수 있다. 일 실시 예에 의하면 제1 손실 함수는 평균 제곱 오차 또는 교차 엔트로피 오차를 포함할 수 있다.

[0046] 도 4는 일 실시 예에 따른 인공 신경망의 가중치를 정규화하는 구체적인 방법을 설명하기 위한 도면이다.

[0047] S420에서, 전자 장치(1000)는 부분 L1 제약 조건 및 L2 제약 조건에 기초하여 제1 가중치 정규화 함수를 결정할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)가 이용하는 제1 가중치 정규화 함수는 하기의 수학적 식 4에 기초하여 결정될 수 있다.

수학적 식 4

[0048]
$$WR(w, \lambda_1, \lambda_2) = \lambda_1(L2(w) + \lambda_2 \cdot M_{w_p} \cdot pL1(w))$$

[0049] 여기에서, WR은 가중치 w, 제1 하이퍼파라미터 및 제2 하이퍼파라미터를 입력으로 하는 제1 가중치 정규화 함수이다. 또한, λ_1 및 λ_2 는 하이퍼 파라미터들로서 제1 가중치 정규화 함수의 정규화(또는 규제화)의 강도(intensity)를 조절하기 위한 변수이고, pL1은 과적합(over fitting)을 막기 위한 회귀 방법론 중 하나로써 부분 L1 제약 조건(part of L1, pL1)을 따르는 정규화 함수이고, L2는 L2 제약 조건(Ridge)을 따르는 정규화 함수이며, M_{w_p} 는 인공 신경망 내 가중치들이 양자화 되기 전 사전 학습된 가중치들의 최대 절대 값을 나타낸다. L2 및 pL1 제약 조건에 따른 가중치의 손실은 후술하는 도 5를 참조하여 더 구체적으로 설명하기로 한다.

[0050] S440에서, 전자 장치(1000)는 제1 가중치 정규화 함수에 따라 정규화된 가중치들을 포함하는 인공 신경망의 제2 손실 함수를 최소화하도록 가중치를 정규화할 수 있다. 예를 들어, 전자 장치(1000)는 상기 수학적 식 4에 기초하여 결정되는 제1 가중치 정규화 함수를 이용하여 양자화된 가중치를 정규화한 후, 정규화된 가중치를 포함하는 인공 신경망에 입력 데이터를 인가한 후, 입력 데이터에 대한 인공 신경망의 출력 값 및 정답 값의 차이에 관한 제2 손실 함수를 결정하고, 결정된 제2 손실 함수가 최소화되도록 가중치를 정규화할 수 있다.

[0051] 도 5는 일 실시 예에 따른 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 나타내는 도면이다.

[0052] 도 5의 차트 502를 참조하면 L2 제약 조건을 따르는 정규화 함수의 손실 곡선(506)과 부분 L1 제약 조건(pL1)을 따르는 정규화 함수의 손실 곡선(508)이 도시되어 있다. 차트 502의 점선은 제1 타입의 가중치 및 제2 타입의 가중치를 구별하기 위한 임계치(thresh)를 나타낸다. 부분 L1 제약 조건(pL1)을 따르는 정규화 함수는 하기 수학적 식 5에 기초하여 결정될 수 있다.

수학적 식 5

[0053]
$$pL1(w) = \begin{cases} |w| - thresh & \text{if } |w| > thresh \\ 0 & \text{if } |w| \leq thresh \end{cases}, \quad (thresh = M_{w_p} \cdot \frac{s_t}{s_t + s_{sl} - 1})$$

[0054] 여기에서 pL1은 부분 L1 제약 조건을 따르는 정규화 함수로써, 가중치 w의 절대 값이 임계치보다 큰 경우, 가중치의 절대 값에서 임계치만큼의 차이 값을 출력 값으로 하고, 가중치 w의 절대 값이 임계치 보다 작은 경우, 0의 값을 출력으로 하는 정규화 함수이다. 또한, 여기에서 M_{w_p} 는 사전 학습된 가중치들의 최대 절대 값이고, s_t 는 제1 타입의 가중치를 표현하는데 필요한 비트 수, s_{sl} 은 제2 타입의 가중치를 표현하는데 필요한 비트 수를 나타낼 수 있다.

[0055] 도 5의 차트 512를 참조하면, 전자 장치(1000)가 이용하는 제1 가중치 정규화 함수의 손실 곡선이 도시된다. 일 실시 예에 의하면, 제1 가중치 정규화 함수는 L2 제약 조건을 따르는 정규화 함수 및 부분 L1 제약 조건(pL1)을 따르는 정규화 함수를 더 함으로써 정의될 수 있다. 도 5의 차트 512를 참조하면 제1 가중치 정규화 함수는 L2 제약 조건을 따르는 정규화 함수 및 부분 L1 제약 조건(pL1)을 따르는 정규화 함수의 합으로 정의되므로, 임계치 부근에서 더 가파른 기울기를 가지는 손실 곡선을 나타낼 수 있다.

[0056] 도 5의 차트 522를 참조하면, 본 개시에 따른 제1 가중치 정규화 함수의 기울기가 도시된다. 전자 장치(1000)는 도 5의 차트 522의 기울기를 나타내는 제1 가중치 정규화 함수를 이용하여 가중치들을 정규화 하기 때문에, 임계치(thresh) 이상의 가중치에 더 큰 강도의 정규화 는 규제화)를 수행함으로써, 상당수의 가중치가 임계치 사이에 존재하도록 변환할 수 있다.

[0057] 도 6은 일 실시 예에 따른 가중치의 타입을 식별하는 방법을 구체적으로 설명하기 위한 도면이다.

[0058] S620에서, 전자 장치(1000)는 양자화 과정에서 정규화된 가중치의 절대 값 및 기 설정된 임계치를 비교한다. S640에서, 전자 장치(1000)는 비교 결과에 기초하여 정규화된 가중치를 -1, 0 또는 1 중 하나 또는, -1 에서 1 사이의 임의의 값을 나타내는 제1 타입의 가중치로 식별하거나, 상기 가중치 중, 제1 타입의 가중치가 아닌 가중치를 제2 타입의 가중치로 식별할 수 있다. 전자 장치(1000)가 가중치의 타입을 식별하는 동작은 하기의 수학적 4에 기초하여 정의될 수 있다.

수학적 6

$$mask(w) = \begin{cases} TW & \text{if } |w| \leq thresh \\ SLW & \text{if } |w| > thresh \end{cases}$$

[0060] 여기에서 mask(w)는 가중치의 타입을 식별하기 위한 mask함수이고, 양자화 과정에서 정규화된 가중치의 절대 값이 기 설정된 임계치 보다 작은 경우, 가중치를 삼항 가중치(Ternary Weight, TW)로 분류하고, 가중치의 절대 값이 기 설정된 임계치 보다 작은 경우, 큰 가중치(Sparse-large weight)로 분류하는 함수이다. 일 실시 예에 의하면, 삼항 가중치(TW)는 -1과 1사이의 범위를 나타내는 가중치 또는 -1, 0 및 1 중 하나를 나타내는 가중치의 타입을 나타내고, 큰 가중치(SLW)는 가중치들 중, -1 보다 작거나 1보다 큰 값을 나타내는 가중치의 타입을 나타낼 수 있다. 이하에서는, 편의상 삼항 가중치(TW)를 제1 타입의 가중치, 큰 가중치(SLW)를 제2 타입의 가중치로 정의하여 설명하기로 한다.

[0061] 도 7은 일 실시 예에 따른 가중치의 적어도 일부를 부호화하는 방법을 구체적으로 설명하기 위한 도면이다.

[0062] S720에서, 전자 장치(1000)는 제1 타입의 가중치를 2비트로 표현 가능한 양자화 상태에 각각 대응되는 이진 값으로 변환한다. 예를 들어, 전자 장치(1000)는 양자화 과정에서 정규화된 가중치가 제1 타입의 가중치로 식별되는 경우, 제1 타입의 가중치를 0 또는 1 값에 기초하여 이진화할 수 있다. 보다 상세하게는, 전자 장치(1000)는 가중치 w가 1인 경우 01로, 가중치가 0인 경우 00으로, 가중치가 -1인 경우 11의 값을 이용하여 가중치를 이진 값으로 나타낼 수 있다.

[0063] S740에서, 전자 장치(1000)는 제2 타입의 가중치를 상기 이진 값으로 변환된 제1 타입의 가중치에 대응되지 않는 양자화 상태를 이용하여 제2 타입의 가중치를 부호화한다. 전자 장치(1000)가 제1 타입의 가중치를 이진 값으로 변환하는 동작 및 가중치들 중, 제1 타입의 가중치가 아닌 가중치인 제2 타입의 가중치를 부호화하는 동작은 후술하는 도 8을 참조하여 구체적으로 설명하기로 한다.

[0064] 도 8은 일 실시 예에 따른 전자 장치가 가중치의 적어도 일부를 부호화하는 방법을 설명하기 위한 도면이다.

[0065] 본 개시에 따른 전자 장치(1000)는 제1 타입의 가중치는 이진 값으로 변환함으로써 값(value)으로 저장하고, 제2 타입의 가중치는, 제1 타입의 가중치에 대응되지 않는 별도의 양자화 상태 각각에 대응되는 지시자를 이용하여 가중치를 표현(representation)할 수 있다. 즉, 본 개시에 따른 전자 장치(1000)는 제1 타입의 가중치를 값으로 표현하고, 제2 타입의 가중치를 지시자(indices)로 표현함으로써, 인공 신경망 내 가중치들을 혼합 가중치 표현 방법(Hybrid weight representation method)을 이용하여 나타낼 수 있다.

[0066] 먼저, 도 8에 도시된 실시 예 (802)를 참조하여 전자 장치(1000)가 제1 타입의 가중치로써 1, 0 및 -1 를 식별하고, 제2 타입의 가중치로써 3, 2, -2, -3을 식별한 경우를 가정하여 설명하기로 한다. 즉, 제1 타입의 가중치의 비트 수 b_1 는 2이고, 제1 타입의 가중치들의 양자화 상태수 s_1 는 3이며, 제2 타입의 가중치의 비트 수 b_{s1} 은 2이고, 제2 타입의 가중치들의 양자화 상태수 s_{s1} 은 4인 경우를 가정한다.

[0067] 전자 장치(1000)는 인공 신경망의 가중치들의 절대 값이 임계치 1보다 작은 가중치들을 양자화함으로써 -1, 0 또는 1로 식별할 수 있다. 전자 장치(1000)는 제1 타입의 가중치(812)들(예컨대 -1, 0, 1)을 양자화 상태 수(quantization level number)가 4인 양자화 상태들(806)을 이용하여 나타낼 수 있다. 양자화 상태 수가 4인 양자화 상태들은 2비트로 표현 가능한 양자화 상태를 의미할 수 있다. 예를 들어, 전자 장치(1000)는 제1 타입의 가중치(812) 중, 1을 제1 양자화 상태 (01)로, 0을 제2 양자화 상태 (00)로, -1을 제3 양자화 상태 (11)에 각각 매핑(mapping)할 수 있다.

[0068] 또한, 전자 장치(1000)는 제1 타입의 가중치를 나타내는데 사용되지 않는 양자화 상태인 제4 양자화 상태 (00)(808)을 이용하여, 제2 타입의 가중치를 부호화할 수 있다. 예를 들어, 전자 장치(1000)는 제1 타입의 가중

치를 이진 값으로 변환하는데 사용되지 않은 제4 양자화 상태(808)에 대응되는 상태 값 (10)을 이용하여 제2 타입의 가중치(818)중, 가중치 3을 1001로, 가중치 2를 1000으로, 가중치 -2를 1010으로, 가중치 -3을 1011와 같은 지시자(indices)로 부호화(encoding)할 수 있다. 본 개시에 따르면, 전자 장치(1000)가 제2 타입의 가중치를 부호화하기 위하여 사용하는 지시자(indices)들은 제2 타입의 가중치들의 양자화 상태에 각각 대응될 수 있다.

[0069] 도 8의 실시 예 802에 도시된 바와 같이, 본 개시에 따른 전자 장치(1000)는 제1 타입의 가중치는 이진 값으로 변환하여 저장하나, 제2 타입의 가중치는 이진 값으로 변환된 제1 타입의 가중치를 나타내는데 사용되지 않은 양자화 상태를 이용하여 부호화함으로써, 인공 신경망의 정확도를 더 향상시킬 수 있다.

[0070] 도 8에 도시된 실시 예 (822)를 참조하여 전자 장치(1000)가 제1 타입의 가중치로써 1, 0 및 -1을 식별하고, 제2 타입의 가중치로써 5, 4, 3, 2, -2, -3, -4, -5를 식별하는 경우를 가정한다. 즉, 제1 타입의 가중치의 비트 수 b_1 는 2이고, 제1 타입의 가중치들의 양자화 상태수 s_1 는 3으로 실시 예 (802)와 동일하지만, 제2 타입의 가중치의 비트 수 b_{s1} 은 3이고, 제2 타입의 가중치들의 양자화 상태수 s_{s1} 은 8인 경우를 가정하여 설명하기로 한다.

[0071] 도 8의 실시 예 (822)를 참조하면, 전자 장치(1000)가 제1 타입의 가중치를 이진 값으로 변환하는 동작은 실시 예 (802)에 대응될 수 있다. 즉, 전자 장치(1000)는 제1 타입의 가중치(832)들(예컨대 -1, 0, 1)을 양자화 상태 수(quantization level number)가 4인 양자화 상태들(826)을 이용하여 나타낼 수 있다. 예를 들어, 전자 장치(1000)는 제1 타입의 가중치(832) 중, 1을 제1 양자화 상태 (01)로, 0을 제2 양자화 상태 (00)로, -1을 제3 양자화 상태 (11)에 각각 매핑(mapping)할 수 있다. 또한, 전자 장치(1000)는 제1 타입의 가중치를 나타내는데 사용되지 않는 양자화 상태인 제4 양자화 상태(00)(828)을 이용하여, 제2 타입의 가중치를 부호화할 수 있다.

[0072] 예를 들어, 전자 장치(1000)는 제1 타입의 가중치를 이진 값으로 변환하는데 사용되지 않은 제4 양자화 상태 (828)에 대응되는 상태 값 (10)을 이용하여 제2 타입의 가중치(838)중, 가중치 5를 10011로, 가중치 4를 10010으로, 가중치 3을 10001로, 가중치 2를 10000로, 가중치 -2를 10100으로, 가중치 -3을 10101으로, 가중치 -4를 10110으로, 가중치 -5를 10111와 같은 지시자(indices)로 부호화(encoding)할 수 있다.

[0073] 따라서, 본 개시에 따른 전자 장치(1000)는 인공 신경망의 가중치들을 양자화 하는 과정에서 제2 타입의 가중치들을 정규화함으로써, 상당수의 가중치들을 제1 타입의 가중치로 변경하여 사용하기 때문에, 인공 신경망의 가중치 연산을 효율적으로 수행함과 함께, 제1 타입의 가중치는 이진 값으로 표현하고, 제2 타입의 가중치는 이진 값으로 변환된 제1 타입의 가중치를 나타내는데 사용되지 않은 양자화 상태(808, 828)에 대응되는 지시자를 이용하여 부호화함으로써 인공 신경망의 정확도를 향상시킬 수 있다.

[0074] 도 9는 또 다른 실시 예에 따른 전자 장치가 인공 신경망의 가중치를 처리하는 방법의 흐름도이다.

[0075] S910에서, 전자 장치(1000)는 가중치의 타입을 1차로 식별할 수 있다. S910은 도 2의 S230에 대응될 수 있으므로 구체적인 설명은 생략하기로 한다.

[0076] S920에서, 전자 장치(1000)는 1차 식별된 가중치의 타입에 기초하여 제1 양자화 함수와 다른 제2 양자화 함수를 결정할 수 있다. 예를 들어, 전자 장치(1000)는 가중치의 타입을 식별하고, 식별된 가중치의 타입에 따라 서로 다른 양자화 함수를 결정하며, 서로 다른 양자화 함수를 가중치의 타입 별로 다르게 적용함으로써 S210단계에서 양자화된 가중치들을 재 양자화(re-quantizing) 할 수 있다. 전자 장치(1000)가 가중치의 타입에 따라 서로 다른 양자화 함수들은 하기의 수학적 식에 기초하여 결정될 수 있다.

수학적 식 7

$$Q_s(w) = \begin{cases} Q_w(\text{clip}(w, -\text{thresh}, \text{thresh}), \text{thresh}, s_t + 1) & \text{if } \text{mask}(w) = TW \\ Q_w(\text{clip}(w, -M_{w_p}, M_{w_p}), M_{w_p}, s_t + s_{sl}) & \text{if } \text{mask}(w) = SLW \end{cases}$$

[0078] 여기에서, $\text{mask}(w)$ 는 전자 장치(1000)가 가중치의 타입을 1차로 식별하기 위해, 수학적 식 6에서 정의된 mask 함수이고, Q_w 는 식별된 가중치의 타입 별로 생성되는 양자화 함수들이며, s_t 는 제1 타입의 가중치들의 양자화 상태수이고, s_{sl} 은 제2 타입의 가중치들의 양자화 상태수를 나타내며, thresh 는 사전 학습 단계에서 인공 신경망 가중치의 절대 값의 최대 크기에 기초하여 결정되는 임계치를 나타낸다.

[0079] 예를 들어, 전자 장치(1000)는 S910에서 식별된 가중치가 제1 타입의 가중치(예컨대 삼항 가중치, TW)인 경우, 가중치가 $-\text{thresh}$ 보다 작은 경우 $-\text{thresh}$ 를 출력하고, 가중치가 $-\text{thresh}$ 및 thresh 사이의 값을 가지는 경우

가중치를 그대로 출력하며, 가중치가 thresh 보다 큰 경우, thresh를 출력하는 클립 함수, thresh 및 제2 타입의 가중치들의 양자화 상태수 + 1의 값을 입력으로 하는 제2 양자화 함수를 결정할 수 있다.

[0080] 또한, 전자 장치(1000)는 S910에서 식별된 가중치가 제2 타입의 가중치(예컨대 큰 가중치, SLW)인 경우, 가중치가 $-M_{wp}$ 보다 작은 경우 $??M_{wp}$ 를 출력하고, 가중치가 $??M_{wp}$ 및 M_{wp} 사이의 값을 가지는 경우 가중치를 그대로 출력하며, 가중치가 M_{wp} 보다 큰 경우, M_{wp} 를 출력하는 클립 함수, M_{wp} 및 제1 타입의 가중치들의 양자화 상태수 s_1 와 제2 타입의 가중치들의 양자화 상태수 ss_1 의 합을 입력으로 하는 제1 양자화 함수를 결정할 수 있다. 상기 수학식 7에서, 전자 장치(1000)가 식별한 가중치가 제2 타입의 가중치인 경우 이용하는 제1 양자화 함수는 전술한 수학식 3에 기초하여 정의되는 제1 양자화 함수에 대응될 수 있다.

[0081] S930에서, 전자 장치(1000)는 1차로 식별된 가중치의 타입에 따라 제1 양자화 함수 및 제2 양자화 함수를 선택적으로 적용함으로써 가중치를 재 양자화 할 수 있다. 즉, 상기 수학식 7에서 표기한 바와 같이, 전자 장치(1000)는 가중치의 타입에 따라 제1 양자화 함수 또는 제2 양자화 함수를 선택적으로 적용함으로써 가중치를 양자화 하기 때문에, S930에서 전자 장치가 가중치를 재 양자화 하는 동작은, 가중치의 선택적 양자화(Selective Quantization) 동작에 대응될 수 있다.

[0082] 예를 들어, 전자 장치(1000)는 1차로 식별된 가중치의 타입이 제1 가중치 타입에 해당하는 경우, 제2 양자화 함수를 이용하여 가중치를 재 양자화 할 수 있다. 보다 상세하게는, 전자 장치(1000)는 제2 양자화 함수에 따라 양자화된 가중치들을 포함하는 인공 신경망의 제3 손실 함수가 최소화되도록 가중치를 양자화할 수 있다. 예를 들어, 전자 장치(1000)는 제2 양자화 함수를 이용하여 가중치를 양자화한 후, 양자화된 가중치들을 이용하여 인공 신경망의 가중치를 수정 및 갱신하고, 양자화된 가중치를 포함하는 인공 신경망에 입력 데이터를 인가한 후, 입력 데이터에 대한 인공 신경망의 출력 값 및 정답 값의 차이에 관한 손실 함수를 결정하며, 결정된 손실 함수가 최소화되도록 가중치를 양자화한다.

[0083] 또한, 전자 장치(1000)는, 1차로 식별된 가중치의 타입이 제2 가중치 타입에 해당하는 경우, 제1 양자화 함수를 이용하여 가중치를 재 양자화 할 수 있다. 보다 상세하게는, 전자 장치(1000)는 제1 양자화 함수에 따라 양자화된 가중치들을 포함하는 인공 신경망의 손실 함수가 최소화되도록 가중치를 양자화할 수 있다. 예를 들어, 전자 장치(1000)는 제1 양자화 함수를 이용하여 가중치를 양자화한 후, 양자화된 가중치들을 이용하여 인공 신경망의 가중치를 수정 및 갱신하고, 양자화된 가중치를 포함하는 인공 신경망에 입력 데이터를 인가한 후, 입력 데이터에 대한 인공 신경망의 출력 값 및 정답 값의 차이에 관한 손실 함수를 결정하며, 결정된 손실 함수가 최소화되도록 가중치를 양자화할 수 있다.

[0084] 본 개시에 따른 전자 장치(1000)는 S210단계에서 가중치를 양자화 한 후, 양자화된 가중치들의 타입에 기초하여 가중치들을 다시 양자화함으로써 이전 학습 단계에서 학습된 인공 신경망의 가중치들의 값을 미세 조정(fine tune)할 수 있고, 결과적으로 인공 신경망의 성능을 향상시킬 수 있다.

[0085] S940에서, 전자 장치(1000)는 재 양자화된 가중치를 L2 제약 조건에 대응되는 제2 가중치 정규화 함수를 이용하여 재 정규화 할 수 있다. 예를 들어, 전자 장치(1000)는 도 5의 차트 502에서 도시된 pL1 제약 조건에 대응되는 정규화 함수를 제1 가중치 정규화 함수에서 제거함으로써 제2 가중치 정규화 함수를 결정할 수 있다. 전자 장치(1000)가 도 2의 S220 단계를 수행한 이후의 인공 신경망의 가중치들은 임계치 이하로 집중 분포되어 있는 상태이다. 따라서, 본 개시에 따른 전자 장치(1000)는 제1 정규화 함수로부터 pL1 제약 조건에 따른 정규화 함수의 영향을 제거함으로써 결정되는 제2 정규화 함수를 이용하여 가중치를 재 정규화함으로써 가중치 식별의 정확성을 향상시킬 수 있다.

[0086] 일 실시 예에 의하면, 전자 장치(1000)는 제2 가중치 정규화 함수에 따라 재 정규화된 가중치들을 포함하는 인공 신경망의 제4 손실 함수를 최소화하도록 재 양자화된(예컨대 selective quantized)가중치를 재 정규화할 수 있다. 보다 상세하게는, 전자 장치(1000)는 제2 가중치 정규화 함수에 따라 정규화된 가중치들을 재 정규화한 후, 정규화된 가중치를 포함하는 인공 신경망에 입력 데이터를 인가한 후, 입력 데이터에 대한 인공 신경망의 출력 값 및 정답 값의 차이에 관한 제4 손실 함수를 결정하고, 결정된 제4 손실 함수가 최소화되도록 가중치를 수정 및 갱신함으로써 가중치를 재 정규화할 수 있다.

[0087] S950에서, 전자 장치(1000)는 재 정규화된 가중치의 타입을 식별할 수 있다. 전자 장치(1000)가 재 정규화된 가중치의 타입을 식별하는 동작은 도 6의 전자 장치(1000)가 가중치의 타입을 식별하는 동작 S620 내지 S640에 대응될 수 있으므로 구체적인 설명은 생략하기로 한다.

- [0088] S960에서, 전자 장치(1000)는 재 정규화된 가중치의 타입에 기초하여 재 정규화된 가중치의 적어도 일부를 부호화할 수 있다. 전자 장치(1000)가 재 정규화된 가중치의 타입에 기초하여, 가중치의 적어도 일부를 부호화하는 동작은 도 7 내지 도 8에서 전자 장치(1000)가 가중치의 적어도 일부를 부호화하는 동작에 대응될 수 있으므로 구체적인 설명은 생략하기로 한다.
- [0089] 또한, 일 실시 예에 의하면 도 9에는 도시되지 않았지만, 전자 장치(1000)는 S960단계 이후에, 이진 값으로 변환된 제1 타입의 가중치 및 지시자(indices)를 이용하여 부호화된 제2 타입의 가중치를 모두 전자 장치(1000)의 메모리에 저장하는 단계를 더 수행할 수도 있다.
- [0090] 도 10은 일 실시 예에 따른 전자 장치가 식별된 가중치의 타입에 따라 서로 다르게 양자화된 가중치들을 설명하기 위한 도면이다.
- [0091] 도 10을 참조하면, 전자 장치(1000)가 식별한 가중치의 타입에 따라 클리핑된 가중치(1004, 1014) 및 양자화된 가중치(1006, 1016)가 도시된다. 일 실시 예에 의하면, 전자 장치(1000)가 식별한 가중치의 타입이 제1 타입의 가중치(1002)인 경우, 수학적 7의 클리핑 함수 clip에 따라 클리핑된 가중치들(1004)은 가중치 범위가 -1 보다 기 설정된 간격만큼 이격된 곳에서 -1보다 소정의 임계치 보다 작은 값에 수렴하고, -1 에서 1사이에서는 선형으로 증가하며, 가중치 범위가 1보다 큰 범위에서는 1보다 소정의 임계치만큼 큰 값에 수렴할 수 있다. 또한, 클리핑 함수를 입력으로 하는 제2 양자화 함수에 따라 양자화된 가중치들(1006)은 제1 타입의 가중치로써, -1, 0 또는 1에 수렴하는 것을 관측할 수 있다.
- [0092] 또한, 일 실시 예에 의하면, 전자 장치(1000)가 식별한 가중치의 타입이 제2 타입의 가중치(1012)인 경우, 수학적 7의 클리핑 함수 clip에 따라 클리핑된 가중치들(1014)은 가중치 범위가 -3 보다 작은 범위에서 -3 값에 수렴하고, -3 에서 3사이에서는 선형으로 증가하며, 가중치 범위가 3보다 큰 범위에서는 3에 수렴할 수 있다. 또한, 클리핑 함수를 입력으로 하는 제1 양자화 함수에 따라 양자화된 가중치들(1016)은 제2 타입의 가중치로써, -3, -2, 2, 또는 3에 수렴하는 것을 관측할 수 있다.
- [0093] 도 11은 일 실시 예에 따른 인공 신경망의 가중치를 처리하는 전자 장치의 블록도이다.
- [0094] 도 11에 도시된 바와 같이, 인공 신경망의 가중치를 처리하는 전자 장치는 프로세서(1400) 및 메모리(1402)를 포함할 수 있다. 그러나, 도시된 구성 요소가 모두 필수구성요소인 것은 아니고, 도시된 구성 요소보다 많은 구성 요소에 의해 전자 장치(1000)가 구현될 수도 있고, 그보다 적은 구성 요소에 의해서도 전자 장치(1000)는 구현될 수도 있다. 일 실시 예에 의하면, 전자 장치(1000)는 프로세서(1400) 및 메모리(1402)외에 통신부(미도시)를 더 포함할 수도 있다.
- [0095] 프로세서(1400)는, 통상적으로 전자 장치(1000)의 전반적인 동작을 제어한다.
- [0096] 일 실시 예에 의하면, 본 개시에 따른 프로세서(1400)는 메모리(1402)에 저장된 프로그램들을 실행함으로써, 도 1 내지 도 10에 기재된 전자 장치(1000)의 기능을 수행할 수 있다. 또한, 프로세서(1400)는 하나 또는 복수의 프로세서로 구성될 수 있고, 하나 또는 복수의 프로세서는 CPU, AP, DSP(Digital Signal Processor) 등과 같은 범용 프로세서, GPU와 같은 그래픽 전용 프로세서 또는 인공지능(AI) 전용 프로세서일 수 있다. 일 실시 예에 의하면, 프로세서(1400)가 범용 프로세서, 인공지능 프로세서 및 그래픽 전용 프로세서를 포함하는 경우, 인공지능 프로세서는 범용 프로세서 또는 그래픽 전용 프로세서와 별도의 칩으로 구현될 수도 있다.
- [0097] 일 실시 예에 의하면, 프로세서(1400)가 복수의 프로세서 또는 그래픽 전용 프로세서 또는 인공지능 전용 프로세서로 구현될 때, 복수의 프로세서 또는 그래픽 전용 프로세서 또는 인공지능 전용 프로세서 중 적어도 일부는 전자 장치(1000) 및 전자 장치(1000)와 연결된 다른 전자 장치 또는 서버에 탑재될 수도 있다.
- [0098] 예를 들어, 프로세서(1400)는, 메모리(1402)에 저장된 프로그램들을 실행함으로써, 전자 장치에 저장된 인공 신경망 내 가중치들을 제1 타입의 가중치 또는 제2 타입의 가중치로 양자화하고, 양자화 하는 과정에서 제2 타입의 가중치를 희소화할 수 있으며, 제1 타입의 가중치를 값으로 나타내고, 제2 타입의 가중치를 지시자를 이용하여 부호화함으로써 인공 신경망 내 가중치를 빠르게 처리함과 함께 인공 신경망의 정확도를 향상시키도록 할 수 있다.
- [0099] 일 실시 예에 의하면, 프로세서(1400)는 상기 인공 신경망 내 레이어들 및 상기 레이어들 간의 연결 강도에 관한 가중치(weight)를 양자화 하고, 상기 양자화된 가중치를 상기 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 이용하여 정규화 하고, 상기 정규화된 가중치의 타입을 식별하고, 상기 식별된 가중치의 타입에 기초하여, 상기 가중치의 적어도 일부를 부호화할 수 있다.

- [0100] 또한, 프로세서(1400)는 상기 인공 신경망 내 가중치들이 양자화 되기 전, 사전 학습된(pre-trained) 가중치에 기초하여 상기 인공 신경망 내 가중치들을 초기화 하고, 상기 초기화된 인공 신경망 내 가중치들을 양자화할 수 있다.
- [0101] 또한, 프로세서(1400)는 인공 신경망 내 사전 학습된 가중치의 크기에 기초하여 제1 양자화 함수를 결정하고, 상기 결정된 제1 양자화 함수에 따라 양자화된 가중치들을 포함하는 상기 인공 신경망의 제1 손실 함수를 최소화하도록 상기 가중치를 양자화할 수 있다.
- [0102] 또한, 프로세서(1400)는 부분 L1 제약 조건 및 L2 제약 조건에 기초하여 제1 가중치 정규화 함수를 결정하고, 상기 결정된 제1 가중치 정규화 함수에 따라 정규화된 가중치들을 포함하는 상기 인공 신경망의 제2 손실 함수를 최소화하도록 상기 가중치를 정규화할 수 있다.
- [0103] 또한, 프로세서(1400)는 상기 가중치의 절대 값 및 기 설정된 임계치를 비교하고, 상기 비교 결과에 기초하여, 상기 정규화된 가중치를 -1, 0 또는 1 중 하나를 나타내는 제1 타입의 가중치 또는 상기 가중치 중, 상기 제1 타입의 가중치가 아닌 가중치를 제2 타입의 가중치로 식별할 수 있다.
- [0104] 또한, 프로세서(1400)는 제1 타입의 가중치를, 2비트로 표현 가능한 양자화 상태에 각각 대응되는 이진 값으로 변환하고, 상기 제2 타입의 가중치를, 상기 이진 값으로 변환된 제1 타입의 가중치에 대응되지 않는 양자화 상태를 이용하여 상기 제2 타입의 가중치를 부호화할 수 있다.
- [0105] 또한, 프로세서(1400)는 상기 식별된 가중치의 타입에 기초하여 상기 제1 양자화 함수와 다른 제2 양자화 함수를 결정하고, 상기 식별된 가중치의 타입에 따라 상기 제1 양자화 함수 및 상기 제2 양자화 함수를 선택적으로 적용함으로써, 상기 가중치를 재 양자화 하고, 상기 재 양자화된 가중치를 상기 L2 제약 조건에 대응되는 제2 가중치 정규화 함수를 이용하여 재 정규화할 수 있다.
- [0106] 통신부(미도시)는, 전자 장치(1000)가 다른 장치(미도시) 및 서버(2000)와 통신을 하게 하는 하나 이상의 구성요소를 포함할 수 있다. 다른 장치(미도시)는 전자 장치(1000)와 같은 컴퓨팅 장치이거나, 센싱 장치일 수 있으나, 이에 제한되지 않는다. 예를 들어, 통신부(미도시)는, 근거리 통신부, 이동 통신부를 포함할 수 있다.
- [0107] 근거리 통신부(short-range wireless communication unit) 는, 블루투스 통신부, BLE(Bluetooth Low Energy) 통신부, 근거리 무선 통신부(Near Field Communication unit), WLAN(와이파이) 통신부, 지그비(Zigbee) 통신부, 적외선(IrDA, infrared Data Association) 통신부, WFD(Wi-Fi Direct) 통신부, UWB(ultra wideband) 통신부, 등을 포함할 수 있으나, 이에 한정되는 것은 아니다. 이동 통신부는, 이동 통신망 상에서 기지국, 외부의 단말, 서버 중 적어도 하나와 무선 신호를 송수신한다.
- [0108] 일 실시 예에 의하면, 통신부(미도시)는 프로세서의 제어에 의하여, 서버로 인공 신경망 내 가중치 값들, 가중치들을 포함하는 인공 신경망의 손실 함수의 값, 손실 기울기 값 등을 전송할 수 있고, 서버로부터 수정된 가중치 값들, 손실 함수의 값, 기울기 값 등을 수신할 수도 있다.
- [0109] 메모리(1402)는, 프로세서(1400)의 처리 및 제어를 위한 프로그램을 저장할 수 있고, 전자 장치(1000)로 입력되거나 전자 장치(1000)로부터 출력되는 데이터를 저장할 수도 있다. 또한, 메모리(1402)는 인공 신경망을 구성하는 레이어들, 레이어들에 포함된 노드들 및 레이어들의 연결 강도에 관한 가중치들에 대한 정보와 가중치 정규화 함수, 양자화 함수, 식별된 가중치의 타입 및 식별된 가중치의 타입에 따라 부호화된 가중치 값들을 저장할 수 있다. 즉, 메모리(1402)는 인공 신경망 내 가중치들이 수정 및 갱신될 경우, 수정 및 갱신된 가중치에 관한 정보를 더 저장할 수 있다.
- [0110] 메모리(1402)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(RAM, Random Access Memory) SRAM(Static Random Access Memory), 롬(ROM, Read-Only Memory), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.
- [0111] 도 12는 일 실시 예에 따른 전자 장치와 통신 가능한 서버의 블록도이다.
- [0112] 도 12는 일 실시 예에 따른 보이스 어시스턴트 서비스를 제공하는 서버의 블록도이다.
- [0113] 일 실시 예에 따르면, 서버(2000)는 통신부(2100), 데이터 베이스(Data Base, 2200) 및 프로세서(2300)를 포함할 수 있다.

- [0114] 통신부(2100)는 상술한 전자 장치(1000)의 통신부(미도시)에 대응될 수 있다. 예를 들어, 통신부(2100)는 전자 장치(1000)로부터 인공 신경망의 레이어들 및 레이어들에 포함된 노드에 관한 정보 또는 신경망 내 레이어들의 연결 강도에 관한 가중치 값들을 수신할 수 있다.
- [0115] 데이터 베이스(2200)는 도 11에 도시된 전자 장치의 메모리(1402)에 대응될 수 있다. 예를 들어, 데이터 베이스(2200)는 프로세서(2300)의 처리 및 제어를 위한 프로그램을 저장할 수 있고, 전자 장치(1000)로 입력되거나 전자 장치(1000)로부터 출력되는 데이터를 저장할 수도 있다. 또한, 데이터 베이스(2200)는 인공 신경망을 구성하는 레이어들, 레이어들에 포함된 노드들 및 레이어들의 연결 강도에 관한 가중치들에 대한 정보와 가중치 정규화 함수, 양자화 함수, 식별된 가중치의 타입 및 식별된 가중치의 타입에 따라 부호화된 가중치 값들을 저장할 수 있다. 또한, 데이터 베이스(2200)는 인공 신경망 내 가중치들이 수정 및 갱신될 경우, 수정 및 갱신된 가중치에 관한 정보를 더 저장할 수도 있다.
- [0116] 프로세서(2300)는 통상적으로 서버(2000)의 전반적인 동작을 제어한다. 예를 들어, 프로세서(2300)는, 서버(2000)의 DB(2200)에 저장된 프로그램들을 실행함으로써, DB(2200) 및 통신부(2100) 등을 전반적으로 제어할 수 있다. 또한, 프로세서(2300)는 DB(2100)에 저장된 프로그램들을 실행함으로써, 도 1 내지 도11에서의 전자 장치(1000)의 동작의 일부를 수행할 수 있다.
- [0117] 예를 들어, 프로세서(2300)는 전자 장치(1000)에 저장된 인공 신경망의 초기화된 가중치들을 수신함으로써, 수신된 가중치들을 양자화하고, 양자화 과정에서 가중치들을 정규화하며, 정규화된 가중치의 타입을 식별하고, 식별된 가중치의 타입에 따라 가중치의 적어도 일부를 부호화할 수 있다.
- [0118] 도 13은 일 실시 예에 따라 전자 장치가 이용하는 인공 신경망의 성능을 설명하기 위한 도면이다.
- [0119] 도 13을 참조하면, 전자 장치(1000)가 ImageNet dataset의 데이터를 기초로 인공 신경망을 학습시킨 후, 모델명(1322), 비트 폭(1324), 활성화 함수의 출력 값을 양자화하는데 필요한 비트수(1326), 제2 하이퍼파라미터(1328), 제2 타입의 가중치인 SLW 가중치의 비율(1332), 전자 장치가 가중치를 부호화함에 따라 부호화된 평균 비트 길이(1332), 양자화 및 정규화 방법(1336)에 따라 다른 인공 신경망의 정확도(Accuracy)가 도시된다.
- [0120] 일 실시 예에 의하면, 전자 장치(1000)는 딥러닝 기반의 인공 신경망으로써 ResNet-18를 기반으로, 비트 폭이 2/3이며, 활성화 함수의 출력 값을 양자화하는데 필요한 비트수가 4이고, SLW 가중치가 0.54% 존재하며, 평균 부호화된 비트 길이가 2.016이고, 식별된 가중치의 타입에 따라 서로 다른 양자화 함수를 적용하며, 양자화 과정에서 선택적 양자화(Selective Quantization)만을 수행할 경우, 57.57%의 정확도를 나타내는 신경망을 학습시킬 수 있다.
- [0121] 또 다른 실시 예에 의하면, 전자 장치(1000)가 학습시키는 인공 신경망은, 전자 장치(1000)가 딥러닝 기반의 인공 신경망으로써 ResNet-18를 기반으로, 비트 폭이 2/4이며, 활성화 함수의 출력 값을 양자화하는데 필요한 비트수가 4, SLW 가중치가 0.269% 존재하며, 평균 부호화된 비트 길이가 2.011이고, 식별된 가중치의 타입에 따라 서로 다른 양자화 함수를 적용함으로써 선택적 양자화 (Selective Quantization, SQ)만을 수행할 경우, 59.00%의 정확도를 나타낼 수 있다.
- [0122] 도 14는 또 다른 실시 예에 따라 전자 장치가 이용하는 인공 신경망의 성능을 설명하기 위한 도면이다.
- [0123] 도 14를 참조하면, 도 14에는 전자 장치(1000)가 학습시킨 인공 신경망의 모델명(1422), 양자 플랜 (Quantization Plan)(1424), 제2 하이퍼 파라미터(1425) 및 이용된 가중치 정규화 함수를 결정하기 위한 제약조건(1428)에 따른 인공 신경망의 정확도가 도시된다.
- [0124] 일 실시 예에 의하면, 전자 장치(1000)가 Quantization plan(예컨대, 비트폭, 1424)가 2/4이고, 제2 하이퍼 파라미터가 1로 설정하며, L2 제약 조건을 따르는 정규화 함수 및 부분 L1 제약 조건(pL1)을 따르는 정규화 함수를 모두 이용하여, AlexNet 모델학습 시키는 경우, 인공 신경망은 76.88%의 정확도를 나타낼 수 있다.
- [0125] 도 15는 일 실시 예에 따른 전자 장치 내 인공 신경망의 동작을 설명하기 위한 도면이다.
- [0126] 도 15를 참조하여, 일 실시 예에 따른, 전자 장치(1000)가 이용하는 인공 신경망의 구체적인 동작을 설명하기로 한다. 예를 들어, 본 개시에 따른 전자 장치(1000)가 이용하는 인공 신경망이 ResNet구조의 신경망인 경우, ResNet의 add operation 전, ResNet의 각 블록 으로부터, 마지막 컨벌루션 레이어들의 출력은 컨벌루션 레이어들의 최대 절대 값으로부터 유도되는 스케일 팩터 r에 의하여 스케일링될 수 있다.
- [0127] 본 개시에 따른 전자 장치(1000)가 가중치를 처리하는 방법에 따르면, 인공 신경망 내 가중치들 및 활성화 함수

의 출력은 동일한 간격($w_q=i_w??I_w$, $a_q=i_a??I_a$)으로 양자화될 수 있다. 여기에서 I_w 및 i_a 는 fixed float 간격이고, i_w 및 i_a 는 정수 변수이다. 따라서, 인공 신경망 내 컨벌루션 레이어 및 풀리 커넥티드 레이어는 하기 수학식 8에 의해, integer operation에 의해 추론(inferred)될 수 있다.

수학식 8

$$o_q = layer(a_q, w_q) = \sum \sum a_q w_q = I_a I_w \cdot \sum \sum i_a i_w$$

여기에서, a_q 는 양자화된 활성화 함수의 출력이고, w_q 는 양자화된 가중치이며, I_w 및 i_a 는 fixed float 간격이고, i_w 및 i_a 는 정수 변수이다. 예를 들어, 배치 정규화(Batch Normalization, BN), ReLU1 활성화 함수(1504) 및 활성화 함수의 출력을 양자화하기 위한 양자화 함수 $Qa(??)$ 는 integer comparators에 의해, 누적(integrated)되거나 압축(compresed)될 수 있다. 활성화 함수의 출력을 양자화 하기 위한 함수, 활성화 함수의 경계 및 배치 정규화의 계수들은 미리 정렬되기 위하여 고정(fixed)될 수 있다.

그러나, ResNet은 layer-wise intervals로 인하여, 데이터 스트림들이 동일한 양자화 간격들을 유지하지 못하게 하는 특성(identity)이 있다. 따라서, 이러한 문제를 해결하기 위해, 전자 장치(1000)는 하기 수학식 9에 의하여 정의되는 스케일 팩터 r 에 기초하여, add operation의 입력을 스케일링할 수 있다. 스케일링 팩터 r 은 blk1/shortcut 컨벌루션 레이어(1506)의 최대 절대 값 및 마지막 컨벌루션 레이어 blk1/conv-b(1512)의 최대 절대 값에 기초하여 결정될 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 로그 함수 및 floor 함수를 사용하기 때문에, 스케일링 팩터 r 은 항상 0.5에서 1의 범위를 가질 수 있고, 결과적으로 전자 장치(1000)는 ResNet의 맵핑 동작을 효과적으로 할 수 있다. 만약 스케일링 팩터 r 이 0인 경우, add operation에 추가될 것이 없고, 스케일링 팩터 r 이 1 보다 큰 경우, shortcut은 큰 레지듀얼에 의해 섭동(perturbed)될 수 있다.

수학식 9

$$r = f(m_{short}, m_{conv_b}) = \frac{m_{short}}{m_{conv_b}} \cdot 2^{\left\lfloor \log_2 \frac{m_{conv_b}}{m_{short}} \right\rfloor}, \quad (0.5 < r \leq 1)$$

여기에서 r 은 스케일링 팩터이고, m_{short} 및 m_{conv_b} 는 blk1/shortcut 컨벌루션 레이어(1506)의 최대 절대 값 및 마지막 컨벌루션 레이어 blk1/conv-b(1512)의 최대 절대 값을 각각 나타낼 수 있다.

수학식 10

$$add(o_q^j, o_q^k) = m_{short} \cdot (2^{\left\lfloor \log_2 \frac{I_a^k I_w^k}{I_a^j I_w^j} \right\rfloor} \cdot \sum \sum i_a^j i_w^j + \sum \sum i_a^k i_w^k)$$

상기 수학식 10을 이용하여, 전자 장치(1000)는 스케일링 팩터 r 을 조정함으로써, j 번째 shortcut 및 k 번째 컨벌루션 레이어와의 간격을 조정할 수 있고, 시프트 연산(shift operation)(2의 곱)만으로도 두가지 입력을 추가할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 인공 신경망 내 첫번째 컨벌루션 레이어를 양자화 하지 않을 수 있다. 인공 신경망 내 첫번째 컨벌루션 레이어가 양자화 되지 않을 경우, add operation 이후, ResNet의 정밀 출력이 mixed 되는 것을 방지하기 위해, 첫째 레지듀얼 블록은 블록의 입력 모양이 첫번째 레이어의 출력 모양과 동일하더라도 skip connection을 가지지 않을 수 있다.

도 16은 일 실시 예에 따른 전자 장치 내 인공 신경망의 동작을 설명하기 위한 도면이다.

도 16을 참조하면, 인공 신경망의 추론 시간(inference time)동안, 혼합 가중치 표현(hybrid weight representation, HWR)을 처리하기 위해 전자 장치가 이용하는 논리 회로가 도시된다. 도 16을 참조하면, 전자 장치가 이용하는 논리 회로는 크게, 이전 레이어의 입력 수신 노드(1612), 가중치(1614)중, 삼항 가중치(TW)로 식별된 가중치를 수신하기 위한 블록(1616), 가중치(1614)중, 큰 가중치(SLW)로 식별된 가중치를 수신하기 위한

블록(1618), 부호를 변환하기 위한 Sign asgmt 블록(1622) 및 입력과 가중치를 곱하기 위한 MUTT 블록(1624), Sign asgmt 블록(1622) 및 MUTT 블록(1624)의 출력을 선택하기 위한 스위치 블록(1626) 및 입력과 가중치의 곱합을 누적합 하기 위한 시그마 블록(1628)을 포함할 수 있다.

[0137] 일 실시 예에 의하면, 전자 장치는 삼항 가중치(TW)를 사용할 ??, 이진 가중치를 활용하기 위해, XNPR 게이트를 사용할 수 있다. 인공 신경망의 입력 값 또는 가중치가 0인 경우, 곱셈 연산은 생략되고, 0이 아닌 경우, XNOR 연산이 적용될 수 있다. 도 16을 참조하면, 삼항 가중치가 012 또는 112이고, 입력 값이 0이 아닌 경우, 부호를 변환하기 위한 Sign asgmt 블록(1622)에 의해, 오직 부호 변환 연산만이 수행될 수 있다. 그러나, 삼항 가중치가 102인 경우, MUTT 블록(1624)에 의해 오직 integer 곱만이 수행될 수 있다. 부호 변환 또는 곱셈의 결과 값은 하나의 뉴런의 출력 값이 될 수 있다.

[0138] 도 17은 또 다른 실시 예에 따른 가중치의 분포를 조절하기 위한 가중치 정규화 함수를 나타내는 도면이다.

[0139] 일 실시 예에 의하면, 전자 장치(1000)는 도 5에서 설명한 바와 같이, L2 제약 조건을 따르는 정규화 함수 및 부분 L1 제약 조건(pL1)을 따르는 정규화 함수를 더 함으로써 결정되는 제1 가중치 정규화 함수를 이용하여 가중치를 정규화할 수 있다. 그러나 또 다른 실시 예에 의하면, 전자 장치(1000)는 제1 가중치 정규화 함수와 다른 제3 가중치 정규화 함수를 이용하여 가중치를 정규화할 수도 있다. 예를 들어, 도 17의 차트 (1702)를 참조하면, L2 제약 조건을 따르는 정규화 함수의 손실 곡선(1706)과 부분 지수 lasso (part of exponential lasso, PeL1) 제약 조건을 따르는 정규화 함수의 손실 곡선(1708)이 도시되어 있다. 차트 1702의 점선은 제1 타입의 가중치 및 제2 타입의 가중치를 구별하기 위한 임계치(thresh)를 나타낼 수 있다.

[0140] 또한, 도 17의 차트 1712를 참조하면, 전자 장치(1000)가 이용하는 제3 가중치 정규화 함수의 손실 곡이 도시된다. 제3 가중치 정규화 함수는 L2 제약 조건을 따르는 정규화 함수 및 부분 지수 lasso (part of exponential lasso, PeL1) 제약 조건을 따르는 정규화 함수를 더 함으로써 정의될 수 있다. 도 17의 차트 1712를 참조하면 제3 가중치 정규화 함수는 L2 제약 조건을 따르는 정규화 함수 및 부분 지수 lasso (part of exponential lasso, PeL1) 제약 조건을 따르는 정규화 함수의 합으로 정의되므로, 임계치 부근에서 더 가파른 기울기를 가지는 손실 곡선을 나타낼 수 있다. 도 17의 차트 1722를 참조하면, 본 개시에 따른 제3 가중치 정규화 함수의 기울기가 도시된다.

[0141] 일 실시 예에 의하면, 전자 장치(1000)는 도 17의 차트 1722와 같은 기울기를 나타내는 제3 가중치 정규화 함수를 이용하여 가중치들을 정규화함으로써, 임계치(thresh) 이상의 가중치에 더 효과적으로 규제(penalty)를 수행할 수 있다. 본 개시에 따른 전자 장치(1000)는 가중치의 타입을 1차로 식별한 후, 식별된 가중치의 타입에 따라 서로 다른 양자화 함수를 결정하고, 서로 다른 양자화 함수를 가중치의 타입 별로 다르게 적용함으로써 인공 신경망의 성능을 더 향상시킬 수 있다.

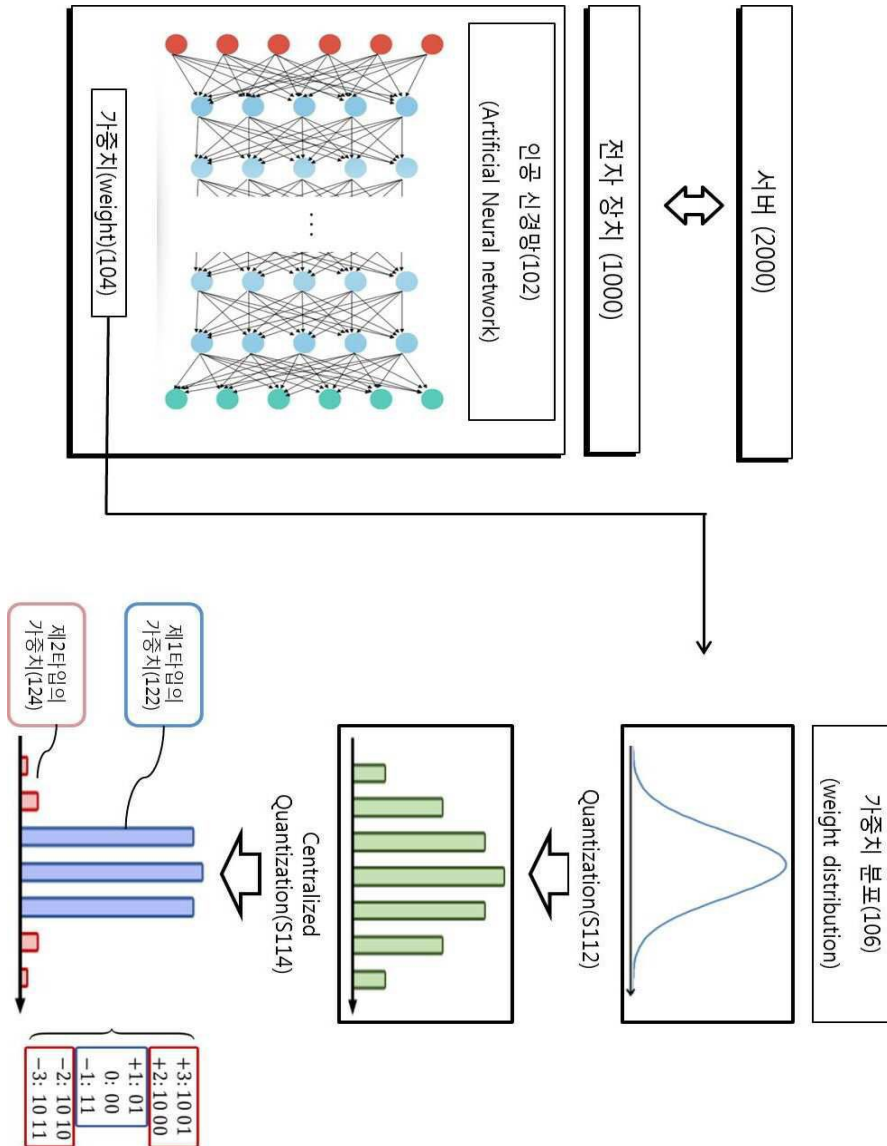
[0142] 일 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 개시를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0143] 또한, 상기 일 실시 예에 다른 방법을 수행하도록 하는 프로그램이 저장된 기록매체를 포함하는 컴퓨터 프로그램 장치가 제공될 수 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.

[0144] 이상에서 본 개시의 실시예에 대하여 상세하게 설명하였지만 본 개시의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 개시의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 개시의 권리범위에 속한다.

도면

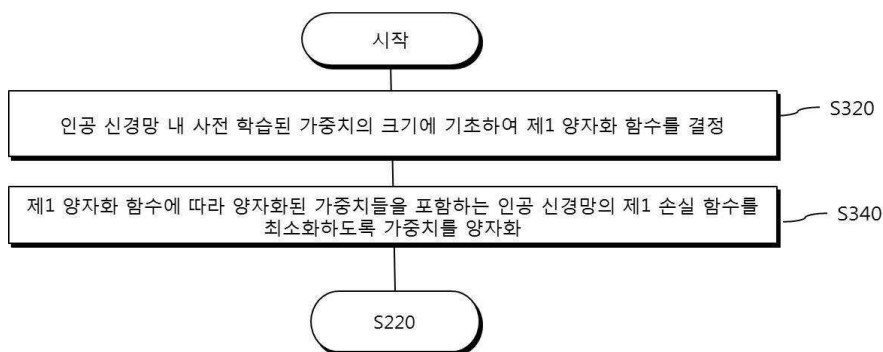
도면1



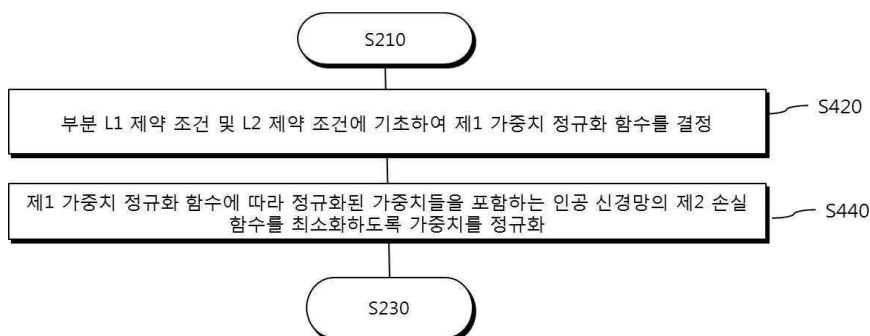
도면2



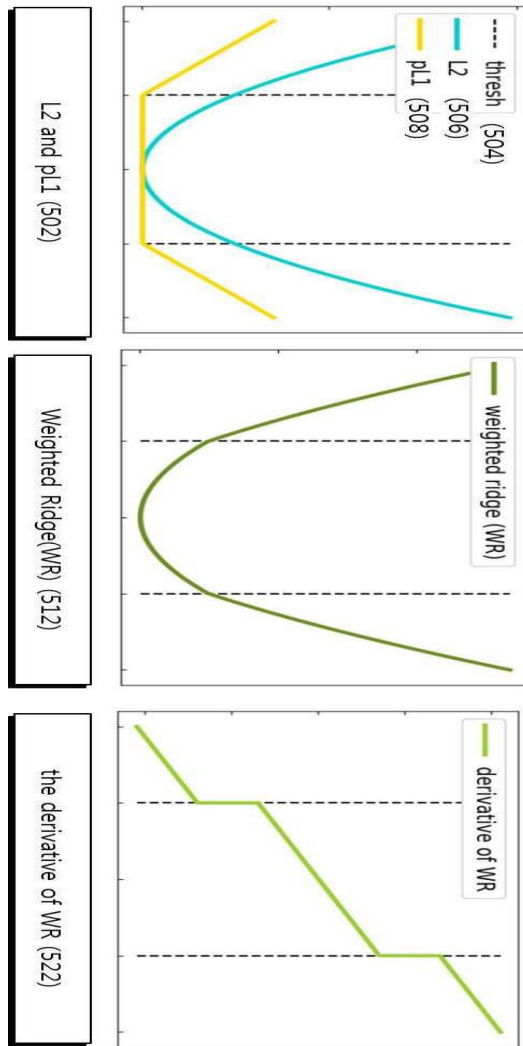
도면3



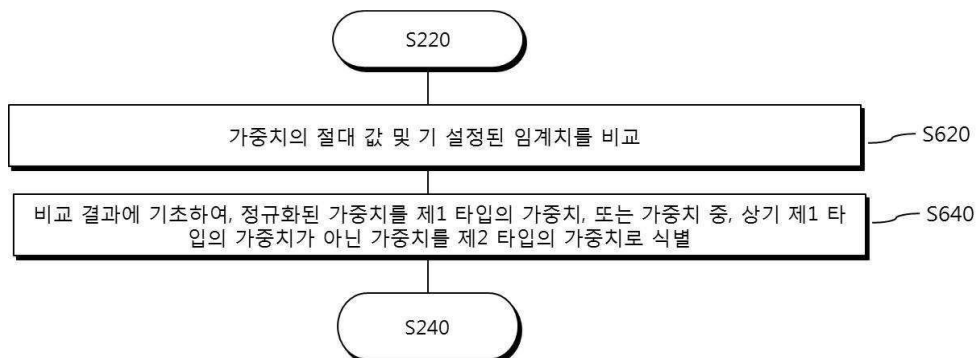
도면4



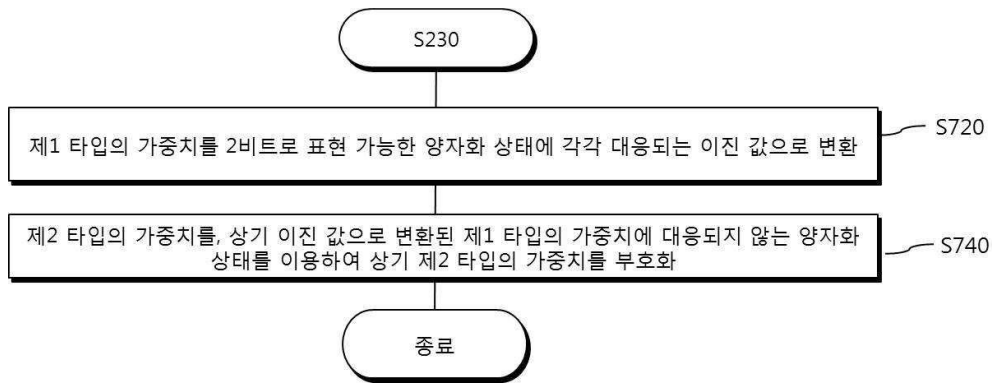
도면5



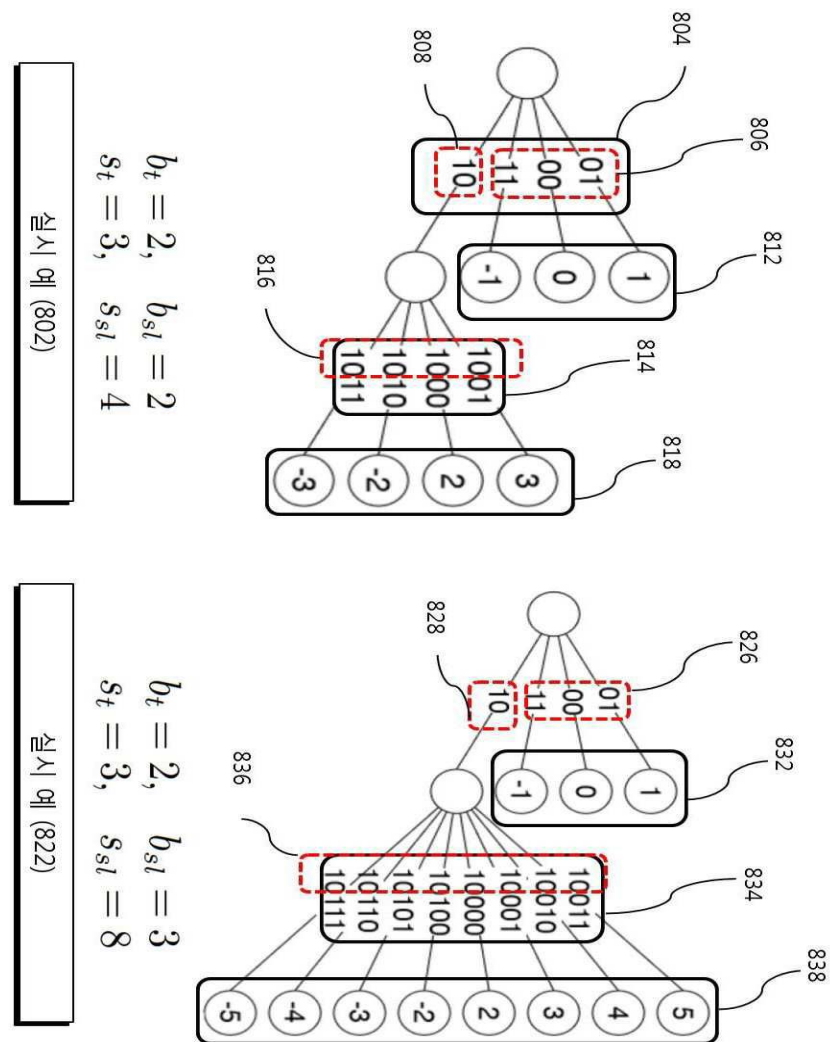
도면6



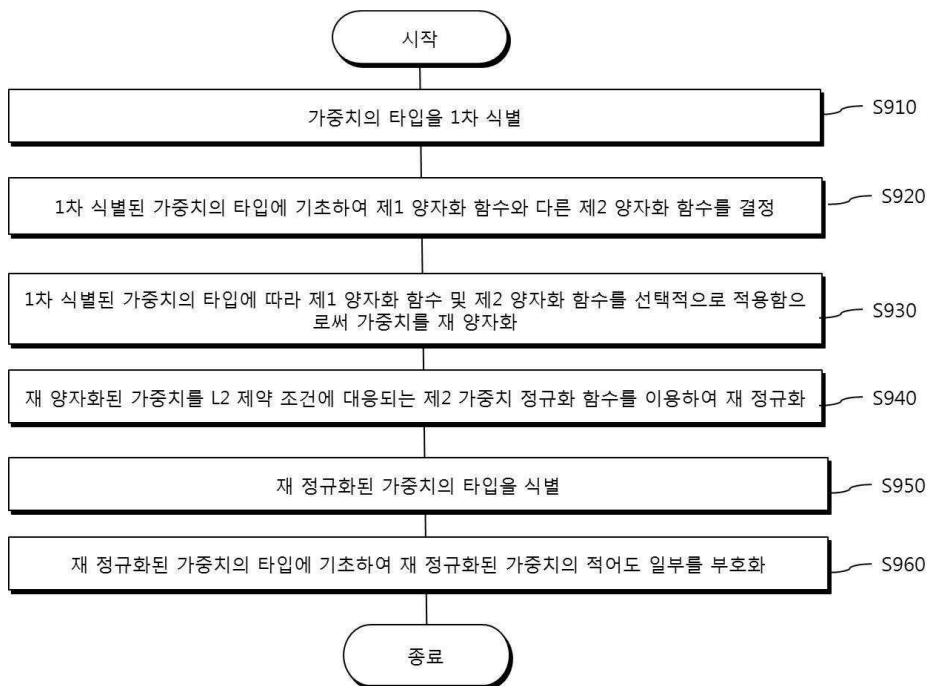
도면7



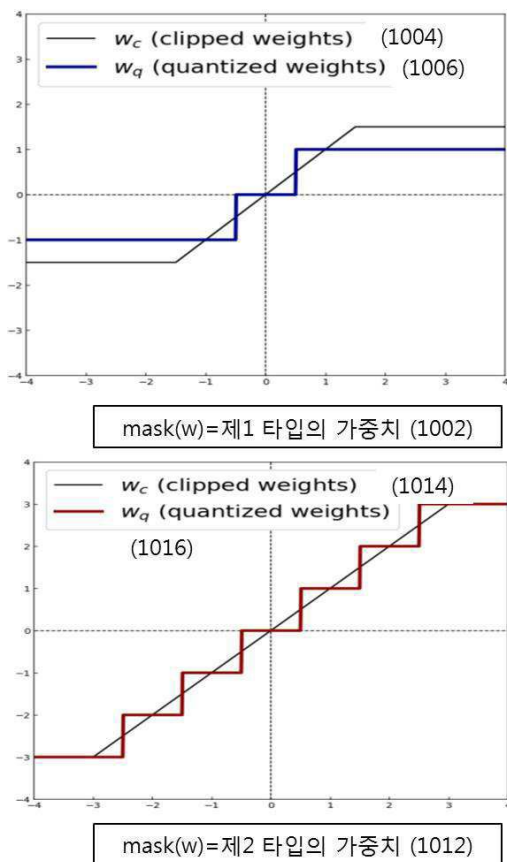
도면8



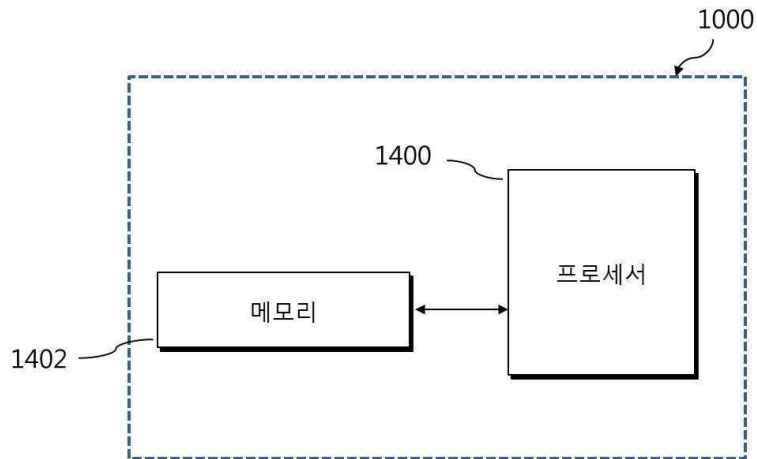
도면9



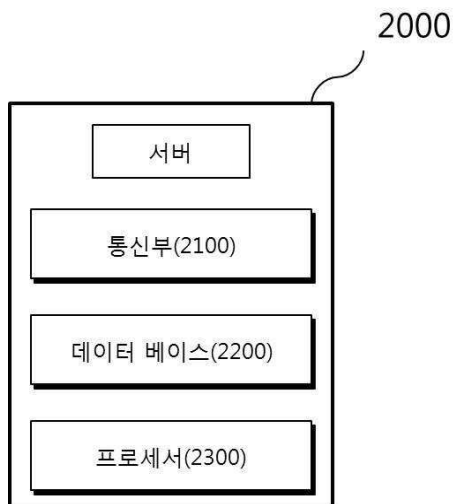
도면10



도면11



도면12



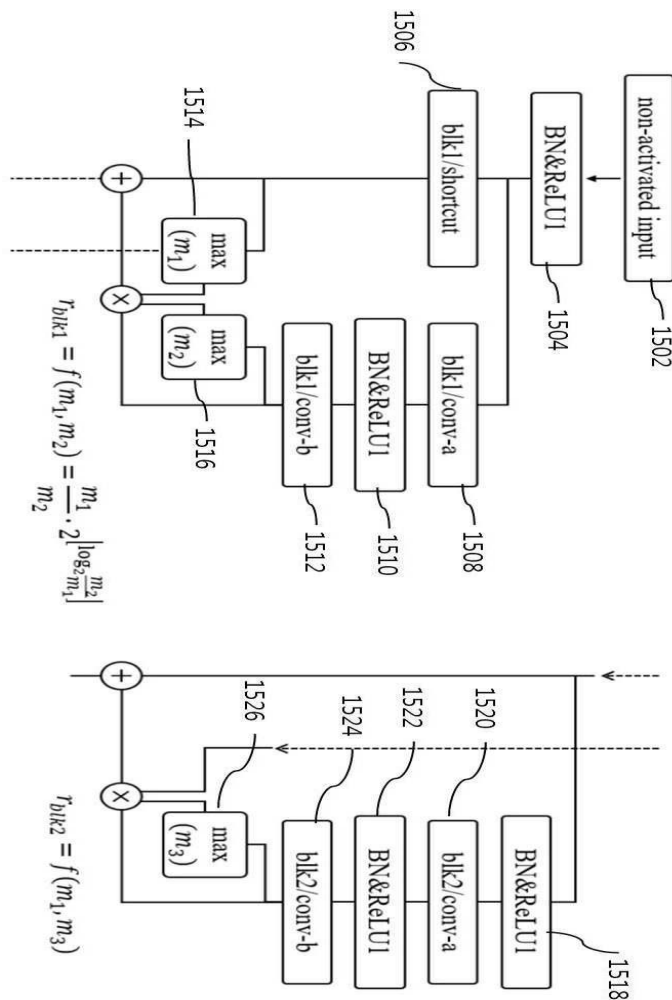
도면13

1322	1324	1326	1328	1330	1332	1334	1336
Model Name	Bit-W (b_t/b_{sl})	Bit-A	λ_2	Percentage of SLW	Average Bit Length	Top-1 Accuracy	Remarks
ResNet-18	32	32	-	-	-	70.41%	FP
ResNet-18	32	32	-	-	-	62.51%	FP(ReLU1)
ResNet-18	2	32	-	0%	2	66.6%	TTQ
ResNet-18	2	4	-	0%	2	52.59%	TTQ*
ResNet-18	2 / 3	4	-	1.11%	2.033	55.51%	BQ
ResNet-18	2 / 3	4	0.25	0.84%	2.025	55.20%	BQ+WR
ResNet-18	2 / 3	4	-	0.54%	2.016	57.67%	SQ
ResNet-18	2 / 4	4	-	4.012%	2.161	58.43%	BQ
ResNet-18	2 / 4	4	1	0.293%	2.012	56.70%	BQ+WR
ResNet-18	2 / 4	4	-	0.269%	2.011	59.00%	SQ

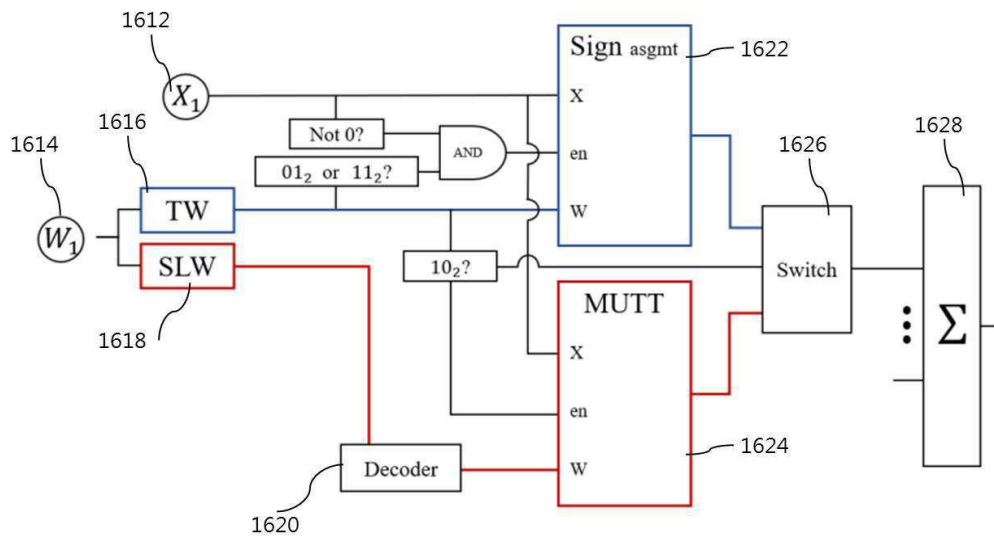
도면14

Model Name	Quantization Plan (b_t/b_{sl})	λ_2	Top-1 Accuracy	Remarks
AlexNet	-	-	75.06%	FP
AlexNet	2 / 3	1.2	75.32%	FP+WR
ResNet-18	-	-	76.07%	FP
ResNet-18	2 / 2	1	76.49%	FP+WR
ResNet-18	2 / 3	1	76.28%	FP+WR
ResNet-18	2 / 4	1	76.88%	FP+WR

도면15



도면16



도면17

