(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0092286
(43) 공개일자 2023년06월26일

(51) 국제특허분류(Int. Cl.)
G06N 3/04 (2023.01) *G06N 3/063* (2023.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/049 (2023.01)
G06N 3/063 (2013.01)
(21) 출원번호 10-2021-0181554
(22) 출원일자 2021년12월17일
심사청구일자 2021년12월17일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
서영교
경기도 부천시 길주로77번길 19-45, 920호(상동, 디아뜨갤러리4)
최수름
인천광역시 연수구 선학로 100, 5동 1204호(선학동, 금호아파트)
이연진
인천광역시 미추홀구 매소홀로 171, 105동 2102호(용현동, 신창미션힐)
(74) 대리인
양성보

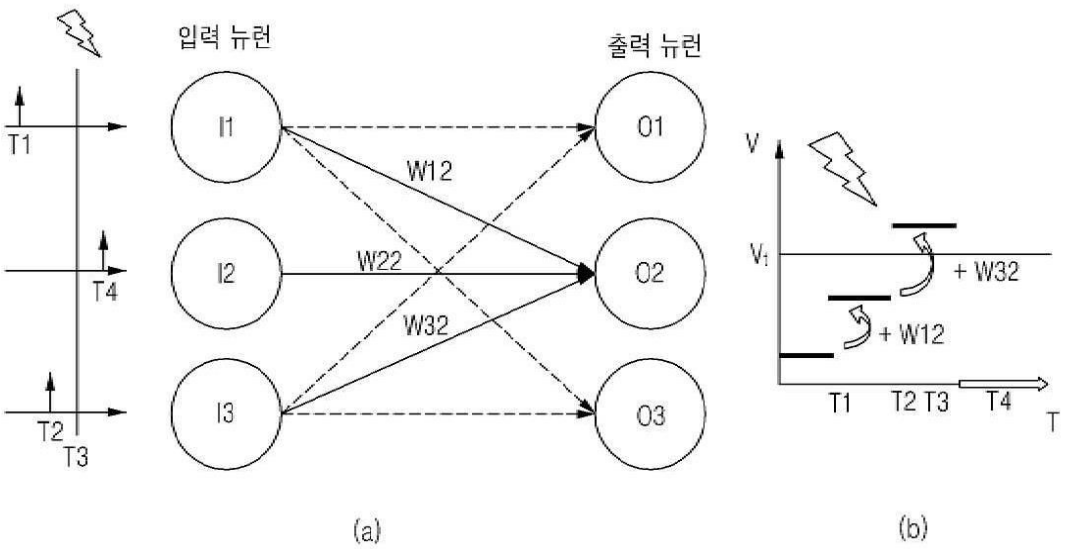
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **효율적이고 정확한 이진화된 심층 스파이킹 신경망**

(57) 요 약

효율적이고 정확한 이진화된 심층 스파이킹 신경망 및 그 동작 방법이 제시된다. 본 발명에서 제안하는 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 동작 방법은 입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF(Integrate and Fire) 모델에 기초한 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하는 단계, 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런 회로부를 통해 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위한 누설 효과 팀(leaky effect term)을 추가하는 단계 및 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G06N 3/084 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711139247
과제번호 2021-0-02052-001
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 대학 ICT연구센터 육성·지원사업
연구과제명 [Ezbaro][정부] 스마트 모빌리티를 위한 인공지능 시스템반도체 핵심 기술 개발 및 인력 양성 (1차년도)
기 여 율 1/1
과제수행기관명 인하대학교 산학협력단
연구기간 2021.07.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF(Integrate and Fire) 모델에 기초한 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하는 단계;

역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런 회로부를 통해 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위한 누설 효과 텀(leaky effect term)을 추가하는 단계; 및

학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계

를 포함하는 심층 스파이킹 신경망의 동작 방법.

청구항 2

제1항에 있어서,

상기 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계는,

배치 정규화 계층을 이용하지 않은 스파이킹 신경망의 역전파를 이용한 학습을 수행하고,

가중치 및 입력과 출력의 이진화를 적용함으로써 메모리 크기와 사용량을 감소시키고 연산 하드웨어를 경량화하는

심층 스파이킹 신경망의 동작 방법.

청구항 3

제2항에 있어서,

가중치 및 입력과 출력을 +1과 -1로 제한하여 산술 연산을 경량화된 비트 간의 연산으로 대체하는

심층 스파이킹 신경망의 동작 방법.

청구항 4

제1항에 있어서,

상기 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계는,

이진화된 역전파 기반의 스파이킹 신견망을 통해 추론(inference) 과정의 실행 시간을 감소시키고, 딥러닝 모델의 정확성을 증가시키는

심층 스파이킹 신경망의 동작 방법.

청구항 5

입력 데이터가 인가되면, 인가된 입력 데이터를 미리 지정된 방식에 따라 인코딩하여 입력 스파이크로 변환하여 대응하는 입력 뉴런으로 인가하는 입력 뉴런부;

IF(Integrate and Fire) 모델에 기초하여 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하고, 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위해 누설 효과 텀(leaky effect term)을 추가하는 LIF 뉴런 회로부;

가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 학습부; 및

복수의 출력 뉴런 각각은 LIF 뉴런 회로부에서 출력되는 가중치의 누적 합에 기초하여 복수의 출력 스파이크를 생성하는 출력 뉴런부

를 포함하는 심층 스파이킹 신경망.

청구항 6

제5항에 있어서,

상기 학습부는,

배치 정규화 계층을 이용하지 않은 스파이킹 신경망의 역전파를 이용한 학습을 수행하고,

가중치 및 입력과 출력의 이진화를 적용함으로써 메모리 크기와 사용량을 감소시키고 연산 하드웨어를 경량화하는

심층 스파이킹 신경망.

청구항 7

제6항에 있어서,

상기 학습부는,

가중치 및 입력과 출력을 +1과 -1로 제한하여 산술 연산을 경량화된 비트 간의 연산으로 대체하는

심층 스파이킹 신경망.

청구항 8

제1항에 있어서,

상기 학습부는,

이진화된 역전파 기반의 스파이킹 신견망을 통해 추론(inference) 과정의 실행 시간을 감소시키고, 딥러닝 모델의 정확성을 증가시키는

심층 스파이킹 신경망.

발명의 설명

기 술 분 야

[0001] 본 발명은 효율적이고 정확한 이진화된 심층 스파이킹 신경망에 관한 것으로, 더욱 상세하게는 메모리 용량을 줄이면서도 높은 정확성을 가지는 배치 정규화 계층이 없는 이진화 스파이킹 신경망에 관한 것이다.

배 경 기 술

[0002] 지난 몇 년간, 인공지능은 사물 인식, 음성인식 등 다양한 인지 산업 분야에 비약적인 발전을 이루었다. 그러나, 증가하는 심층 신경망의 복잡성으로 인해 모바일이나 임베디드 디바이스의 적용이 어려워졌다[1].

[0003] 이러한 난제를 해결하기 위해, 심층 신경망을 효율적으로 만드는 방법들이 제안되었으며, 그 중에서 입력, 출력, 가중치 값들을 이진수로 처리하는 방법이 메모리 비용을 줄이고, 신경망을 경량화하는 효과적인 방법으로 많은 주목을 받고 있다 [2]. 기존 XNOR-Net 기반의 이진화된 신경망(Binary Neural Network)은 성능을 유지하기 위해 배치 정규화(Batch Norm) 계층을 사용한다[3, 4].

[0004] 하지만, 배치 정규화 계층은 입력 데이터를 평균과 표준편차에 의해 정규화를 하는 과정을 필요로 하며, 이것을 구현하기 위해서는 복잡하고 전력 소모가 큰 추가적인 연산 회로와 메모리가 필요하다. 만약, 배치 정규화 계층을 신경망에 삽입하지 않는다면, XNOR-Net 기반의 이진화된 합성곱 신경망의 정확도는 크게 감소하게 된다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적 과제는 이진화된 데이터를 사용하여 메모리 크기와 컨볼루션 연산의 복잡성을 줄일 수 있는 스파이킹 신경망(Spiking Neural Network) 기반의 배치 정규화 계층이 없는 이진화된 심층 스파이킹 신경망 및 그 동작 방법을 제공하는데 있다.

과제의 해결 수단

[0006] 일 측면에 있어서, 본 발명에서 제안하는 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 동작 방법은 입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF(Integrate and Fire) 모델에 기초한 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하는 단계, 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런 회로부를 통해 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위한 누설 효과 텀(leaky effect term)을 추가하는 단계 및 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계를 포함한다.

[0007] 상기 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계는 배치 정규화 계층을 이용하지 않은 스파이킹 신경망의 역전파를 이용한 학습을 수행하고, 가중치 및 입력과 출력의 이진화를 적용함으로써 메모리 크기와 사용량을 감소시키고 연산 하드웨어를 경량화한다.

[0008] 본 발명의 실시예에 따르면, 가중치 및 입력과 출력을 +1과 -1로 제한하여 산술 연산을 경량화된 비트 간의 연산으로 대체한다.

[0009] 상기 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계는 이진화된 역전파 기반의 스파이킹 신견망을 통해 추론(inference) 과정의 실행 시간을 감소시키고, 딥러닝 모델의 정확성을 증가시킨다.

[0010] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 효율적이고 정확한 이진화된 심층 스파이킹 신경망은 입력 데이터가 인가되면, 인가된 입력 데이터를 미리 지정된 방식에 따라 인코딩하여 입력 스파이크로 변환하여 대응하는 입력 뉴런으로 인가하는 입력 뉴런부, IF(Integrate and Fire) 모델에 기초하여 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하고, 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위해 누설 효과 텀(leaky effect term)을 추가하는 LIF 뉴런 회로부, 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 학습부 및 복수의 출력 뉴런 각각은 LIF 뉴런 회로부에서 출력되는 가중치의 누적 합에 기초하여 복수의 출력 스파이크를 생성하는 출력 뉴런부를 포함한다.

발명의 효과

[0011] 본 발명의 실시예들에 따르면 이진화된 데이터를 사용하여 메모리 크기와 컨볼루션 연산의 복잡성을 줄일 수 있는 스파이킹 신경망(Spiking Neural Network) 기반의 배치 정규화 계층이 없는 이진화된 심층 스파이킹 신경망을 통해 메모리 크기와 사용량을 줄이고 연산 하드웨어를 경량화할 수 있다. 또한, 이진화된 역전파 기반의 스파이킹 신견망을 통해서 추론 과정의 실행 시간을 줄일 수 있을 뿐만 아니라, 딥러닝 모델의 정확성 또한 증가시킬 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 스파이킹 신경망의 동작 개념을 설명하기 위한 도면이다.

도 2는 본 발명의 일 실시예에 따른 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 동작 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명의 일 실시예에 따른 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 구성을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 LIF 뉴런 회로부를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0013] 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0015] 도 1은 본 발명의 일 실시예에 따른 스파이킹 신경망의 동작 개념을 설명하기 위한 도면이다.

[0016] 도 1(a)는 스파이킹 신경망의 입력 뉴런과 출력 뉴런의 동작 개념을 설명하기 위한 도면이다.

[0017] 도 1(b)는 시간에 따라 가중치에 의해 변화하는 전압을 나타내는 도면이다.

[0018] 도 1(a) 도시된 바와 같이, 스파이킹 신경망(Spiking Neural Network: SNN)은 복수의 입력 뉴런(I1 ~ I3)와 복수의 출력 뉴런(O1 ~ O3) 및 복수의 입력 뉴런(I1 ~ I3)와 복수의 출력 뉴런(O1 ~ O3)을 서로 잇는 복수의 시냅스를 포함한다. 그리고 복수의 시냅스 각각은 학습에 의해 획득되는 가중치를 갖는다. 도 1에서는 일 예로 3개의 입력 뉴런(I1 ~ I3)과 하나의 출력 뉴런(O2) 사이를 잇는 3개의 시냅스가 갖는 가중치(W12, W22, W32)를 도시하였다.

[0019] 복수의 입력 뉴런(I1 ~ I3)에는 도 1(a)와 같이 서로 다른 타이밍(T1, T2, T4)에 입력 스파이크(Input spike) 신호가 인가된다. 입력 스파이크는 입력 데이터를 미리 지정된 방식으로 인코딩하여 스파이크 트레인(spike train)의 형태로 획득된 신호이다. 그리고 입력 스파이크가 인가된 입력 뉴런(I1 ~ I3)은 시냅스를 따라 복수의 출력 뉴런(O1 ~ O3)으로 스파이크를 전달한다. 이때, 각 시냅스의 가중치가 입력 스파이크에 가중되어 전달되고, 가중치가 가중된 입력 스파이크가 전달된 복수의 출력 뉴런(O1 ~ O3) 각각은 입력 스파이크가 전달된 타이밍과 시냅스에 할당된 가중치에 기반하여, 서로 다른 크기로 누적된 스파이크의 세기가 미리 지정된 발화 문턱값(Vt) 이상이면, 출력 스파이크(Output spike) 신호를 출력한다.

[0020] 도 1(b)를 참조하면, 제1 타이밍(T1)과 제2 타이밍(T2)에 각각 제1 및 제3 입력 뉴런(I1, I3)으로 인가된 입력 스파이크에 시냅스의 가중치(W12, W32)가 가중되어 제2 출력 뉴런(O2)으로 전달되고, 제2 출력 뉴런(O2)은 가중치(W12, W32)가 가중되고 누적된 입력 스파이크의 전압이 발화 문턱값(Vt) 이상이므로, 제3 타이밍(T3)에서 출력 스파이크를 출력한다.

[0021] 이와 같은 스파이킹 신경망에서는 주로 인간 신경망의 동작 특성을 모방한 LIF(Leaky-Integrate and Fire) 뉴런 모델이 이용 된다. LIF 뉴런 모델은 입력되는 입력 스파이크에 따른 가중치를 누적(Integrate)하고, 누적된 가중치가 미리 지정된 기준 레벨 이상이면, 발화(Fire)하여 다음 뉴런으로 전달할 새로운 입력 스파이크를 생성하며, 입력 스파이크가 전달되지 않으면 누적된 입력 스파이크의 세기가 누설(Leaky)되어 시간이 흐를수록 점차 약해지도록 구성된 모델이다.

[0023] 도 2는 본 발명의 일 실시예에 따른 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 동작 방법을 설명하기 위한 흐름도이다.

[0024] 제안하는 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 동작 방법은 입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF(Integrate and Fire) 모델에 기초한 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하는 단계(210), 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런 회로부를 통해 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위한 누설 효과 텀(leaky effect term)을 추가하는 단계(220) 및 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계(230)를 포함한다.

[0025] 단계(210)에서, 입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF 모델에 기초한 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행한다.

[0026] 복수의 입력 뉴런에 서로 다른 타이밍에 입력 스파이크(Input spike) 신호가 인가되고, 이러한 입력 스파이크는 입력 데이터를 미리 지정된 방식으로 인코딩하여 스파이크 트레인(spike train)의 형태로 획득된 신호이다. 그리고 입력 스파이크가 인가된 입력 뉴런은 시냅스를 따라 복수의 출력 뉴런으로 스파이크를 전달한다.

[0027] 단계(220)에서 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런 회로부를 통해 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위한 누설 효과 텀

(leaky effect term)을 추가한다.

[0028] 일반적으로 스파이킹 신경망은 불연속성으로 인해 역전파(Backpropagation)를 통한 신경망 학습이 불가능하다. 하지만, LIF 근사 미분 방정식을 사용 할 경우, 높은 정확성을 가지는 역전파 기법 기반의 신경망 학습이 가능하다[5].

[0029] 먼저, IF(Integrate and Fire) 모델을 바탕으로 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 될 수 있도록 식(1) 및 식(2)에서 보듯이 근사를 진행하였다.

[0030] $$a_{IF}(t) \approx \frac{1}{V_{th}} \sum_{i=1}^{n} (w_i x_i(t)) = \frac{1}{V_{th}} net(t) \quad \text{식(1)}$$

[0031] $$\frac{\partial a_{IF}}{\partial net} \approx \frac{1}{V_{th}} \cdot 1 = \frac{1}{V_{th}} \quad \text{식(2)}$$

[0032] 이 모델을 바탕으로 LIF(Leaky-Integrate and Fire)의 활성화 함수를 표현하기 위해 누설 효과 텀(leaky effect term)을 추가하여 LIF 모델에서의 역전파 기법을 제안한다.

[0033] 단계(230)에서, 학습부를 통해 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행한다.

[0034] 각 시냅스의 가중치가 입력 스파이크에 가중되어 전달되고, 가중치가 가중된 입력 스파이크가 전달된 복수의 출력 뉴런 각각은 입력 스파이크가 전달된 타이밍과 시냅스에 할당된 가중치에 기반하여, 서로 다른 크기로 누적된 스파이크의 세기가 미리 지정된 발화 문턱값 이상이면, 출력 스파이크(Output spike) 신호를 출력한다.

[0035] 본 발명의 실시예에 따르면, 배치 정규화 계층을 이용하지 않은 스파이킹 신경망의 역전파를 이용한 학습을 수행할 수 있다. 또한, 메모리 크기와 사용량을 줄이고 연산 하드웨어의 경량화를 위해서 가중치 및 입력과 출력의 이진화를 적용한다.

[0036] 본 발명의 실시예에 따르면, 가중치 및 입력과 출력을 +1과 -1로 제한하여 부동 소수점 기반의 기존 신경망과 달리 산술 연산을 경량화된 비트 간의 연산으로 대체하여 딥러닝 모델을 경량화할 수다.

[0037] 따라서 이진화된 역전파 기반의 스파이킹 신견망을 통해서 추론(inference) 과정의 실행 시간을 또한 매우 줄일 수 있을 뿐만 아니라, 딥러닝 모델의 정확성 또한 XNOR-Net보다 크며, 부동 소수점 기반의 기존 CNN에 근접한 매우 높은 정확성을 가질 수 있다.

[0039] 도 3은 본 발명의 일 실시예에 따른 효율적이고 정확한 이진화된 심층 스파이킹 신경망의 구성을 설명하기 위한 도면이다.

[0040] 제안하는 효율적이고 정확한 이진화된 심층 스파이킹 신경망(300)은 입력 뉴런부(310), LIF 뉴런 회로부(320), 출력 뉴런부(330) 및 학습부(340)를 포함한다.

[0041] 본 발명의 실시예에 따른 입력 뉴런부(310)는 입력 데이터가 인가되면, 인가된 입력 데이터를 미리 지정된 방식에 따라 인코딩하여 입력 스파이크로 변환하여 대응하는 입력 뉴런으로 인가한다.

[0042] 복수의 입력 뉴런에 서로 다른 타이밍에 입력 스파이크(Input spike) 신호가 인가되고, 이러한 입력 스파이크는 입력 데이터를 미리 지정된 방식으로 인코딩하여 스파이크 트레인(spike train)의 형태로 획득된 신호이다. 그리고 입력 스파이크가 인가된 입력 뉴런은 시냅스를 따라 복수의 출력 뉴런부(330)으로 스파이크를 전달한다.

[0043] 본 발명의 실시예에 따른 LIF 뉴런 회로부(320)는 입력 뉴런부로부터 입력 받은 입력 스파이크에 대하여 IF(Integrate and Fire) 모델에 기초하여 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 되도록 근사를 진행하고, 역전파(Backpropagation)를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF(Leaky-Integrate and Fire) 뉴런의 활성화 함수로 표현하기 위해 누설 효과 텀(leaky effect term)을 추가한다.

[0044] 일반적으로 스파이킹 신경망은 불연속성으로 인해 역전파(Backpropagation)를 통한 신경망 학습이 불가능하다. 하지만, LIF 근사 미분 방정식을 사용 할 경우, 높은 정확성을 가지는 역전파 기법 기반의 신경망 학습이 가능하다[5].

[0045] 먼저, IF(Integrate and Fire) 모델을 바탕으로 근사를 사용해 IF 뉴런의 활성화 함수가 선형 함수가 될 수 있도록 식(1) 및 식(2)에서 보듯이 근사를 진행한다.

[0046] 이 모델을 바탕으로 LIF(Leaky-Integrate and Fire)의 활성화 함수를 표현하기 위해 누설 효과 텀(leaky effect term)을 추가하여 LIF 모델에서의 역전파 기법을 제안한다.

[0048] 도 4는 본 발명의 일 실시예에 따른 LIF 뉴런 회로부를 설명하기 위한 도면이다.

[0049] 심층 스파이킹 신경망의 주요 구성 요소인 뉴런은 LIF(leaky integrate and fire) 뉴런 회로를 통하여 구현된다. 일반적인 LIF 뉴런은 축전기를 포함하며, 축전기를 통하여 뉴런에 들어오는 신호를 축적한다. 도 4는 본 발명의 일 실시예에 따른 LIF 뉴런 회로부를 설명하기 위한 도면이다.

[0050] 도 4를 참조하면, 본 발명의 실시예에 따른 LIF 뉴런 회로부는 시냅스로부터 들어온 신호들에 의해 멤브레인(membrane) 전위가 증가하고, 멤브레인 전위가 뉴런의 역치에 도달했을 때, 다음 심층 시냅스층(hidden layer)의 입력 또는 출력에 해당되는 스파이크를 발생시킨다. 스파이크가 발생된 직후엔 멤브레인 전위는 초기값으로 복귀하게 되고, 입력이 들어와도 멤브레인 전위가 증가하기 않는 휴지기(refractory period)가 발생된다. 이러한 휴지기가 지나고 다시 입력이 들어올 때 멤브레인 전위는 증가하게 된다. 즉, LIF 뉴런 회로는 신호가 들어오지 않는 동안 멤브레인 전위가 감소하는 누설(leaky) 특성을 갖는다.

[0052] 다시 도 3을 참조하면, 본 발명의 실시예에 따른 출력 뉴런부(330)는 복수의 출력 뉴런 각각은 LIF 뉴런 회로부에서 출력되는 가중치의 누적 합에 기초하여 복수의 출력 스파이크를 생성한다.

[0053] 본 발명의 실시예에 따른 학습부(340)는 가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행한다.

[0054] 각 시냅스의 가중치가 입력 스파이크에 가중되어 전달되고, 가중치가 가중된 입력 스파이크가 전달된 복수의 출력 뉴런 각각은 입력 스파이크가 전달된 타이밍과 시냅스에 할당된 가중치에 기반하여, 서로 다른 크기로 누적된 스파이크의 세기가 미리 지정된 발화 문턱값 이상이면, 출력 스파이크(Output spike) 신호를 출력한다.

[0055] 본 발명의 실시예에 따르면, 배치 정규화 계층을 이용하지 않은 스파이킹 신경망의 역전파를 이용한 학습을 수행할 수 있다. 또한, 메모리 크기와 사용량을 줄이고 연산 하드웨어의 경량화를 위해서 가중치 및 입력과 출력의 이진화를 적용한다.

[0056] 본 발명의 실시예에 따르면, 가중치 및 입력과 출력을 +1과 -1로 제한하여 부동 소수점 기반의 기존 신경망과 달리 산술 연산을 경량화된 비트 간의 연산으로 대체하여 딥러닝 모델을 경량화할 수다.

[0057] 따라서 이진화된 역전파 기반의 스파이킹 신견망을 통해서 추론(inference) 과정의 실행 시간을 또한 매우 줄일 수 있을 뿐만 아니라, 딥러닝 모델의 정확성 또한 XNOR-Net보다 크며, 부동 소수점 기반의 기존 CNN에 근접한 매우 높은 정확성을 가질 수 있다.

[0058] 표 1은 표준 CNN, XNOR-Net, BSNN에 대한 정확도를 나타낸다.

[0059] <표 1>

[0060]

Network	Accuracy (CIFAR-10)
STANDARD CNN (Without BATCHNORM)	91.8%
XNOR-NET (With BACHTNORM)	88.0%
XNOR-NET (Without BATCHNORM)	73.9%
THIS WORK (BSNN)	90.2%

[0061] 제안된 이진화된 심층 스파이킹 신경망은 VGG8 네트워크에서 100 time-step을 사용하여 훈련되었고 표준 이미지 데이터 셋인 CIFAR-10 데이터셋을 이 용하여 진행되었다. 표준 합성곱 신경망 (Standard CNN) 에서는 91.8%의

정확도를 보였다. 그리고 배치 정규화 계층이 없는 XNOR-Net 기반의 이진화된 합성곱 신경망에서는 73.9%의 분류 정확도로 표준 합성곱 신경망에서 크게 정확성이 감소하는 것을 확인하였다. 배치 정규화 계층이 있는 이진화된 합성곱 신경망은 상대적으로 높은 88.0%의 정확도를 보였지만, 여전히 표준 합성곱 신경망보다는 낮은 정확성을 가지고 있다. 하지만 본 발명에서 제시된 이진화된 스파이킹 신경망은 배치 정규화가 없음에도 불구하고, 배치 정규화 계층이 있는 XNOR-NET보다 높은 90.2%의 정확도를 보였으며, 부동 소수점을 사용하는 표준 합성곱 신경망과 비교하여 매우 근접한 정확성을 가지고 있다.

[0062] 이와 같이 본 발명의 실시예에 따른 배치 정규화 계층을 사용하지 않는 이진화 방식 스파이킹 신경망을 통해 기존의 이진화된 가중치와 입력 및 출력을 사용하며, LIF 뉴런의 활성화 함수를 선형 함수로 근사화하여 스파이킹 신경망의 역전파 학습이 가능해지도록 하였다. 배치 정규화 계층을 사용하지 않기 때문에 추가적인 연산 하드웨어나 메모리 사용이 필요하지 않으면서 스파이킹 신경망에 최적화된 역전파 학습기법을 통해 기존 이진화 합성곱 신경망보다 크게 향상된 정확성을 가질 수 있다.

[0064] 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, 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 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

[0070] <참고문헌>

[0071] [1] T. Simons and D. J. Lee, "A review of binarized neural networks", MDPI Electron., vol. 8, no. 6, pp. 661-686, Jun 2019.

[0072] [2] M.. Courbariaux and Y. Bengio, "Binarized neural networks: Training deepneural networks with weights and activations constrained to +1 or -1", arXiv preprint arXiv:1602.02830, 2016.

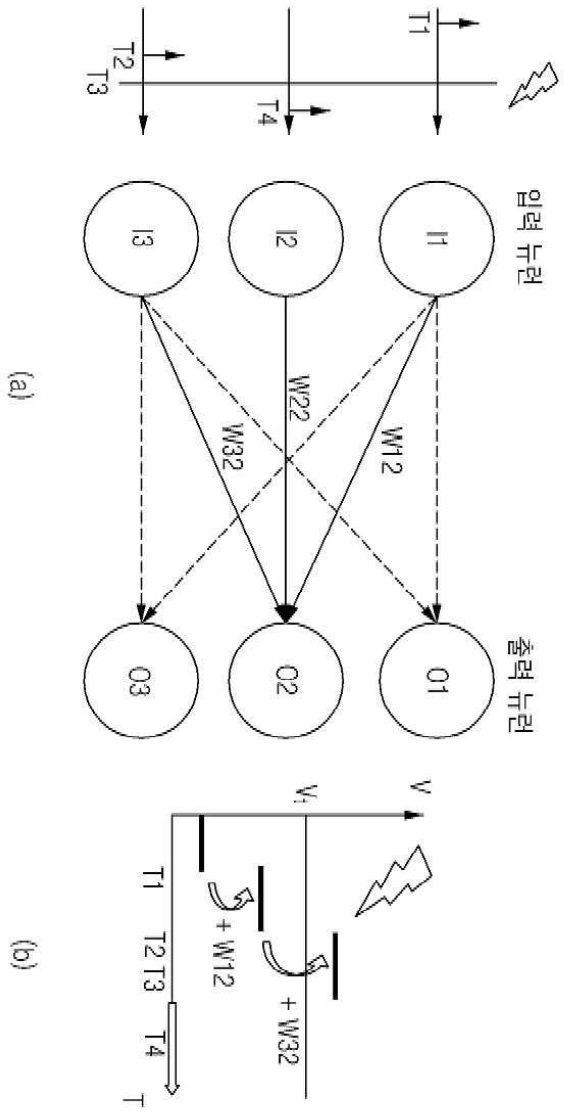
[0073] [3] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "Xnornet: Imagenet classification using binary convolutional neural networks", European Conference on Computer Vision, pp. 525-542. 2016.

[0074] [4] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167, 2015.

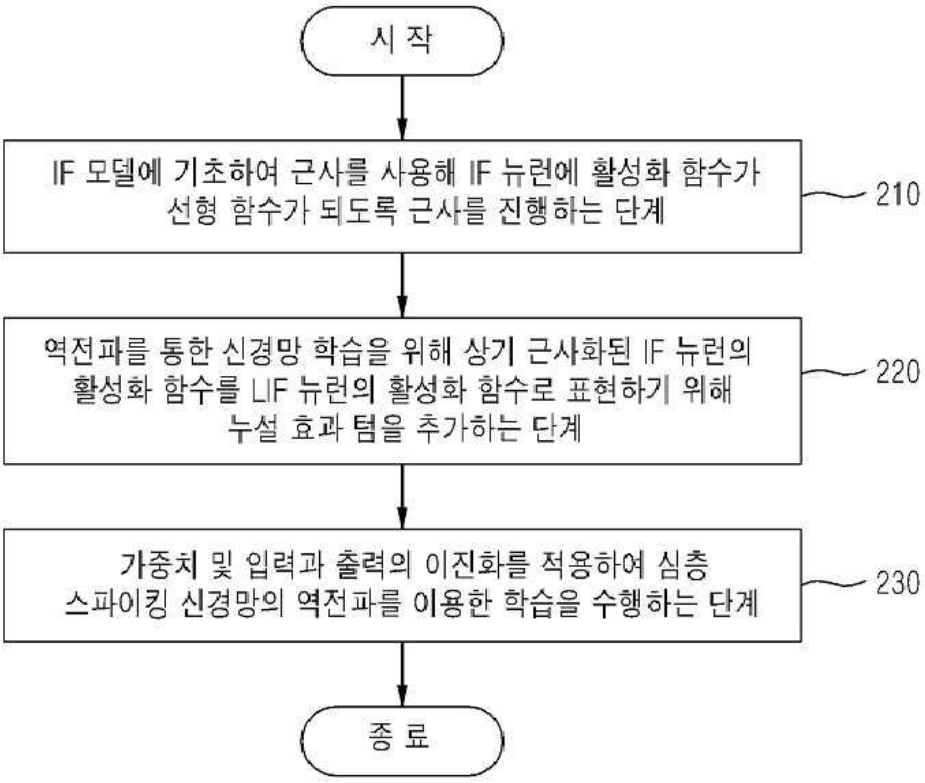
[0075] [5] C Lee, SS Sarwar, P Panda and G Srinivasan. "Enabling Spike-Based Backpropagation for Training Deep Neural Network Architectures", Front. Neurosci., Feburary 2020.

도면

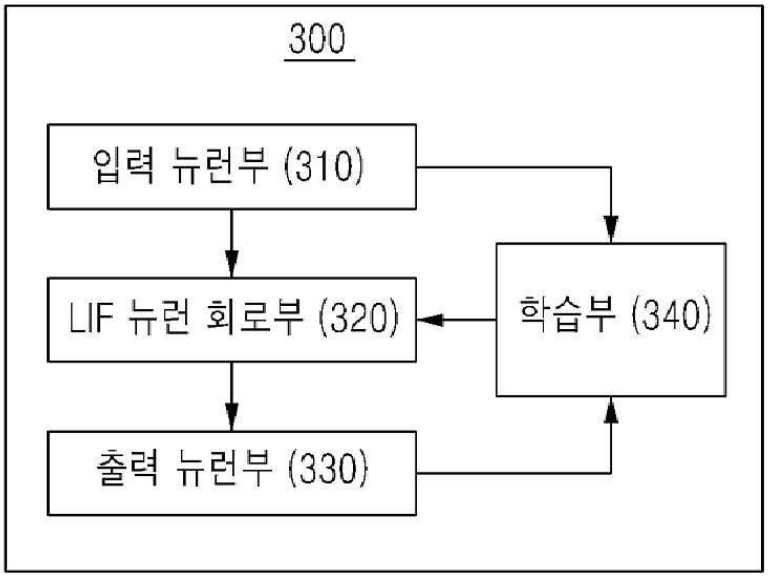
도면1



도면2


시 작
IF 모델에 기초하여 근사를 사용해 IF 뉴런에 활성화 함수가 선형 함수가 되도록 근사를 진행하는 단계
210
역전파를 통한 신경망 학습을 위해 상기 근사화된 IF 뉴런의 활성화 함수를 LIF 뉴런의 활성화 함수로 표현하기 위해 누설 효과 텀을 추가하는 단계
220
가중치 및 입력과 출력의 이진화를 적용하여 심층 스파이킹 신경망의 역전파를 이용한 학습을 수행하는 단계
230
종 료

도면3


300
입력 뉴런부 (310)
LIF 뉴런 회로부 (320)
학습부 (340)
출력 뉴런부 (330)

도면4

