



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0089796
(43) 공개일자 2025년06월19일

(51) 국제특허분류(Int. Cl.)
G06N 3/0495 (2023.01) G06N 3/0464 (2023.01)
G06N 3/047 (2023.01) G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/0495 (2023.01)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0179475
(22) 출원일자 2023년12월12일
심사청구일자 2023년12월12일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
송병철
서울특별시 양천구 목동서로 38, 125동 307호(목동, 목동신시가지아파트1단지)
박지훈
인천광역시 미추홀구 인주대로224번길 17, 반석마을 303호(용현동)
(74) 대리인
양성보

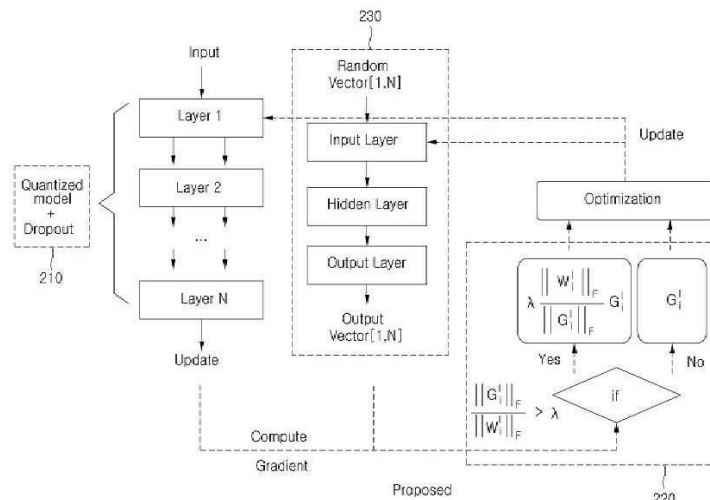
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법 및 시스템

(57) 요약

양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법 및 시스템이 제시된다. 본 발명에서 제안하는 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법은 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃(dropout)을 추가하여 분류기의 포워딩을 수행하는 단계, 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 곱합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계 및 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06N 3/047 (2023.01)

G06N 3/082 (2023.01)

G06N 3/084 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711160648
과제번호	2022-0-00955-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발사업
연구과제명	[Ezbaro] 고화질 CIS 및 DVS 기반 객체 및 행동인지가 가능한 인공지능 반도체 개발
(1차년도)	
기 여 율	33/100
과제수행기관명	(주) 아이닉스
연구기간	2022.04.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179347
과제번호	00155915
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신인재양성(교육운영수익)
연구과제명	[Ezbaro][정부+민간] 인공지능융합혁신인재양성(2차년도)
기 여 율	33/100
과제수행기관명	인하대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711186835
과제번호	2022R1A4A1033549
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	집단연구지원
연구과제명	[Ezbaro] 인공지능 융합형 메타휴먼 플랫폼 연구실
기 여 율	34/100
과제수행기관명	인하대학교
연구기간	2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 분류기의 포워딩을 수행하는 단계;

상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계; 및

상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계

를 포함하는 양자화 후 학습 방법.

청구항 2

제1항에 있어서,

상기 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 분류기의 포워딩을 수행하는 단계는,

편향되지 않은 출력 값을 얻기 위해 상기 드롭아웃을 추가하여 분류기의 포워딩을 수행함으로써 상기 양자화 후 학습 과정에서 일부 뉴런을 확률적으로 제거하는

양자화 후 학습 방법.

청구항 3

제1항에 있어서,

상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계는,

양자화로 인해 표현 가능한 범위가 감소되어 기울기가 가중치보다 큰 경우가 발생하여, 기울기의 방향을 나타내는 상기 기울기의 놈을 상기 가중치의 놈으로 나눈 값이 미리 정해진 임계값을 초과할 경우 해당 기울기의 크기를 제한하는

양자화 후 학습 방법.

청구항 4

제1항에 있어서,

상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계는,

상기 MLP의 입력은 레이어 개수만큼의 값을 가지는 랜덤값의 벡터이고, 입력 레이어, 히든레이어, 출력 레이어를 거쳐 레이어 개수만큼의 값을 가지는 출력 벡터를 가지며, 상기 출력된 벡터의 각각의 스칼라 값을 각 레이어에 할당하여 기울기 클리핑을 수행할 때 사용하고, 상기 MLP는 양자화 모델과 함께 순방향 전파(forward)와 역방향 전파(backward)를 진행하며 업데이트되는

양자화 후 학습 방법.

청구항 5

양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 포워딩을 수행하는 분류기;

상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 곱합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 기울기 클리핑부; 및

상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 MLP 신경망

을 포함하는 양자화 후 학습 시스템.

청구항 6

제5항에 있어서,

상기 분류기는,

편향되지 않은 출력 값을 얻기 위해 상기 드롭아웃을 추가하여 분류기의 포워딩을 수행함으로써 상기 양자화 후 학습 과정에서 일부 뉴런을 확률적으로 제거하는

양자화 후 학습 시스템.

청구항 7

제5항에 있어서,

상기 기울기 클리핑부는,

양자화로 인해 표현 가능한 범위가 감소되어 기울기가 가중치보다 큰 경우가 발생하여, 기울기의 방향을 나타내는 상기 기울기의 놈을 상기 가중치의 놈으로 나눈 값이 미리 정해진 임계값을 초과할 경우 해당 기울기의 크기를 제한하는

양자화 후 학습 시스템.

청구항 8

제5항에 있어서,

상기 MLP 신경망은,

상기 MLP의 입력은 레이어 개수만큼의 값을 가지는 랜덤값의 벡터이고, 입력 레이어, 히든레이어, 출력 레이어를 거쳐 레이어 개수만큼의 값을 가지는 출력 벡터를 가지며, 상기 출력된 벡터의 각각의 스칼라 값을 각 레이어에 할당하여 기울기 클리핑을 수행할 때 사용하고, 상기 MLP는 양자화 모델과 함께 순방향 전파(forward)와 역방향 전파(backward)를 진행하며 업데이트되는

양자화 후 학습 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 최근 신경망의 정확도가 증가함에 따라 모델의 크기도 함께 커지고 있다. 이에 따라 가지치기(pruning), 지식증류(knowledge distillation), 양자화(quantization) 등 다양한 신경망 압축 기법들이 연구되고 있다. 이중 양자화란 모델의 가중치(weight)와 액티베이션(activation)의 비트폭(bit-width)을 32bit 플로팅 포인 트(floating point; fp)에서 그 이하 비트(fp16, int8, int4, ...)로 낮추어 추론 동안 모델의 메모리 사용량을 낮추고 속도를 빠르게 하는 기법이다. 그러나 비트폭을 낮추면 가중치와 액티베이션의 표현범위를 줄이게 되므로 정확도 저하 문제가 발생한다. 4bit 이하의 공격적인 양자화의 경우 문제는 더 심각하다.

[0003] 기존의 덤러닝 기반의 양자화 기법은 크게 PTQ(Post Training Quantization)와 QAT(Quantization-Aware Training) 두 가지로 구분하게 된다.

[0004] 양자화를 하게 되면(PTQ) 기존의 FP32 모델보다 표현하는 비트 수가 줄어들기에 정확도 감소가 발생하게 된다.

이를 해결하기 위해 추가 학습을 진행하여 정확도를 복구하는 방법인 소위 훈련 후 양자화(Quantization Aware Training; QAT)가 있다. QAT는 학습 중 가짜 양자화 노드(fake quantization node)를 추가하여 양자화가 어떻게 진행되는지 모사하는 방법이다. 양자화 후에 추가 학습을 진행하는 QAT를 사용하여 FP32모델에 가까운 정확도를 얻게 된다. 이렇게 추가 학습을 통해서 기존의 양자화를 이용하여 감소하였던 정확도를 최대한 복구하게 된다.

[0005] 하지만, 이 QAT에서 가장 높은 성능을 보였던 HAWQ기법은 학습 에포크(epoch)가 증가할수록 과적합되어 낮은 정확도에 수렴하게 된다. 추가적으로 QAT에서는 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 양자화 후 정확도 감소를 줄이게 되는데 이때 배치 정규화 항을 결합함으로써 여러 정규화 효과들을 사용할 수 없게 된다. 이런 양자화와 추가 학습 과정에서 발생하는 과적합과 배치 정규화 항에 대한 문제점을 해결하지 못하면, 양자화에 대한 정확도 감소를 복구하기가 힘들다.

[0006] 또한, 기존의 양자화 신경망은 레이어마다 비트폭(bit-width)에 대해 민감도가 다르다. 즉, 특정 레이어는 낮은 비트로 양자화를 하게 되면 성능감소가 심하다는 의미이다. 대부분의 양자화 신경망은 이러한 민감도를 반영하지 않고 있고, 반영하더라도 학습 시 많은 시간과 메모리가 소요된다.

선행기술문헌

비특허문헌

[0007] (비특허문헌 0001) [1] Amir Gholami et al. A Survey of Quantization Methods for Efficient Neural Network Inference, arXiv preprint arXiv: 2103.13630.2021

(비특허문헌 0002) [2] Zhewei Yao et al, HAWQV3: Dyadic Neural Network Quantization. In ICML, 2021.

(비특허문헌 0003) [3] Andrew Brock et al, High-Performance Large-Scale Image Recognition Without Normalization, arXiv preprint arXiv:2102. 0617.

발명의 내용

해결하려는 과제

[0008] 본 발명이 이루고자 하는 기술적 과제는 QAT 기법을 활용한 HAWQ의 학습 과정에서 오버피팅(overfitting)되는 문제를 해결하기 위해 AGC(Adaptive Gradient Clipping)과 드롭아웃(dropout)을 적용한 기울기 클리핑 방법 및 시스템을 제공하는데 있다. 또한, 배치 정규화 층의 효과를 대신하며 정확도 감소를 해결하기 위한 기울기 클리핑 방법과 드롭아웃을 적용하며 최적화를 위해 각 레이어에 적응적으로 기울기 클리핑 값을 할당하기 위한 MLP(Multi-Layer Perceptron)를 추가한 신경망을 제안한다.

과제의 해결 수단

[0009] 일 측면에 있어서, 본 발명에서 제안하는 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법은 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃(dropout)을 추가하여 분류기의 포워딩을 수행하는 단계, 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계 및 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계를 포함한다.

[0010] 상기 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 분류기의 포워딩을 수행하는 단계는 편향되지 않은 출력 값을 얻기 위해 상기 드롭아웃을 추가하여 분류기의 포워딩을 수행함으로써 상기 양자화 후 학습 과정에서 일부 뉴런을 확률적으로 제거한다.

[0011] 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계는 양자화로 인해 표현 가능한 범위가 감소되어 기울기가 가중치보다 큰 경우가 발생하여, 기울기의 방향을 나타내는 상기 기울기

의 높을 상기 가중치의 높으로 나눈 값이 미리 정해진 임계값을 초과할 경우 해당 기울기의 크기를 제한한다.

[0012] 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계는 상기 MLP의 입력은 레이어 개수만큼의 값을 가지는 랜덤값의 벡터이고, 입력 레이어, 히든레이어, 출력 레이어를 거쳐 레이어 개수만큼의 값을 가지는 출력 벡터를 가지며, 상기 출력된 벡터의 각각의 스칼라 값을 각 레이어에 할당하여 기울기 클리핑을 수행할 때 사용하고, 상기 MLP는 양자화 모델과 함께 순방향 전파(forward)와 역방향 전파(backward)를 진행하며 업데이트된다.

[0013] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 시스템은 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 포워딩을 수행하는 분류기, 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 높과 가중치의 높의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 기울기 클리핑부 및 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 MLP 신경망을 포함한다.

발명의 효과

[0014] 본 발명의 실시예들에 따르면 AGC(Adaptive Gradient Clipping)과 드롭아웃(dropout)을 적용한 기울기 클리핑 방법 및 시스템을 통해 QAT 기법을 활용한 HAWQ의 학습 과정에서 오버피팅(overfitting)되는 문제를 해결할 수 있다. 또한, 각 레이어에 적응적으로 기울기 클리핑 값을 할당하기 위한 MLP를 추가한 신경망을 제안하여 배치 정규화 층의 효과를 대신하며 정확도 감소를 해결하기 위한 기울기 클리핑 방법과 드롭아웃을 적용하여 각 레이어에 최적화된 값을 배치할 수 있다. 본 발명의 실시예들에 따르면 QAT에 AGC를 적용하고 또한 AGC를 레이어에 적응적으로 적용하기 위해 학습 동안 AGC의 파라미터들을 MLP를 이용하여 함께 학습함으로써 ImageNet 데이터셋을 활용한 분류(classification)문제에 우수한 성능을 보여준다.

도면의 간단한 설명

[0015] 도 1은 종래기술에 따른 양자화 후 학습 알고리즘을 설명하기 위한 도면이다.
 도 2는 본 발명의 일 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법을 설명하기 위한 도면이다.
 도 3은 본 발명의 일 실시예에 따른 드롭아웃을 설명하기 위한 도면이다.
 도 4는 본 발명의 일 실시예에 따른 기울기 클리핑 제어 신경망을 설명하기 위한 도면이다.
 도 5는 본 발명의 일 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 시스템의 구성을 나타내는 도면이다.
 도 6은 본 발명의 일 실시예에 따른 8비트에서의 AGC와 드롭아웃을 포함한 QAT의 검증 데이터셋의 정확도를 비교한 그래프이다.
 도 7은 본 발명의 일 실시예에 따른 8비트 양자화에서 AGC와 MLP를 추가한 모델 정확도를 비교한 그래프이다.
 도 8은 본 발명의 일 실시예에 따른 4비트 양자화에서의 MLP를 추가한 AGC와 기존 기법의 정확도를 비교한 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0016] 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0018] 도 1은 종래기술에 따른 양자화 후 학습 알고리즘을 설명하기 위한 도면이다.

[0019] 기존의 양자화 후 학습(Quantization-Aware Training; QAT) 기법에서는 양자화 한 모델에 추가 학습을 하게 된다. 이때 학습 중 가짜 양자화 노드(fake quantization node)를 추가하여 실제 양자화를 하는 것처럼 동작하게 만들어 양자화를 시뮬레이션하게 된다. 이렇게 추가 학습을 통해서 기존의 양자화를 이용하여 감소하였던 정확도를 최대한 복구하게 된다.

- [0020] 하지만, 이 QAT에서 가장 높은 성능을 보였던 HAWQ기법은 학습 에포크(epoch)가 증가할수록 과적합되어 낮은 정확도에 수렴하게 된다. 추가적으로 QAT에서는 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 양자화 후 정확도 감소를 줄이게 되는데 이때 배치 정규화 항을 결합함으로써 여러 정규화 효과들을 사용할 수 없게 된다. 이런 양자화와 추가 학습 과정에서 발생하는 과적합과 배치 정규화 항에 대한 문제점을 해결하지 못하면, 양자화에 대한 정확도 감소를 복구하기가 힘들다.
- [0021] 또한, 기존의 양자화 신경망은 레이어마다 비트폭(bit-width)에 대해 민감도가 다르다. 즉, 특정 레이어는 낮은 비트로 양자화를 하게 되면 성능감소가 심하다는 의미이다. 대부분의 양자화 신경망은 이러한 민감도를 반영하지 않고 있고, 반영하더라도 학습 시 많은 시간과 메모리가 소요된다.
- [0023] 도 2는 본 발명의 일 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법을 설명하기 위한 도면이다.
- [0024] 본 발명의 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 방법은 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃(dropout)을 추가하여 분류기의 포워딩을 수행하는 단계(210), 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한하는 단계(220) 및 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP(Multi-Layer Perceptron)를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화하는 단계(230)를 포함한다.
- [0025] 단계(210)에서, 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 분류기의 포워딩을 수행한다.
- [0026] 본 발명의 실시예에 따르면, 양자화 후 학습에 있어서 배치 정규화 층의 폴딩(folding)으로 정규화 효과를 상실했기에 드롭아웃으로 정규화 효과를 대체한다.
- [0027] 단계(220)에서, 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한한다.
- [0028] 본 발명의 실시예에 따르면, 안정적인 학습을 위한 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 넘어가면 기울기가 과하게 커진 것으로 판단하여 기울기의 크기를 제한하게 된다.
- [0029] 단계(230)에서, 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화한다.
- [0030] 본 발명의 실시예에 따르면, 기울기 클리핑을 통제하여 레이어에 적응적으로 기울기 클리핑을 수행하기 위해 랜덤한 벡터를 MLP에 넣어 레이어 개수만큼의 스칼라 값을 얻어낸 후, 상기 스칼라 값들을 레이어마다 할당하여 기울기 클리핑을 진행한다.
- [0031] 본 발명의 실시예에 따른 레이어에 적응적인 기울기 클리핑 구성을 갖는 제안 알고리즘은 양자화 과정에서 과한 기울기를 제한함으로써 레이어에 적응적이고 안정적인 학습을 하는 동시에 정규화 효과를 추가하여 성능의 상승을 기대할 수 있다.
- [0033] 도 3은 본 발명의 일 실시예에 따른 드롭아웃을 설명하기 위한 도면이다.
- [0034] 도 3(a)는 종래기술의 배치 정규화 폴딩에 관하여 설명하기 위한 도면이다.
- [0035] QAT에서 가장 높은 성능을 보였던 HAWQ기법은 학습 에포크(epoch)가 증가할수록 과적합되어 낮은 정확도에 수렴하게 된다. 추가적으로 QAT에서는 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 양자화 후 정확도 감소를 줄이게 되는데 이때 배치 정규화 항을 결합함으로써 여러 정규화 효과들을 사용할 수 없게 된다. 이런 양자화와 추가 학습 과정에서 발생하는 과적합과 배치 정규화 항에 대한 문제점을 해결하지 못하면, 양자화에 대한 정확도 감소를 복구하기가 힘들다.

[0036] 도 3(b)는 본 발명의 실시예에 따른 드롭아웃을 설명하기 위한 도면이다.

[0037] 본 발명의 실시예에 따르면, 배치 정규화 층이 가지고 있는 정규화 효과를 대체하기 위해 드롭아웃을 이용한다. 드롭아웃은 배치 정규화가 등장하기 전부터 좋은 정규화 효과를 보이는 기법이었다. 편향되지 않은 출력 값을 얻기 위해 일부 뉴런을 확률적으로 제거하는 기법이다. 본 발명에서는 QAT에서 추론처럼 동작하기 위해 배치 정규화 층이 가중치와 결합하여 하나의 가중치처럼 동작하도록 하는 폴딩(Folding)이 수행되기 때문에 정규화 효과를 제대로 사용할 수 없게 된다. 따라서 드롭아웃으로 정규화 효과를 대체한다.

[0039] 도 4는 본 발명의 일 실시예에 따른 기울기 클리핑 제어 신경망을 설명하기 위한 도면이다.

[0040] 본 발명의 실시예에 따른 기울기 클리핑은 특정 조건이 임계값을 넘어가게 되면 기울기 값을 제한하게 된다. 특정 조건이라 함은 기울기의 높을 가중치의 높으로 나눈 값이 일정 임계값을 초과하는 경우를 말한다. 기울기의 높을 기울기의 방향을 나타내며 가중치의 높과 임계값이 제한하는 기울기의 값을 결정하게 된다. 기본적으로 학습과정에서는 초기 학습 때에는 많은 것을 학습해야 하기에 기울기의 값이 큰 편이며 가중치의 값은 작은편이다. 반대로 학습이 어느 정도 완료되었을 시점에는 기울기는 작은 값이 되며 가중치의 값은 커지게 된다. 양자화의 경우 표현 가능한 범위가 줄어들기에 기울기가 가중치보다 큰 경우가 발생할 수 있고, 이때의 경우 과하게 학습했다고 판단하게 되며 기울기를 제한한다.

[0041] 기존 QAT는 배치 정규화(batch normalization) 레이어와 가중치를 폴드(fold)하여 양자화를 진행한다. 이는 배치 정규화 레이어의 여러 정규화 효과들을 제대로 활용하지 못하게 할 수 있다. 배치 정규화를 제거하여 경쟁력 있는 성능을 보여준 NF-ResNet은 AGC(Adaptive Gradient Clipping)와 드롭아웃(dropout)을 통해 여러 정규화 효과들을 대체하였다.

[0042] AGC의 수식은 다음과 같다:

$$G^l \rightarrow \begin{cases} \lambda \frac{\|W^l\|_F}{\|G^l\|_F} & \text{if } \frac{\|G^l\|_F}{\|W^l\|_F} > \lambda, \\ G^l & \text{otherwise.} \end{cases} \quad (1)$$

[0044] 여기서 $W^l \in R^{N \times M}$, $G^l \in R^{N \times M}$ 이며 각각 가중치와 기울기를 나타내고 l 은 레이어, $\|\dots\|$ 은 프로베니우스 놈(Frobenius norm)이다. 클리핑 문턱값 λ 는 하이퍼파라미터이다.

[0045] 본 발명의 실시예에 따르면, 베이스라인 QAT 기법 중 하나인 HAWQ 모델이 오버피팅되는 현상을 확인하였고, AGC를 적용하여 안정적인 성능을 얻었다. 또한 기존의 양자화 신경망은 레이어마다 민감도가 다르다. 따라서, 같은 하이퍼파라미터를 사용한 AGC는 레이어마다의 민감도를 포함하고 있지 않다.

[0046] 본 발명의 실시예에 따른 기울기 클리핑을 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑 임계값을 조절한다. MLP의 입력으로는 레이어 개수만큼의 값을 가지는 랜덤값의 벡터이며, 입력 레이어, 히든레이어, 출력 레이어를 거쳐 레이어 개수만큼의 값을 가지는 출력 벡터를 가지게 된다. 이 출력된 벡터의 각각의 스칼라 값을 각 레이어에 할당하여 기울기 클리핑을 수행할 때 사용한다. 이 MLP는 양자화 모델과 함께 순방향 전파(forward)와 역방향 전파(backward)를 진행하며 업데이트된다.

[0047] 본 발명의 실시예에 따르면, AGC를 적용한 모델의 각 레이어에 적응적인 기울기 클리핑을 위해 클리핑 문턱값 λ 를 학습하는 MLP를 추가한다. MLP의 입력은 임의의 값을 가지는 클리핑 벡터 $v^{(k)}$ 이고 이 입력 벡터를 MLP에 통과시켜 모델의 레이어 개수와 동일한 n개의 양수값을 가지는 벡터 $\Lambda^{(k)} = [\lambda_1^{(k)}, \lambda_2^{(k)}, \dots, \lambda_n^{(k)}]$ 를 출력한다.

$$\Lambda^{(k)} = M(v^{(k)_c}) \quad (2)$$

[0049] 여기서 λ 는 스칼라 값이며 k 는 현재 파라미터 상태를, M 은 MLP 모델을 나타낸다. 이 MLP를 통과시켜 얻은 양수값들을 각각 클리핑 문턱값 λ 로 하여 모델의 레이어 마다 매칭시키고 AGC를 적용함으로써 각 레이어의 민감

도를 반영한다. QAT 과정에서 레이어마다 다른 클리핑 문턱값 λ 를 가진 AGC로 모델 파라미터를 업데이트함(수식 (3))과 동시에 MLP의 입력인 클리핑 벡터도 함께 업데이트한다(수식 (4)).

$$\theta_n^{(k)} \leftarrow \theta_n^{(k-1)} - \lambda_n \nabla_{\theta_n^{(k-1)}} L_m(\theta^{(k-1)}) \quad (3)$$

$$v_c^{(k)} = - \nabla_{v_c^{(k-1)}} L_c(v_c^{(k-1)}) \quad (4)$$

$$L = L_m + \alpha L_c \quad (5)$$

θ_n 은 모델의 n번째 레이어의 파라미터이다. 전체 손실함 수는 L_m 과 L_c 의 합으로 구성되며, L_m 과 L_c 는 각각 모델과 MLP의 손실함수이며 α 는 하이퍼파라미터이다.

도 5는 본 발명의 일 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 시스템의 구성을 나타내는 도면이다.

본 발명의 실시예에 따른 양자화 후 학습에서의 레이어에 적응적인 기울기 클리핑 시스템은 분류기(510), 기울기 클리핑부(520) 및 MLP 신경망(530)을 포함한다.

본 발명의 실시예에 따른 분류기(510)는 양자화 후 학습에 있어서 배치 정규화 층을 가중치와 함께 컨볼루션(convolution)하는 폴딩(Folding)을 수행함으로써 감소되는 정규화 효과를 개선하기 위해 드롭아웃을 추가하여 포워딩을 수행한다.

본 발명의 실시예에 따른 분류기(510)는 편향되지 않은 출력 값을 얻기 위해 상기 드롭아웃을 추가하여 분류기의 포워딩을 수행함으로써 상기 양자화 후 학습 과정에서 일부 뉴런을 확률적으로 제거한다.

본 발명의 실시예에 따른 기울기 클리핑부(520)는 상기 드롭아웃을 추가한 분류기의 포워딩을 수행하고 난 후, 기울기 클리핑을 수행하기 위해 기울기의 놈과 가중치의 놈의 결합이 미리 정해진 임계값을 초과할 경우, 해당 기울기의 크기를 제한한다.

본 발명의 실시예에 따른 기울기 클리핑부(520)는 양자화로 인해 표현 가능한 범위가 감소되어 기울기가 가중치보다 큰 경우가 발생하여, 기울기의 방향을 나타내는 상기 기울기의 놈을 상기 가중치의 놈으로 나눈 값이 미리 정해진 임계값을 초과할 경우 해당 기울기의 크기를 제한한다.

본 발명의 실시예에 따른 MLP 신경망(530)은 상기 기울기 클리핑을 분류기의 레이어마다 적응적으로 적용하기 위해 MLP를 추가하여 기울기 클리핑을 위한 임계값을 조절하여 상기 분류기의 파라미터를 최적화한다.

본 발명의 실시예에 따른 MLP 신경망(530)은 상기 MLP의 입력은 레이어 개수만큼의 값을 가지는 랜덤값의 벡터이고, 입력 레이어, 히든레이어, 출력 레이어를 거쳐 레이어 개수만큼의 값을 가지는 출력 벡터를 가지며, 상기 출력된 벡터의 각각의 스칼라 값을 각 레이어에 할당하여 기울기 클리핑을 수행할 때 사용하고, 상기 MLP는 양자화 모델과 함께 순방향 전파(forward)와 역방향 전파(backward)를 진행하며 업데이트된다.

도 6은 본 발명의 일 실시예에 따른 8비트에서의 AGC와 드롭아웃을 포함한 QAT의 검증 데이터셋의 정확도를 비교한 그래프이다.

종래기술에서 오버피팅의 발생 원인으로 폴드된 배치 정규화 레이어가 제 역할을 하지 못했기 때문이라고 판단했고, 본 발명의 실시예에 따른 AGC를 적용하여 모델을 안정시켰다. ImageNet 데이터셋을 사용하여 ResNet18 모델에서 8비트 양자화 실험을 진행하였다.

도 6과 같이, AGC를 적용하였을 때 에포크가 증가하여도 오버피팅되지 않고 71.47%의 안정적인 정확도를 가지며, 69.88%까지 내려가는 종래기술과는 큰 차이를 보여준다. NF-ResNet에서는 AGC 기법과 함께 드롭아웃을 적용하여 배치 정규화 레이어의 정규화 효과를 대체하였기에 본 발명의 실시예에 따른 실험에서도 드롭아웃을 추가하여 성능을 비교하였다. 실험결과 도 6과 표 1에서 볼 수 있듯이 AGC만 적용했을 때보다 약 간 더 높은

71.55%의 정확도를 보인다.

<표 1>

	AGC	Dropout	MLP	Accuracy(%)
QAT	x	x	x	69.81
QAT_Dropout	x	x	x	69.87
AGC	o	x	x	71.49
AGC_Dropout	o	o	x	71.55
MLP	o	x	o	71.57
MLP_Dropout	o	o	o	71.57

도 7은 본 발명의 일 실시예에 따른 8비트 양자화에서 AGC와 MLP를 추가한 모델 정확도를 비교한 그래프이다.

본 발명의 실시예에 따르면, AGC를 적용한 모델의 클리핑 문턱값 λ 를 각 레이어에 적응적인 그레디언트 클리핑을 위해 λ 를 학습하는 MLP를 추가한다. 먼저 MLP의 입력인 무작위의 값을 가지는 클리핑 벡터의 변화를 관찰하였다. 벡터들이 0.0145의 값을 가지며 오버피팅된다. 이를 해결하기 본 발명의 실시예에서는 클리핑 벡터의 초기값을 0.0145에 가깝게 설정한 후 실험을 진행했다. 도 7의 실험 결과와 표 1의 각 기법을 추가한 8비트 양자화에 대한 정확도 비교에서 볼 수 있듯이 클리핑 벡터를 적용한 모델의 정확도가 71.57%로 AGC만 사용한 모델보다 우수하다. 따라서 제안 기법이 전반적으로 높은 정확도를 보임을 확인할 수 있다.

도 8은 본 발명의 일 실시예에 따른 4비트 양자화에서의 MLP를 추가한 AGC와 기존 기법의 정확도를 비교한 그래프이다.

본 발명의 실시예에 따르면, MLP를 추가한 AGC가 다른 비트에서도 잘 작동함을 확인하기 위하여 4비트 양자화에서의 경향성을 관찰하였다. ImageNet의 1% 데이터셋을 사용하여 ResNet18 모델에서 4비트 양자화를 진행하였다. 도 8에서 관찰 할 수 있듯이 기존 기법이 오버피팅이 되며 61.23%의 정확도를 가지는 반면, MLP와 AGC를 적용한 모델이 안정적인 정확도 향상을 보여주며 65.16%의 기존 기법 대비 높은 정확도를 가짐을 보여주었다

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

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 네트워크

로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

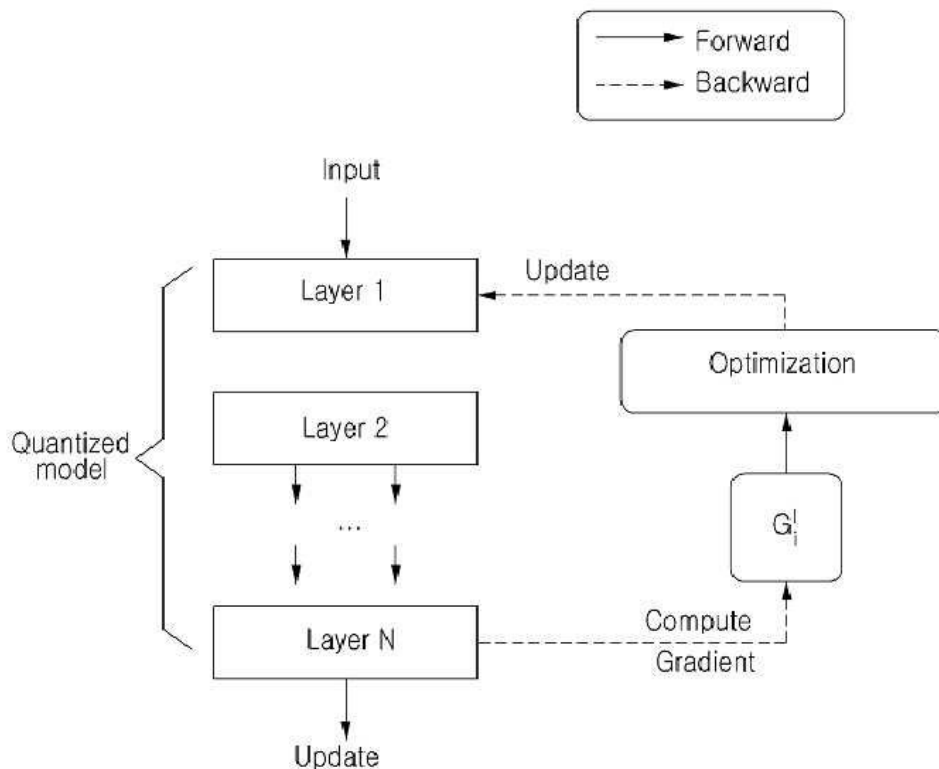
[0078] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic 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.

[0079] 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.

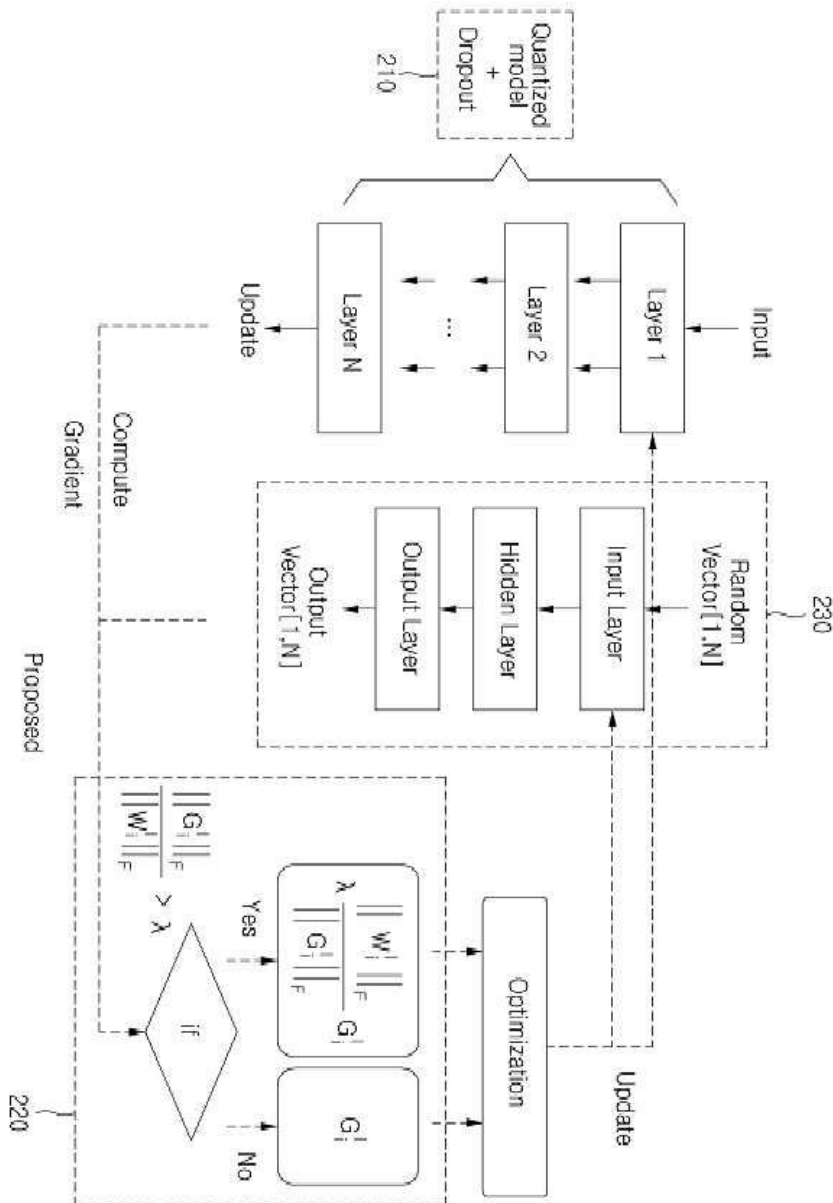
[0080] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

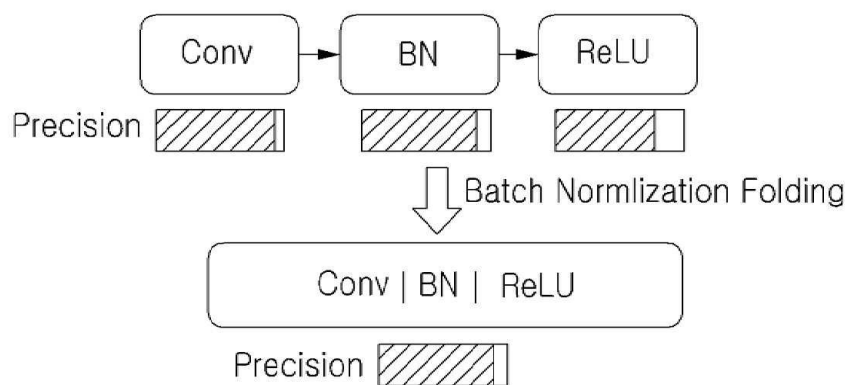
도면1



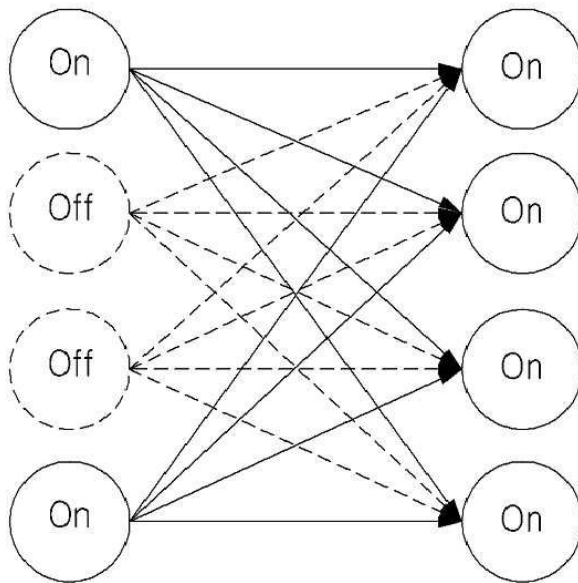
도면2



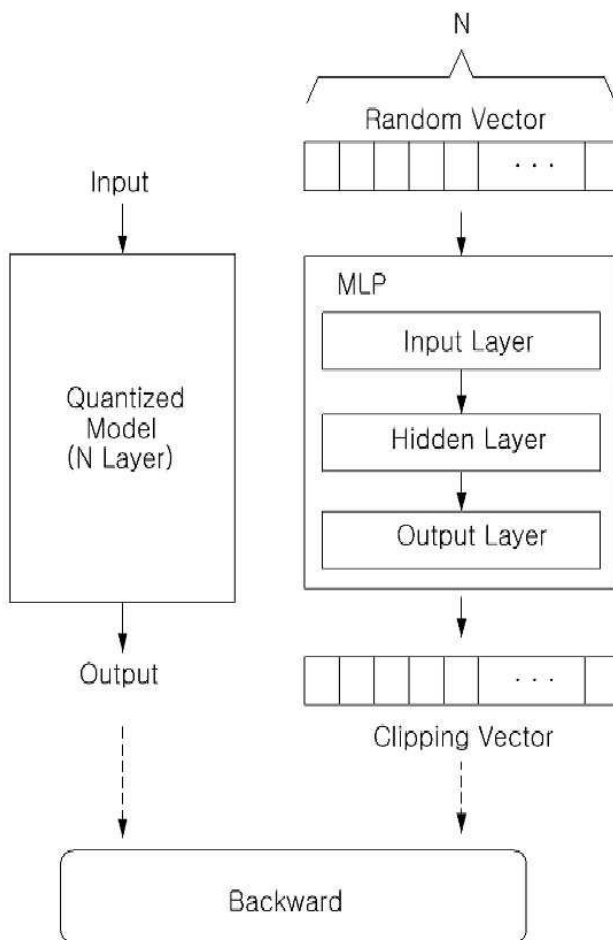
도면3a



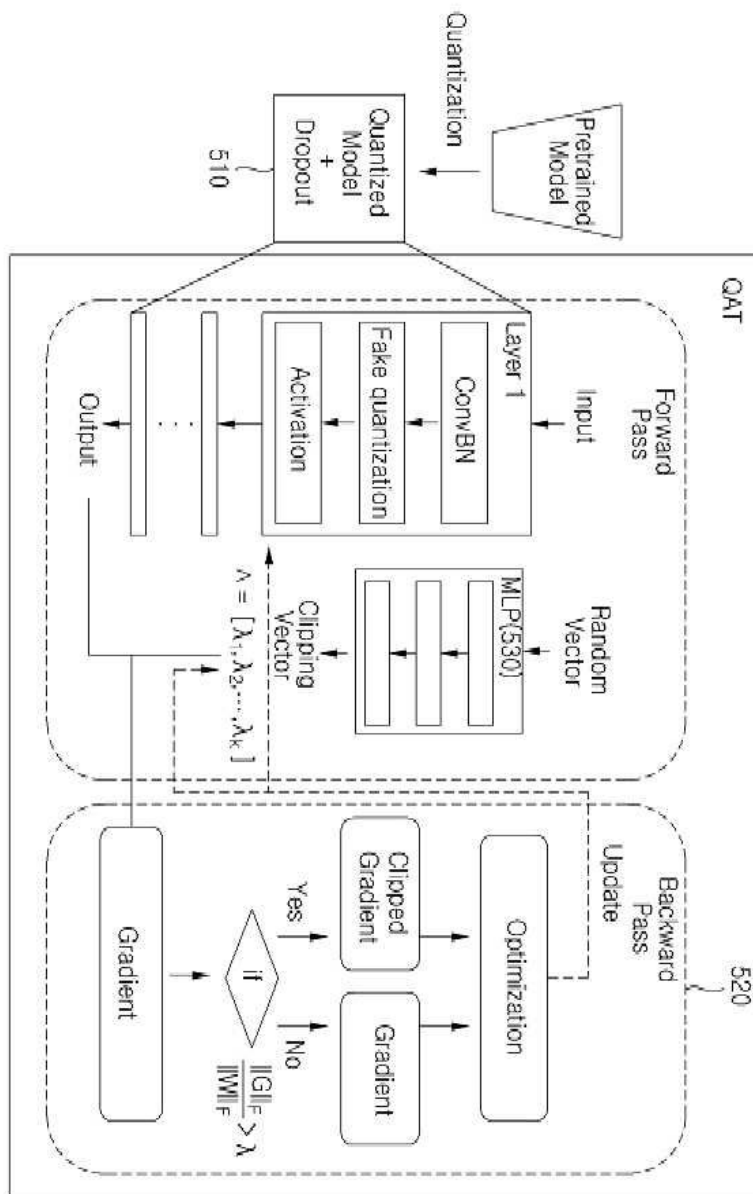
도면3b



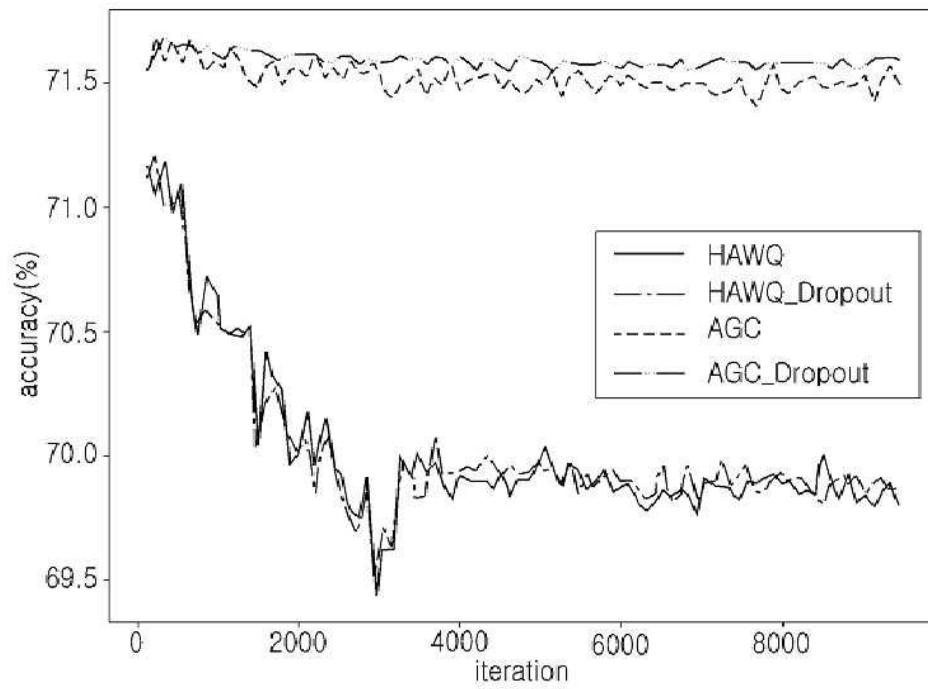
도면4



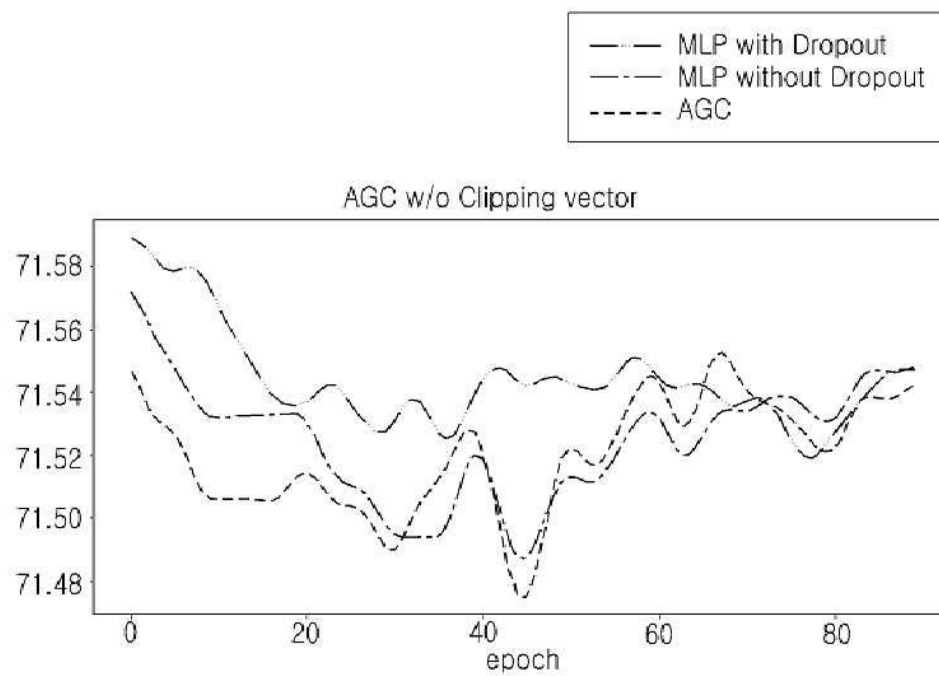
도면5



도면6



도면7



도면8

