



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년12월13일
(11) 등록번호 10-2612695
(24) 등록일자 2023년12월07일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) G06N 3/04 (2023.01)
(52) CPC특허분류
G06N 3/08 (2023.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2022-0136035
(22) 출원일자 2022년10월20일
심사청구일자 2022년10월20일
(56) 선행기술조사문헌
Zhiqiang Shen et al., "S2-BNN: Bridging the Gap Between Self-Supervised Real and 1-bit Neural Networks via Guided Distribution Calibration," arXiv:2102.08946v2 [cs.CV] 21 Jun 2021 (2021.06.21.)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
최종현
서울특별시 서대문구 증가로 191, 101동 1603호 (남가좌동, DMC래미안클라시스)
김다현
경기도 용인시 수지구 광교중앙로 305, C동 613호 (상현동, 광고2차푸르지오시티)
(74) 대리인
특허법인 수

전체 청구항 수 : 총 22 항

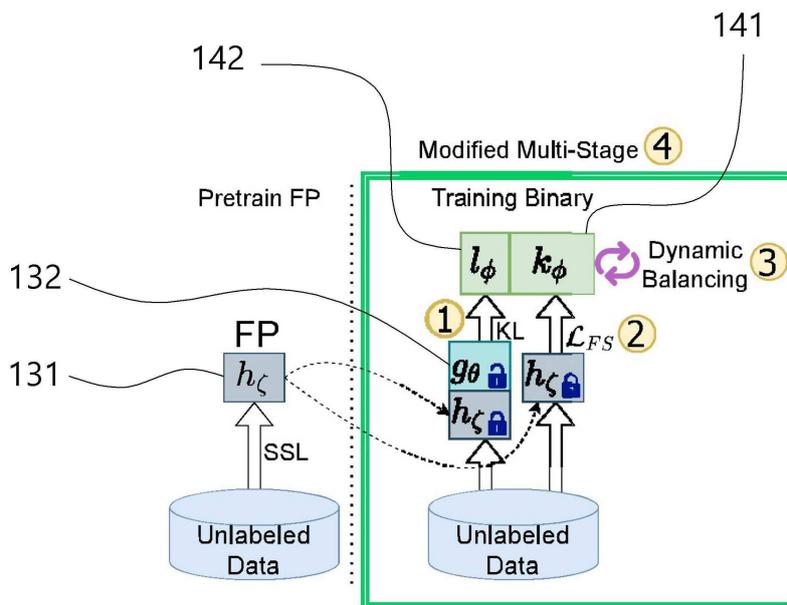
심사관 : 송근배

(54) 발명의 명칭 변동성 플로팅 포인트 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기지도학습 방법 및 장치, 그리고 이를 이용한 테스트 방법 및 테스트 장치

(57) 요약

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기지도학습(Self-Supervised Learning, SSL) 방법이 개시된다. 즉, (a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습 (뒷면에 계속)

대표도 - 도4



용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및 (b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 단계를 포함하는 것을 특징으로 하는 방법이 개시된다.

- 이 발명을 지원한 국가연구개발사업
- | | |
|-------------|---|
| 과제고유번호 | 1711160525 |
| 과제번호 | 2022-0-00959-001 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 사람중심인공지능핵심원천기술개발 |
| 연구과제명 | (2세부) 의사결정 지원을 위한 퓨샷 학습 기반시각 및 언어에 대한 인과관계 추론기술개발 |
| 기여율 | 20/100 |
| 과제수행기관명 | 서울대학교 산학협력단 |
| 연구기간 | 2022.04.01 ~ 2022.12.31 |
- 이 발명을 지원한 국가연구개발사업
- | | |
|-------------|--|
| 과제고유번호 | 1711160535 |
| 과제번호 | 2022-0-00871-001 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 사람중심인공지능핵심원천기술개발 |
| 연구과제명 | (1세부) 인공지능 에이전트 협업기반 신경망변이 및 지능 강화 기술 개발 |
| 기여율 | 20/100 |
| 과제수행기관명 | 한국전자통신연구원 |
| 연구기간 | 2022.04.01 ~ 2022.12.31 |
- 이 발명을 지원한 국가연구개발사업
- | | |
|-------------|--------------------------------------|
| 과제고유번호 | 1711153253 |
| 과제번호 | 2022R1A2C4002300 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 한국연구재단 |
| 연구사업명 | 개인기초연구(과기정통부) |
| 연구과제명 | 인간 수준의 인공지능을 위한 증가적 멀티 모달 인지 에이전트 연구 |
| 기여율 | 20/100 |
| 과제수행기관명 | 연세대학교 |
| 연구기간 | 2022.03.01 ~ 2023.02.28 |
- 이 발명을 지원한 국가연구개발사업
- | | |
|-------------|-------------------------------------|
| 과제고유번호 | 1711160620 |
| 과제번호 | 2022-0-00077-001 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 사람중심인공지능핵심원천기술개발 |
| 연구과제명 | 이종데이터기반 상식 추출, 이해, 추론을 위한 인공지능 기술개발 |
| 기여율 | 20/100 |
| 과제수행기관명 | 한국과학기술원 |
| 연구기간 | 2022.04.01 ~ 2022.12.31 |
- 이 발명을 지원한 국가연구개발사업
- | | |
|-------------|-------------------------|
| 과제고유번호 | 1711152718 |
| 과제번호 | 2020-0-01361-003 |
| 부처명 | 과학기술정보통신부 |
| 과제관리(전문)기관명 | 정보통신기획평가원 |
| 연구사업명 | 정보통신방송혁신인재양성(R&D) |
| 연구과제명 | 인공지능대학원지원(연세대학교) |
| 기여율 | 10/100 |
| 과제수행기관명 | 연세대학교 산학협력단 |
| 연구기간 | 2022.01.01 ~ 2022.12.31 |
- 이 발명을 지원한 국가연구개발사업

과제고유번호	1711152863
과제번호	2021-0-02068-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능 혁신 허브 연구 개발
기여율	5/100
과제수행기관명	경희대학교산학협력단
연구기간	2022.01.01 ~ 2022.12.31
이 발명을 지원한 국가연구개발사업	
과제고유번호	1711160465
과제번호	2022-0-00113-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	사람중심인공지능핵심원천기술개발
연구과제명	지속 가능한 협업형 멀티 모달 평생 학습 프레임워크 개발
기여율	5/100
과제수행기관명	연세대학교 산학협력단
연구기간	2022.04.01 ~ 2022.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 방법에 있어서,

(a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 단계

를 포함하되,

상기 (a) 단계는,

상기 학습 장치가, 상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하는 것을 특징으로 하는 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[\mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), b_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터 세트를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, h_{ζ} 는 상기 FP 특징추출 연산을 의미하며, b_{ϕ} 는 상기 바이너리 연산을 의미하는 것을 특징으로 하는 방법.

청구항 4

제1항에 있어서,

상기 (a) 단계는,

상기 학습 장치가, 상기 바이너리 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 특징부 파라미터를 포함하는 바이너리 특징부로 하여금, 상기 학습용 벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 특징추출 연산을 인가하여 학습용 바이너리 특징벡터를 생성하도록 한 다음, (ii) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 기능부 파라미터를 포함하는 바이너리 기능부로 하여금, 상기 학습용 바이너리 특징벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 기능 연산을 인가하여 상기 학습용 바이너리 출력을 생성하도록 하는 것을 특징으로 하는 방법.

청구항 5

제4항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, (i) 상기 학습용 바이너리 출력 및 상기 학습용 FP 출력을 참조하여 KL 발산 로스를 생성하는 프로세스 및 (ii) 상기 학습용 FP 특징벡터 및 상기 학습용 바이너리 특징벡터를 참조하여 피처 유사성 로스를 생성하는 프로세스를 수행한 다음, 상기 KL 발산 로스 및 상기 피처 유사성 로스를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 방법.

청구항 6

제5항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, λ 는 기설정된 정적 밸런싱 파라미터를 의미하고, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하며, 상기 평균 연산의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피처 유사성 로스를 의미하는 것을 특징으로 하는 방법.

청구항 7

제6항에 있어서,

상기 학습 장치가, 하기 수식에 따라 상기 피처 유사성 연산을 수행하되,

$$\mathcal{L}_{FS}(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$$

상기 수식에서, v_1 및 v_2 는 임의의 입력 벡터이고, $\frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$ 는 v_1 및 v_2 간의 코사인 유사도를 의미하는 것을 특징으로 하는 방법.

청구항 8

제5항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, 상기 KL 발산 로스 및 상기 피쳐 유사성 로스와 함께, Iteration 수에 따라 변동되는 동적 밸런싱 파라미터를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 방법.

청구항 9

제8항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda(t)) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda(t) \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, $\lambda(t)$ 는 상기 동적 밸런싱 파라미터를 의미하고, t 는 상기 Iteration 수를 의미하며, \mathcal{L}_{FS} 는 피쳐 유사성 연산을 의미하고, 상기 평균 연산 내부의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피쳐 유사성 로스를 의미하는 것을 특징으로 하는 방법.

청구항 10

제9항에 있어서,

상기 (b) 단계는,

상기 학습 장치가, 하기 수식에 따라 상기 동적 밸런싱 파라미터를 계산하되,

$$\lambda(t) = \lambda_{T_{max}} - (\lambda_{T_{max}} - \lambda_0) \cdot \frac{\cos\left(\frac{\pi t}{T_{max}}\right) + 1}{2}$$

상기 수식에서, T_{max} 는 기설정된 최대 Iteration 수를 의미하고, $\lambda_{T_{max}}$ 는 상기 최대 Iteration 시의 기설정된 밸런싱 파라미터를 의미하며, λ_0 는 최초 Iteration 시의 기설정된 밸런싱 파라미터를 의미하고, $\lambda_{T_{max}}$ 가 λ_0 에 비해 작은 것을 특징으로 하는 방법.

청구항 11

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기지도학습(Self-Supervised Learning, SSL) 방법에 있어서,

(a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상

기 FP 파라미터 중 적어도 일부를 학습하는 단계

를 포함하되,

멀티스테이지 학습 프로세스를 적용하는 경우,

상기 (a) 단계 이전에,

(a0) 상기 학습 장치가, 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 단계;

를 더 포함하고,

(c) 상기 학습 장치가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(d) 상기 학습 장치가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 단계

를 더 포함하는 것을 특징으로 하는 방법.

청구항 12

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 자기지도학습(Self-Supervised Learning, SSL)된 바이너리 뉴럴넷을 테스트하는 방법에 있어서,

(a) (1) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식의 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (2) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행함으로써 학습이 완료된 상태에서, 테스트용 벡터가 획득되면, 테스트 장치가, 상기 바이너리 뉴럴넷으로 하여금, 상기 바이너리 뉴럴넷에 적어도 하나의 바이너리 연산을 인가하도록 하는 단계; 및

(b) 상기 테스트 장치가, 상기 바이너리 뉴럴넷으로부터 획득된 테스트용 바이너리 출력을 사용자에게 제공하는 단계

를 포함하되,

상기 (a) 단계는,

상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하여 학습을 완료한 상태인 것을 특징으로 하는 방법.

청구항 13

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 장치에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (II) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행하되,

상기 (I) 프로세스는,

상기 프로세서가, 상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하는 것을 특징으로 하는 장치.

청구항 14

삭제

청구항 15

제13항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[\mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), b_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터 세트를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, h_{ζ} 는 상기 FP 특징추출 연산을 의미하며, b_{ϕ} 는 상기 바이너리 연산을 의미하는 것을 특징으로 하는 장치.

청구항 16

제13항에 있어서,

상기 (I) 프로세스는,

상기 프로세서가, 상기 바이너리 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 특징부 파라미터를 포함하는 바이너리 특징부로 하여금, 상기 학습용 벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 특징추출 연산을 인가하여 학습용 바이너리 특징벡터를 생성하도록 한 다음, (ii) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인

바이너리 기능부 파라미터를 포함하는 바이너리 기능부로 하여금, 상기 학습용 바이너리 특징벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 기능 연산을 인가하여 상기 학습용 바이너리 출력을 생성하도록 하는 것을 특징으로 하는 장치.

청구항 17

제16항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, (i) 상기 학습용 바이너리 출력 및 상기 학습용 FP 출력을 참조하여 KL 발산 로스를 생성하는 프로세스 및 (ii) 상기 학습용 FP 특징벡터 및 상기 학습용 바이너리 특징벡터를 참조하여 피쳐 유사성 로스를 생성하는 프로세스를 수행한 다음, 상기 KL 발산 로스 및 상기 피쳐 유사성 로스를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 장치.

청구항 18

제17항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, λ 는 기설정된 정적 밸런싱 파라미터를 의미하고, \mathcal{L}_{FS} 는 피쳐 유사성 연산을 의미하며, 상기 평균 연산의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피쳐 유사성 로스를 의미하는 것을 특징으로 하는 장치.

청구항 19

제18항에 있어서,

상기 프로세서가, 하기 수식에 따라 상기 피쳐 유사성 연산을 수행하되,

$$\mathcal{L}_{FS}(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$$

상기 수식에서, v_1 및 v_2 는 임의의 입력 벡터이고, $\frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$ 는 v_1 및 v_2 간의 코사인 유사도를 의미하는 것을 특징으로 하는 장치.

청구항 20

제17항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, 상기 KL 발산 로스 및 상기 피쳐 유사성 로스와 함께, Iteration 수에 따라 변동되는 동적 밸런싱 파라미터를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 장치.

청구항 21

제20항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda(t)) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda(t) \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, $\lambda(t)$ 는 상기 동적 밸런싱 파라미터를 의미하고, t 는 상기 Iteration 수를 의미하며, \mathcal{L}_{FS} 는 피쳐 유사성 연산을 의미하고, 상기 평균 연산 내부의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피쳐 유사성 로스를 의미하는 것을 특징으로 하는 장치.

청구항 22

제21항에 있어서,

상기 (II) 프로세스는,

상기 프로세서가, 하기 수식에 따라 상기 동적 밸런싱 파라미터를 계산하되,

$$\lambda(t) = \lambda_{T_{max}} - (\lambda_{T_{max}} - \lambda_0) \cdot \frac{\cos\left(\frac{\pi t}{T_{max}}\right) + 1}{2}$$

상기 수식에서, T_{max} 는 기설정된 최대 Iteration 수를 의미하고, $\lambda_{T_{max}}$ 는 상기 최대 Iteration 시의 기설정된 밸런싱 파라미터를 의미하며, λ_0 는 최초 Iteration 시의 기설정된 밸런싱 파라미터를 의미하고, $\lambda_{T_{max}}$ 와 λ_0 에 비해 작은 것을 특징으로 하는 장치.

청구항 23

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 장치에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식의 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (II) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행하되,

멀티스태이지 학습 프로세스를 적용하는 경우,

상기 (I) 프로세스 이전에,

(I0) 상기 프로세서가, 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 프로세스;

를 더 수행하고,

(III) 상기 프로세서가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (IV) 상기 프로세서가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 프로세스를 더 수행하는 것을 특징으로 하는 장치.

청구항 24

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 자기지도학습(Self-Supervised Learning, SSL)된 바이너리 뉴럴넷의 테스트 장치에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) (1) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (2) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행함으로써 학습이 완료된 상태에서, 테스트용 벡터가 획득되면, 상기 바이너리 뉴럴넷으로 하여금, 상기 바이너리 뉴럴넷에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스; 및 (II) 상기 바이너리 뉴럴넷으로부터 획득된 테스트용 바이너리 출력을 사용자에게 제공하는 프로세스를 수행하도록 설정된 것을 특징으로 하되,

상기 (I) 프로세스는,

상기 프로세서가, 상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하여 학습을 완료한 상태인 것을 특징으로 하는 장치.

발명의 설명

기술 분야

[0001] 본 발명은 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법 및 장치, 그리고 이를 이용한 테스트 방법 및 테스트 장치에 관한 것이다.

배경 기술

[0002] FP 뉴럴넷이란, 웨이트 및 액티베이션 등의 값들을 FP 형태로 사용하는 네트워크로서, 정확한 값을 연산에 사용한다는 점에서 정확도 면의 장점이 있으나, 컴퓨팅 파워를 많이 소모하는 단점이 있다. 이와 반대로, 웨이트 및

액티베이션 등의 값들을 바이너리 형태로 사용하는 바이너리 뉴럴넷은, FP 뉴럴넷에 비해 정확도가 떨어질 수 있으나, 일반적으로 모델의 크기도 FP 뉴럴넷에 비해 훨씬 작고, 컴퓨팅 파워를 훨씬 덜 소모하는 장점이 있는 네트워크이다. 최근 저전력 기기의 사용이 많아짐에 따라, 효율성 면의 장점이 있는 바이너리 뉴럴넷에 대한 수요가 높아지고 있는 상황이다.

[0003] 한편, 정답 데이터가 존재하지 않는 학습 데이터만을 사용하여 학습을 수행하는 비지도학습의 한 갈래인, 학습 데이터를 사용하여 pretext task를 학습한 다음 downstream task로 transfer하여 학습을 수행하는 방식인 SSL은 최근 엄청난 발전을 이루어, 라벨링 비용이 들지 않으면서도 일반적인 지도학습과 동일하거나 오히려 더 나은 성능을 보이고 있는 방법으로, 이를 바이너리 뉴럴넷에 적용할 수 있다면 상당한 기술적 진전을 이룰 수 있을 것이다. 지금까지 연구된 SSL 방법은 주로 FP 뉴럴넷과 같은 큰 모델을 기준으로 한 것이 대부분이다. 이와 같은 상황에서, Shen et al.은 S2-bnn: Bridging the gap between self-supervised real and 1-bit neural networks via guided distribution calibration(2021)에서 SSL 방법을 바이너리 뉴럴넷에 적용한 S2-BNN을 제안하였는데, 이에 대해 이하 설명하도록 한다.

[0004] 도 6a 및 도 6b는 각각 종래의 Supervised KL Div. 학습 방법 및 이에 SSL 방법을 적용하여 응용한 S2-BNN 학습 방법을 나타낸 도면이다.

[0005] 도 6a를 참조하면, 일반적으로 바이너리 뉴럴넷을 학습하는 데에 사용되는 Supervised KL Div. 방법을 확인할 수 있는데, 개략적으로 설명하면, 이는 FP 뉴럴넷을 지도학습한 상태에서, 학습된 FP 뉴럴넷의 출력과 바이너리 뉴럴넷의 출력을 비교하여 로스를 생성한 다음 바이너리 뉴럴넷의 출력이 학습된 FP 뉴럴넷의 출력과 유사해지도록, 즉 FP 뉴럴넷을 바이너리 뉴럴넷의 타겟으로 삼아 학습을 수행하는 방식이다. 여기서 KL 발산 로스 함수는 통상의 기술자에게 널리 알려진 것이므로 설명을 생략하도록 한다.

[0006] 또한, 도 6b를 참조하면, S2-BNN은 이와 같은 방식을 응용하여, FP 뉴럴넷을 SSL 방법을 통해 학습시킨 다음 학습된 FP 뉴럴넷을 바이너리 뉴럴넷의 타겟으로 삼아 학습을 수행하는 방식인데, SSL을 바이너리 뉴럴넷에 적용하고자 한 시도들 중에는 성능이 좋은 편이나, 학습이 완료되어 불변하는 FP 뉴럴넷을 고정된 타겟으로 삼는 방식 때문에 다소의 비효율이 초래되는 문제가 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 상술한 문제점을 해결하는 것을 목적으로 한다.

[0008] 또한 본 발명은 FP 뉴럴넷의 FP 기능부 파라미터가 고정되지 않은 상태에서 바이너리 뉴럴넷을 학습하면서 함께 학습되도록 함으로써 학습 시나리오의 진행에 따라 변화되는 FP 뉴럴넷을 이동 표적으로 사용하는, 바이너리 뉴럴넷의 SSL 방법을 제공하는 것을 목적으로 한다.

[0009] 또한 본 발명은 피쳐 유사성 로스, 동적 밸런싱 및 멀티스테이지 학습 방식을 추가적으로 적용하여 바이너리 뉴럴넷의 정확도를 향상시킬 수 있는 SSL 방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0010] 상기한 바와 같은 본 발명의 목적을 달성하고, 후술하는 본 발명의 특징적인 효과를 실현하기 위한 본 발명의 특징적인 구성은 하기와 같다.

[0011] 본 발명의 일 태양에 따르면, 변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기지도학습(Self-Supervised Learning, SSL) 방법에 있어서, (a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및 (b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도

도 일부를 학습하는 단계를 포함하는 것을 특징으로 하는 방법이 개시된다.

[0012] 일례로서, 상기 (a) 단계는, 상기 학습 장치가, 상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하는 것을 특징으로 하는 방법이 개시된다.

[0013] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} [\mathcal{L}_{KL}(g_{\theta}(h_{\zeta}(x)), b_{\phi}(x))]$$

[0015] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터 세트를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, h_{ζ} 는 상기 FP 특징추출 연산을 의미하며, b_{ϕ} 는 상기 바이너리 연산을 의미하는 것을 특징으로 하는 방법이 개시된다.

[0016] 일례로서, 상기 (a) 단계는, 상기 학습 장치가, 상기 바이너리 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 특징부 파라미터를 포함하는 바이너리 특징부로 하여금, 상기 학습용 벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 특징추출 연산을 인가하여 학습용 바이너리 특징벡터를 생성하도록 한 다음, (ii) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 기능부 파라미터를 포함하는 바이너리 기능부로 하여금, 상기 학습용 바이너리 특징벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 기능 연산을 인가하여 상기 학습용 바이너리 출력을 생성하도록 하는 것을 특징으로 하는 방법이 개시된다.

[0017] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, (i) 상기 학습용 바이너리 출력 및 상기 학습용 FP 출력을 참조하여 KL 발산 로스를 생성하는 프로세스 및 (ii) 상기 학습용 FP 특징벡터 및 상기 학습용 바이너리 특징벡터를 참조하여 피쳐 유사성 로스를 생성하는 프로세스를 수행한 다음, 상기 KL 발산 로스 및 상기 피쳐 유사성 로스를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 방법이 개시된다.

[0018] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} [(1 - \lambda)\mathcal{L}_{KL}(g_{\theta}(h_{\zeta}(x)), l_{\phi}(k_{\phi}(x))) + \lambda\mathcal{L}_{FS}(h_{\zeta}(x), k_{\phi}(x))]$$

[0020] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, λ 는 기설정된 정적 밸런싱 파라미터를 의미하고, \mathcal{L}_{FS} 는 피쳐 유사성 연산을 의미하며, 상기 평균 연산의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피쳐 유사성 로스를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0021] 일례로서, 상기 학습 장치가, 하기 수식에 따라 상기 피처 유사성 연산을 수행하되,

$$[0022] \mathcal{L}_{FS}(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$$

[0023] 상기 수식에서, v_1 및 v_2 는 임의의 입력 벡터이고, $\frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$ 는 v_1 및 v_2 간의 코사인 유사도를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0024] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, 상기 KL 발산 로스 및 상기 피처 유사성 로스와 함께, Iteration 수에 따라 변동되는 동적 밸런싱 파라미터를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 방법이 개시된다.

[0025] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$[0026] \min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda(t)) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda(t) \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

[0027] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, $\lambda(t)$ 는 상기 동적 밸런싱 파라미터를 의미하고, t 는 상기 Iteration 수를 의미하며, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하고, 상기 평균 연산 내부의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피처 유사성 로스를 의미하는 것을 특징으로 하는 방법이 개시된다.

[0028] 일례로서, 상기 (b) 단계는, 상기 학습 장치가, 하기 수식에 따라 상기 동적 밸런싱 파라미터를 계산하되,

$$[0029] \lambda(t) = \lambda_{T_{max}} - (\lambda_{T_{max}} - \lambda_0) \cdot \frac{\cos\left(\frac{\pi t}{T_{max}}\right) + 1}{2}$$

[0030] 상기 수식에서, T_{max} 는 기설정된 최대 Iteration 수를 의미하고, $\lambda_{T_{max}}$ 는 상기 최대 Iteration 시의 기설정된 밸런싱 파라미터를 의미하며, λ_0 는 최초 Iteration 시의 기설정된 밸런싱 파라미터를 의미하고, $\lambda_{T_{max}}$ 와 λ_0 에 비해 작은 것을 특징으로 하는 방법이 개시된다.

[0031] 일례로서, 멀티스테이지 학습 프로세스를 적용하는 경우, 상기 (a) 단계 이전에, (a0) 상기 학습 장치가, 상기 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 단계; 를 더 포함하고, (c) 상기 학습 장치가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 상기 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및 (d) 상기 학습 장치가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 단계를 더 포함하는 것을 특징으로 하는 방법이 개시된다.

[0032] 본 발명의 다른 태양에 따르면, 변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 자기지도학습(Self-Supervised Learning, SSL)된 바이너리 뉴럴넷을 테스트하는 방법에 있어서, (a) 테스트 장치가, (1) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바

이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (2) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행함으로써 학습이 완료된 상태에서, 테스트용 벡터가 획득되면, 상기 바이너리 뉴럴넷으로 하여금, 상기 바이너리 뉴럴넷에 적어도 하나의 바이너리 연산을 인가하도록 하는 단계; 및 (b) 상기 테스트 장치가, 상기 바이너리 뉴럴넷으로부터 획득된 테스트용 바이너리 출력을 사용자에게 제공하는 단계를 포함하는 것을 특징으로 하는 방법이 개시된다.

[0033] 본 발명의 또다른 태양에 따르면, 변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기지도학습(Self-Supervised Learning, SSL) 장치에 있어서, 인스트럭션들을 저장하는 하나 이상의 메모리; 및 상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (II) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행하는 것을 특징으로 하는 장치가 개시된다.

[0034] 일례로서, 상기 (I) 프로세스는, 상기 프로세서가, 상기 변동성 FP 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부로 하여금, 상기 학습용 벡터에 상기 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 한 다음, (ii) 상기 변동성 FP 뉴럴넷에 포함된, 상기 FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부로 하여금, 상기 학습용 FP 특징벡터에 상기 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 상기 학습용 FP 출력을 생성하도록 하는 것을 특징으로 하는 장치가 개시된다.

[0035] 일례로서, 상기 (II) 프로세스는, 상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[\mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), b_{\phi}(x) \right) \right]$$

[0037] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터 세트를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, h_{ζ} 는 상기 FP 특징추출 연산을 의미하며, b_{ϕ} 는 상기 바이너리 연산을 의미하는 것을 특징으로 하는 장치가 개시된다.

[0038] 일례로서, 상기 (I) 프로세스는, 상기 프로세서가, 상기 바이너리 뉴럴넷에 상기 학습용 벡터를 입력하여, (i) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 특징부 파라미터를 포함하는 바이너리 특징부로 하여금, 상기 학습용 벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 특징추출 연산을 인가하여 학습용 바이너리 특징벡터를 생성하도록 한 다음, (ii) 상기 바이너리 뉴럴넷에 포함된, 상기 바이너리 파라미터 중 적어도 일부인 바이너리 기능부 파라미터를 포함하는 바이너리 기능부로 하여금, 상기 학습용 바이너리 특징벡터에 상기 바이너리 연산의 적어도 일부인 바이너리 기능 연산을 인가하여 상기 학습용 바이

너리 출력을 생성하도록 하는 것을 특징으로 하는 장치가 개시된다.

[0039] 일례로서, 상기 (II) 프로세스는, 상기 프로세서가, (i) 상기 학습용 바이너리 출력 및 상기 학습용 FP 출력을 참조하여 KL 발산 로스를 생성하는 프로세스 및 (ii) 상기 학습용 FP 특징벡터 및 상기 학습용 바이너리 특징벡터를 참조하여 피처 유사성 로스를 생성하는 프로세스를 수행한 다음, 상기 KL 발산 로스 및 상기 피처 유사성 로스를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 장치가 개시된다.

[0040] 상기 (II) 프로세스는, 상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하되,

$$[0041] \min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

[0042] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, λ 는 기설정된 정적 밸런싱 파라미터를 의미하고, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하며, 상기 평균 연산의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피처 유사성 로스를 의미하는 것을 특징으로 하는 장치가 개시된다.

[0043] 일례로서, 상기 프로세서가, 하기 수식에 따라 상기 피처 유사성 연산을 수행하되,

$$[0044] \mathcal{L}_{FS}(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$$

[0045] 상기 수식에서, v_1 및 v_2 는 임의의 입력 벡터이고, $\frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$ 는 v_1 및 v_2 간의 코사인 유사도를 의미하는 것을 특징으로 하는 장치가 개시된다.

[0046] 일례로서, 상기 (II) 프로세스는, 상기 프로세서가, 상기 KL 발산 로스 및 상기 피처 유사성 로스와 함께, Iteration 수에 따라 변동되는 동적 밸런싱 파라미터를 참조하여 상기 커플링 로스를 생성하는 것을 특징으로 하는 장치가 개시된다.

[0047] 일례로서, 상기 (II) 프로세스는, 상기 프로세서가, 하기 수식에 따라 상기 커플링 로스를 생성하여 학습을 수행하되,

$$[0048] \min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda(t)) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda(t) \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

[0049] 상기 수식에서, θ 는 상기 FP 기능부 파라미터를 의미하고, ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 상기 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 상기 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 상기 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, $\lambda(t)$ 는 상기 동적 밸런싱 파라미터를 의미하고, t 는 상기 Iteration 수를 의미하며, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하고, 상기 평균 연산 내부의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피처 유사성 로스를 의미하는 것을 특징으로 하는 장치가 개시된다.

[0050] 일례로서, 상기 (II) 프로세스는, 상기 프로세서가, 하기 수식에 따라 상기 동적 밸런싱 파라미터를 계산하되,

$$\lambda(t) = \lambda_{T_{max}} - (\lambda_{T_{max}} - \lambda_0) \cdot \frac{\cos\left(\frac{\pi t}{T_{max}}\right) + 1}{2}$$

[0051] 상기 수식에서, T_{max} 는 기설정된 최대 Iteration 수를 의미하고, $\lambda_{T_{max}}$ 는 상기 최대 Iteration 시의 기설정된 밸런싱 파라미터를 의미하며, λ_0 는 최초 Iteration 시의 기설정된 밸런싱 파라미터를 의미하고, $\lambda_{T_{max}}$ 와 λ_0 에 비해 작은 것을 특징으로 하는 장치가 개시된다.

[0052] 일례로서, 멀티스테이지 학습 프로세스를 적용하는 경우, 상기 (I) 프로세스 이전에, (IO) 상기 프로세서가, 상기 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 프로세스; 를 더 수행하고, (III) 상기 프로세서가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 상기 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (IV) 상기 프로세서가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 프로세스를 더 수행하는 것을 특징으로 하는 장치가 개시된다.

[0053] 본 발명의 또다른 태양에 따르면, 변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 자기지도학습(Self-Supervised Learning, SSL)된 바이너리 뉴럴넷의 테스트 장치에 있어서, 인스트럭션들을 저장하는 하나 이상의 메모리; 및 상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) (1) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (2) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행함으로써 학습이 완료된 상태에서, 테스트용 벡터가 획득되면, 상기 바이너리 뉴럴넷으로 하여금, 상기 바이너리 뉴럴넷에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스; 및 (II) 상기 바이너리 뉴럴넷으로부터 획득된 테스트용 바이너리 출력을 사용자에게 제공하는 프로세스를 수행하도록 설정된 것을 특징으로 하는 장치가 개시된다.

발명의 효과

[0054] 본 발명은 FP 뉴럴넷의 기능부가 고정되지 않은 상태에서 바이너리 뉴럴넷을 학습하면서 함께 학습되도록 함으로써 학습 시나리오의 진행에 따라 변화되는 FP 뉴럴넷을 이동 표적으로 사용하는, 바이너리 뉴럴넷의 SSL 방법을 제공할 수 있는 효과가 있다.

[0055] 또한 본 발명은 피치 유사성 로스, 동적 밸런싱 및 멀티스테이지 학습 방식을 추가적으로 적용하여 바이너리 뉴럴넷의 정확도를 향상시킬 수 있는 SSL 방법을 제공할 수 있는 효과가 있다.

도면의 간단한 설명

[0056] 도 1은 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 수행하는 학습 장치의 구성을 나타낸 도면이다.

도 2a 및 2b는 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 수행하는 데에 사용되는 변동성 FP 뉴럴넷 및 바이너리 뉴럴넷의 구성을 나타낸 도면이다.

도 3은 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 나타낸 흐름도이다.

도 4는 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법의 실시예들을 단계적으로 나타낸 도면이다.

도 5는 본 발명의 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법의 각 실시예를 적용함에 따른 테스트 시의 정확도의 변화를 나타낸 그래프이다.

도 6a 및 도 6b는 각각 종래의 Supervised KL Div. 학습 방법 및 이에 SSL 방법을 적용하여 응용한 S2-BNN 학습 방법을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0058] 후술하는 본 발명에 대한 상세한 설명은, 본 발명이 실시될 수 있는 특정 실시예를 예시로서 도시하는 첨부 도면을 참조한다. 이들 실시예는 당업자가 본 발명을 실시할 수 있기에 충분하도록 상세히 설명된다. 본 발명의 다양한 실시예는 서로 다르지만 상호 배타적일 필요는 없음이 이해되어야 한다. 예를 들어, 여기에 기재되어 있는 특정 형상, 구조 및 특성은 일 실시예에 관련하여 본 발명의 정신 및 범위를 벗어나지 않으면서 다른 실시예로 구현될 수 있다. 또한, 각각의 개시된 실시예 내의 개별 구성요소의 위치 또는 배치는 본 발명의 정신 및 범위를 벗어나지 않으면서 변경될 수 있음이 이해되어야 한다. 따라서, 후술하는 상세한 설명은 한정적인 의미로서 취하려는 것이 아니며, 본 발명의 범위는, 적절하게 설명된다면, 그 청구항들이 주장하는 것과 균등한 모든 범위와 더불어 첨부된 청구항에 의해서만 한정된다. 도면에서 유사한 참조부호는 여러 측면에 걸쳐서 동일하거나 유사한 기능을 지칭한다.

[0059] 이하, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 하기 위하여, 본 발명의 바람직한 실시예들에 관하여 첨부된 도면을 참조하여 상세히 설명하기로 한다.

[0060] 도 1은 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 수행하는 학습 장치의 구성을 나타낸 도면이다.

[0061] 도 1을 참조하면, 학습 장치(100)는 변동성 FP 뉴럴넷(130) 및 바이너리 뉴럴넷(140)을 포함할 수 있다. 이 때, 변동성 FP 뉴럴넷(130) 및 바이너리 뉴럴넷(140)의 입출력 및 연산 과정은 각각 통신부(110) 및 프로세서(120)에 의해 이루어질 수 있다. 다만, 도 1에서는 통신부(110) 및 프로세서(120)의 구체적인 연결 관계를 생략하였다. 또한, 메모리(115)는 후술할 여러 가지 인스트럭션들을 저장한 상태일 수 있고, 프로세서(120)는 메모리에 저장된 인스트럭션들을 수행하도록 됨으로써 추후 설명할 프로세스들을 수행하여 본 발명을 수행할 수 있다. 이와 같이 학습 장치(100)가 묘사되었다고 하여, 학습 장치(100)가 본 발명을 실시하기 위한 미디엄, 프로세서 및 메모리가 통합된 형태인 통합 프로세서를 포함하는 경우를 배제하는 것은 아니다. 이하 변동성 FP 뉴럴넷(130) 및 바이너리 뉴럴넷(140)의 구성에 대해 설명하도록 한다.

[0062] 도 2a 및 2b는 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 수행하는 데에 사용되는 변동성 FP 뉴럴넷 및 바이너리 뉴럴넷의 구성을 나타낸 도면이다.

[0063] 도 2a를 참조하면, 변동성 FP 뉴럴넷(130)은 FP 특징부(131) 및 FP 기능부(132)를 포함함을 알 수 있다. 여기서 FP 특징부(131)는, 컨벌루션 연산 등을 사용하여 입력으로부터 특징 값, 즉 FP 특징벡터를 추출하는 기능을 수행할 수 있으며, 내부적인 구성은 일반적인 뉴럴 네트워크들의 해당 부분과 동일 또는 유사할 수 있다. 또한, FP 기능부(132)는, FP 특징부(131)를 통해 획득된 FP 특징벡터를 사용하여 변동성 FP 뉴럴넷(130)의 구성 목적에 따른 기능을 수행할 수 있는데, 예를 들어 이미지 분류, 이미지 세그멘테이션, 패턴 인식, 제어 등의 다양한 기능들 중 하나일 수 있으며, 이의 내부 구조는 일반적으로 알려진 뉴럴 네트워크가 해당 기능을 수행하는 경우의 해당 기능부와 동일 또는 유사할 수 있다. 이와 같은 FP 특징부(131) 및 FP 기능부(132)의 웨이트 및 액티베이션 값 등의 파라미터들은 FP 형식일 수 있다. 또한, FP 특징부(131)의 FP 특징부 파라미터는 학습이 완료되어 불변할 수 있고, FP 기능부(132)의 FP 기능부 파라미터는 학습이 되지 않은 상태로 랜덤하게 초기화된 상태일 수 있다. 이와 같은 상태가 본 발명의 방법 수행에 어떠한 영향을 미치는지는 추후 설명될 것이다. 그리고, 도면에서와 같이, 이하 변동성 FP 뉴럴넷(130)은 수식에서 f 로, FP 특징부(131)는 h_{ζ} 로, FP 기능부(132)는 g_{θ} 로 표기될 것이다.

[0064] 도 2b를 참조하면, 바이너리 뉴럴넷(140)은 바이너리 특징부(141) 및 바이너리 기능부(142)를 포함함을 알 수

있다. 여기서 바이너리 특징부(141)는, 컨벌루션 연산 등을 사용하여 입력으로부터 특징 값, 즉 바이너리 특징 벡터를 추출하는 기능을 수행할 수 있으며, 내부적인 구성은 일반적인 뉴럴 네트워크들의 해당 부분과 동일 또는 유사할 수 있다. 또한, 바이너리 기능부(142)는, 바이너리 특징부(141)를 통해 획득된 바이너리 특징벡터를 사용하여 바이너리 뉴럴넷(140)의 구성 목적에 따른 기능을 수행할 수 있는데, 예를 들어 이미지 분류, 이미지 세그멘테이션, 패턴 인식, 제어 등의 다양한 기능들 중 하나일 수 있으며, 이의 내부 구조는 일반적으로 알려진 뉴럴 네트워크가 해당 기능을 수행하는 경우의 해당 기능부와 동일 또는 유사할 수 있다. 여기서 바이너리 기능부(142)가 수행하는 기능은 FP 기능부(132)가 수행하는 것과 동일할 것이다. 이와 같은 바이너리 특징부(141) 및 바이너리 기능부(142)의 웨이트 및 액티베이션 값 등의 파라미터들은 바이너리 형식일 수 있는데, 이하 설명하겠지만, 실시예에 따라 액티베이션 값을 먼저 바이너리화한 다음 차후에 웨이트 값을 바이너리화하는 방식 등으로 시간 순서를 달리하여 구현될 수도 있다. 이와 반대로, 웨이트 값을 먼저 바이너리화한 다음 액티베이션 값을 바이너리화하는 방식으로 구현될 수도 있을 것이다. 그리고, 도면에서와 같이, 이하 바이너리 뉴럴넷(140)은 수식에서 b_{ϕ} 로, 바이너리 특징부(141)는 k_{ϕ} 로, 바이너리 기능부(142)는 l_{ϕ} 로 표기될 것이다.

[0065] 이하 전술한 학습 장치(100)를 사용하여 수행되는, 본 발명의 일 실시예에 따른 SSL 방법에 대해 구체적으로 설명하도록 한다. 이를 위해 도 3을 참조한다.

[0066] 도 3은 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법을 나타낸 흐름도이다.

[0067] 도 3을 참조하면, 학습 장치(100)가, 학습용 벡터가 획득되면, 바이너리 형식의 바이너리 파라미터를 포함하는 바이너리 뉴럴넷(140)으로 하여금, 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 할 수 있다(S01-1). 그리고, 이와 병렬적으로, 학습 장치(100)가, 학습용 벡터가 획득되면, 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 FP 파라미터 중 적어도 일부가 조정되는 변동성 FP 뉴럴넷(130)으로 하여금, 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 할 수 있다(S01-2). 이후, 학습 장치(100)가, (i) 바이너리 뉴럴넷(140)으로부터 획득된 학습용 바이너리 출력 및 (ii) 변동성 FP 뉴럴넷(130)으로부터 획득된 학습용 FP 출력을 참조하여 바이너리 뉴럴넷 및 변동성 FP 뉴럴넷 간의 커플링 로스를 생성할 수 있다(S02). 그리고, 학습 장치(100)가, 커플링 로스를 사용하여 바이너리 파라미터 중 적어도 일부를 학습하면서 FP 파라미터 중 적어도 일부를 학습할 수 있다(S03).

[0068] 이하 본 발명의 SSL 방법에 대해 더욱 자세히 설명할 것인데, 이는 상술한 도 3의 프로세스 흐름을 큰 틀로 하여, 다소 단순화된 실시예부터 여러 가지 개선구성이 추가된 실시예 중 하나가 선택적으로 수행될 수 있는바, 각각에 대해 설명하기 위해 도 4를 참조하도록 한다.

[0069] 도 4는 본 발명의 일 실시예에 따른 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법의 실시예들을 단계적으로 나타낸 도면이다.

[0070] 도 4를 참조하면, 본 발명의 각각의 실시예의 구성이 ①②③④로 표시된 것을 확인할 수 있다. 여기서, ①②③④는 시간적 순서를 의미하는 것은 아니다.

[0071] 참고로, 먼저 가장 단순화되어 있으면서 본 발명의 기본적인 컨셉을 가장 잘 보여주는 실시예는 KL 발산 로스를 사용하는 ① 실시예이고, 이를 개선하여 피쳐 유사성 로스를 추가적으로 사용한 것이 ①+② 실시예, 이를 또 다시 개선하여 동적 밸런싱을 적용한 것이 ①+②+③ 실시예일 수 있다. ④ 실시예는, ①, ①+②, ①+②+③ 실시예 중 어느 것을 선택하여 병렬적으로 적용하여 성능을 향상시킬 수 있는 실시예일 수 있다. 이하 각각의 실시예에 대해, ①부터 순서대로 설명하도록 한다.

[0072] 즉, 학습 장치(100)는, 학습 데이터 세트에 포함된 학습용 벡터가 획득되면, 이를 변동성 FP 뉴럴넷(130)과 바이너리 뉴럴넷(140)에 각각 병렬적으로 입력할 수 있다. 즉, 변동성 FP 뉴럴넷(130)의 경우, 학습 장치(100)가, 변동성 FP 뉴럴넷(130)에 포함된, FP 파라미터 중 적어도 일부인 FP 특징부 파라미터가 기학습된 상태이며 불변하는 FP 특징부(131)로 하여금, 학습용 벡터에 FP 연산의 적어도 일부인 FP 특징추출 연산을 인가하여 학습용 FP 특징벡터를 생성하도록 할 수 있다. 이후, 학습 장치(100)는, 변동성 FP 뉴럴넷(130)에 포함된, FP 파라미터 중 적어도 일부인 FP 기능부 파라미터가 기학습되지 않고 랜덤하게 초기화된 상태에서 바이너리 파라미터 중 적어도 일부를 학습하는 과정을 통해 함께 학습되는 FP 기능부(132)로 하여금, 학습용 FP 특징벡터에 FP 연산의 적어도 일부인 FP 기능 연산을 인가하여 학습용 FP 출력을 생성하도록 할 수 있다. 즉, 전술한 S01-2 단계에서 지칭된 SSL 방법론에 따라 학습된 FP 파라미터 중 적어도 일부는 FP 특징부 파라미터일 수 있고, 바이너리 파라미터의 학습 과정에서 함께 학습되는 FP 파라미터 중 적어도 일부는 FP 기능부 파라미터일 수 있다.

[0073] 한편, 바이너리 뉴럴넷(140)의 경우, 이상의 변동성 FP 뉴럴넷(130)의 프로세스와 병렬적으로, 학습 장치(100)가, 바이너리 뉴럴넷(140)에 포함된, 바이너리 파라미터 중 적어도 일부인 바이너리 특징부 파라미터를 포함하는 바이너리 특징부(141)로 하여금, 학습용 벡터에 바이너리 연산의 적어도 일부인 바이너리 특징추출 연산을 인가하여 학습용 바이너리 특징벡터를 생성하도록 할 수 있다. 그리고, 학습 장치(100)가, 바이너리 뉴럴넷(140)에 포함된, 바이너리 파라미터 중 적어도 일부인 바이너리 기능부 파라미터를 포함하는 바이너리 기능부(142)로 하여금, 학습용 바이너리 특징벡터에 바이너리 연산의 적어도 일부인 바이너리 기능 연산을 인가하여 학습용 바이너리 출력을 생성하도록 할 수 있다.

[0074] 이와 같은 병렬적인 과정이 완료되면, 학습 장치(100)는, 학습용 바이너리 출력 및 학습용 FP 출력을 참조하여, 하기 수식에 따라 커플링 로스를 생성하여 학습을 수행할 수 있다.

$$[0075] \min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[\mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), b_{\phi}(x) \right) \right]$$

[0076] 상기 수식에서, θ 는 FP 기능부 파라미터를 의미하고, ϕ 는 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 학습용 벡터를 의미하며, D 는 학습용 벡터를 포함하는 학습 데이터 세트를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 FP 기능 연산을 의미하고, h_{ζ} 는 FP 특징추출 연산을 의미하며, b_{ϕ} 는 바이너리 연산을 의미할 수 있다.

[0077] 이상의 커플링 로스가 생성되면, 학습 장치(100)는 바이너리 파라미터 중 적어도 일부를 학습하면서 FP 파라미터 중 적어도 일부, 구체적으로는 FP 기능부(132)의 FP 기능부 파라미터를 학습할 수 있다. 여기서, FP 기능부 파라미터가 올바르게 학습될 수 있을지에 대한 의문이 들 수 있으나, Chen et al.의 A simple framework for contrastive learning on visual representations(2020), Chen et al.의 Big self-supervised models are strong semi-supervised learners(2020), Chen and He의 Exploring simple Siamese representation learning(2021), Goyal et al.의 Self-supervised pretraining of visual features in the wild(2021) 등의 선행 연구에서 확인할 수 있듯, FP 기능부(132)가 이미지 분류 연산을 수행하는 경우, 이와 같은 방식으로 FP 기능부(132)만을 SSL 방식으로 학습하더라도 옳은 방향으로 학습이 이루어질 수 있다. 또한, 이와 같이 FP 기능부(132)가 학습에 따라 계속 변화함으로써, 바이너리 네트워크(140)가 고정 표적이 아닌 이동 표적을 기준으로 학습을 수행하여, 더욱 나은 성능을 보일 수 있게 된다.

[0078] 이상 도 4의 ① 실시예에 대해 설명하였고, 이하 도 4의 ①+② 실시예에 대해 설명하도록 한다. ①+②는 ①의 커플링 로스를 생성할 때, KL 발산 함수만을 사용하는 것이 아니라, 피처 유사성 연산을 함께 사용하는 실시예이다. 즉, 학습 장치(100)가, (i) 학습용 바이너리 출력 및 학습용 FP 출력을 참조하여 KL 발산 로스를 생성하는 프로세스 및 (ii) 학습용 FP 특징벡터 및 학습용 바이너리 특징벡터를 참조하여 피처 유사성 로스를 생성하는 프로세스를 수행한 다음, KL 발산 로스 및 피처 유사성 로스를 참조하여 커플링 로스를 생성하여 학습을 수행할 수 있다. 이를 수식으로 표현하면 하기와 같다.

$$[0079] \min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[\left((1 - \lambda) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right) \right]$$

[0080] 상기 수식에서, θ 는 FP 기능부 파라미터를 의미하고, ϕ 는 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 학습용 벡터를 의미하며, D 는 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 FP 기능 연산을 의미하고, h_{ζ} 는 FP 특징추출 연산을 의미하며, l_{ϕ} 는 바이너리 기능 연산을 의미하고, k_{ϕ} 는 바이너리 특징추출 연산을 의미하며, λ 는 0 이상 1 이하의 기설정된 정적 밸런싱 파라미터를 의미하고, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하며, 평균 함수의 첫번째 텀(term)은 KL 발산 로스, 두 번째 텀(term)은 피처 유사성 로스를 의미할 수 있다.

[0081] 여기서 KL 발산 로스 부분은 이전에 설명한 것과 동일하므로 설명을 생략하고, 피처 유사성 로스 부분에 대해 설명하도록 한다. 피처 유사성 로스가 필요한 이유에 대해 먼저 설명하면, FP 기능부(132)의 FP 기능부 파라미터가 학습되지 않은 상태이므로, 이를 기준으로 KL 발산 로스를 계산하면 그라디언트가 과도하게 크면서도 변동

성이 크게 된다. 이는 학습의 안정성을 해치므로, 일단 FP 기능부 파라미터가 안정화될 때까지는 바이너리 특징부(141)가 FP 특징부(131)를 기준으로 먼저 학습하도록 하는 프로세스가 필요하다. 이를 위한 로스가 피처 유사성 로스로, 이는 바이너리 특징부(141)의 출력과 FP 특징부(131)의 출력 간의 차이에 비례하며, 이를 최소화하는 바이너리 특징부 파라미터를 찾음으로써 바이너리 특징부(141)가 FP 기능부(132)의 초기 급변동에도 불구하고 올바르게, 즉 FP 특징부(131)의 출력과 유사한 출력을 생성할 수 있도록 한다. 따라서 피처 유사성 연산은, 학습용 바이너리 특징벡터 및 학습용 FP 특징벡터 간의 차이를 나타낼 수 있는 연산이어야 하고, 일 실시예에 따르면, 이하의 수식에 따른 코사인 거리 연산일 수 있다.

$$\mathcal{L}_{FS}(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$$

[0082]

상기 수식에서, v_1 및 v_2 는 임의의 입력 벡터이고, $\frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$ 는 v_1 및 v_2 간의 코사인 유사도를 의미할 수 있다.

[0083]

이상 도 4의 ①+② 실시예에 대해 설명한 바, 이하 도 4의 ①+②+③ 실시예에 대해 설명하도록 한다. 전술한 ①+② 실시예에서는, 밸런싱 파라미터 λ 가 기설정되어 정적인 상태이다. 하지만, FP 기능부(132)에 의한 그래디언트 크기 및 변동성 문제는 학습 초반에 영향을 많이 끼치고, 학습 후반으로 갈수록 영향이 적다. 이에 따라 더욱 효율적인 학습을 위해 밸런싱 파라미터 λ 가 Iteration 수에 따라 변동되도록, 구체적으로는 학습 초반에는 크고, 학습 후반에는 작아지도록 동적으로 조정하는 실시예가 ①+②+③이다. 이하 바로 본 실시예의 수식에 대해 설명하도록 한다.

[0084]

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} \left[(1 - \lambda(t)) \mathcal{L}_{KL} \left(g_{\theta} \left(h_{\zeta}(x) \right), l_{\phi} \left(k_{\phi}(x) \right) \right) + \lambda(t) \mathcal{L}_{FS} \left(h_{\zeta}(x), k_{\phi}(x) \right) \right]$$

[0085]

상기 수식에서, θ 는 FP 기능부 파라미터를 의미하고, 상기 ϕ 는 상기 바이너리 파라미터를 의미하며, \mathbb{E} 는 평균 연산을 의미하고, x 는 상기 학습용 벡터를 의미하며, D 는 상기 학습용 벡터를 포함하는 학습 데이터를 의미하고, \mathcal{L}_{KL} 은 KL 발산 로스 함수를 의미하며, g_{θ} 는 FP 기능 연산을 의미하고, $h_{\zeta}(x)$ 는 FP 특징추출 연산을 의미하며, $l_{\phi}(x)$ 는 바이너리 기능 연산을 의미하고, $k_{\phi}(x)$ 는 상기 바이너리 특징추출 연산을 의미하며, $\lambda(t)$ 는 동적 밸런싱 파라미터를 의미하고, t 는 Iteration 수를 의미하며, \mathcal{L}_{FS} 는 피처 유사성 연산을 의미하고, 평균 연산 내부의 첫번째 텀(term)은 상기 KL 발산 로스, 두번째 텀(term)은 상기 피처 유사성 로스를 의미할 수 있다. 다른 부분은 전술한 것과 동일하므로 설명을 생략하고, 동적 밸런싱 파라미터 $\lambda(t)$ 가 어떻게 학습 초반에는 크고 학습 후반에는 작도록 하는지에 대해 설명하도록 한다.

[0086]

즉, 학습 장치(100)는, 하기 수식에 따라 동적 밸런싱 파라미터를 계산하되,

$$\lambda(t) = \lambda_{T_{max}} - (\lambda_{T_{max}} - \lambda_0) \cdot \frac{\cos\left(\frac{\pi t}{T_{max}}\right) + 1}{2}$$

[0088]

상기 수식에서, T_{max} 는 기설정된 최대 Iteration 수를 의미하고, $\lambda_{T_{max}}$ 는 0 이상 1 이하의 최대 Iteration 시의 기설정된 밸런싱 파라미터를 의미하며, λ_0 는 0 이상 1 이하의 최초 Iteration 시의 기설정된 밸런싱 파라미터를 의미할 수 있다. 여기서 $\lambda_{T_{max}}$ 가 λ_0 에 비해 작게 설정되면, Iteration이 진행될수록 $\lambda(t)$ 가 작아짐은 명백하다.

[0089]

이상 도 4의 ①+②+③ 실시예에 대해 설명한 바, 이하 도 4의 ④ 실시예인 멀티스테이지 학습 프로세스에 대해 설명할 것인데, 전술한 바와 같이 ①, ①+②, ①+②+③ 실시예 중 어느 것을 선택하여 병렬적으로 적용할 수 있는 것이나, 이하 설명에서는 ① 실시예에 적용한 경우를 상정하여 설명하도록 한다. 여기서 ①, ①+②, ①+②+

[0090]

③ 실시예 모두 큰 구조는 비슷하므로, ① 실시예에 적용한 경우만을 상정하여 설명하여도 무방할 것이다.

[0091] 멀티스테이지 학습 프로세스는 Bulat et al.의 Bats: Binary architecture search(2020), Liu et al.의 Reactnet: Towards precise binary neural network with generalized activation functions(2020) 및 Martinez et al.의 Training binary neural networks with real-to-binary convolutions(2020) 등에서 제안된, 바이너리 뉴럴 네트워크에 적용할 수 있는 효과적인 학습 방식으로, 골자는, 액티베이션 값만 바이너리화하고 웨이트 값은 바이너리화하지 않은 상태에서 1차적인 학습을 진행하고, 이후 바이너리화하지 않은 상태에서 학습된 웨이트 값을 바이너리화하여 전체가 바이너리화된 뉴럴 네트워크를 획득한 다음 다시 2차 학습을 진행한다. 이와 같은 멀티스테이지 학습 프로세스는, 2차 학습의 초기값이 최적화 값에 가까운 상태라 수렴이 빨리 일어난다는 장점이 있다. 하지만 이와 같은 종래 기술에 따른 멀티스테이지 학습 프로세스를 직접적으로 본 발명의 실시예들에 적용할 수는 없는데, 이는 2차 학습에서 FP 기능부(132)에 비해 바이너리 뉴럴넷(140)의 수렴 속도가 너무 빨라, 바이너리 뉴럴넷(140)이 최적화 상태에 도달하지 못한 상태에서 학습이 종료될 수 있기 때문이다. 이와 같은 문제점을 해결하기 위해, 본 발명의 실시예들에 멀티스테이지 학습 프로세스를 적용할 때에는, FP 기능부 파라미터의 경우에도 1차 학습에서 획득된 값을 2차 학습에서 그대로 초기값으로 사용하여, FP 기능부(132)와 바이너리 뉴럴넷(140)의 수렴 속도를 유사하게 조정할 수 있다.

[0092] 즉, 학습 장치(100)는, 도 3의 S01-1 및 S01-2 이전에, 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 바이너리 파라미터를 획득할 수 있다. 여기서 액티베이션 값 및 웨이트 값은 랜덤하게 초기화된 값일 수 있고, 또는 변동성 FP 뉴럴넷(130)의 대응되는 값들과 동일할 수도 있다. 이후, S01 내지 S03이 수행됨으로써 1차 학습이 완료되면, 학습 장치(100)가, (i) 커플링 로스를 사용하여 학습된 바이너리 뉴럴넷(140)의 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) FP 파라미터 중 적어도 일부가 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행할 수 있다. 이상의 과정은 조정 바이너리 뉴럴넷 및 조정 변동성 FP 뉴럴넷을 사용했다는 점 외에는 전술한 도 3의 S01-1 및 S01-2와 유사하게 수행될 수 있다. 이후, 학습 장치(100)가, (i) 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 제2 커플링 로스를 사용하여 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습할 수 있다. 이 경우에도, 1차 학습되었던, 조정 변동성 FP 뉴럴넷의 FP 기능부 파라미터가 조정 바이너리 파라미터 중 적어도 일부와 함께 학습될 수 있으며, 도 3의 S02 및 S03과 유사하게 수행될 수 있다. 이상의 ④ 실시예를 적용할 경우, 최종적으로 학습을 통해 획득되는 뉴럴넷은 조정 바이너리 뉴럴넷일 것이고, ①, ①+②, ①+②+③ 실시예를 적용할 경우, 최종적으로 학습을 통해 획득되는 뉴럴넷은 바이너리 뉴럴넷(140)일 수 있다.

[0093] 이상 설명한 바에 따라 획득된 바이너리 뉴럴넷(140)의 테스트 방법에 대해 이하 설명하도록 한다. ④ 실시예를 적용한 경우에 획득되는 조정 바이너리 뉴럴넷의 경우에도 이하의 테스트 방법이 동일하게 적용될 수 있음은 자명할 것이다.

[0094] 즉, (1) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 바이너리 뉴럴넷으로 하여금, 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 적어도 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 FP 파라미터 중 적어도 일부가 조정되는 변동성 FP 뉴럴넷으로 하여금, 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (2) (i) 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 바이너리 뉴럴넷 및 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 커플링 로스를 사용하여 바이너리 파라미터 중 적어도 일부를 학습하면서 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행함으로써 학습이 완료된 상태에서, 테스트용 벡터가 획득되면, 테스트 장치가, 바이너리 뉴럴넷으로 하여금, 바이너리 뉴럴넷에 적어도 하나의 바이너리 연산을 인가하도록 할 수 있다. 그리고, 테스트 장치가, 바이너리 뉴럴넷으로부터 획득된, 테스트용 바이너리 출력을 사용자에게 제공할 수 있다. 이하, 이상의 각 실시예를 적용한 경우의 테스트 시의 성능에 대해 설명하도록 한다.

[0095] 도 5는 본 발명의 변동성 FP 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 SSL 방법의 각 실시예를 적용함에 따른 테스트 시의 정확도의 변화를 나타낸 그래프이다.

[0096] 도 5를 참조하면, FP 기능부(132)가 이미지 분류를 수행하는 경우의 ImageNet Linear Evaluation Accuracy 값

들을 확인할 수 있다. ① 실시예만으로도 57.33%의 상당한 정확도를 보이고, ①+②, ①+②+③ 및 ①+②+③+④로 추가적인 구성을 부가한 실시예일수록 더 나은 정확도를 보임을 확인할 수 있다.

[0097] 이하 ①+②+③+④ 실시예를 선택한 경우의 전체 알고리즘을 소개하고 마무리하도록 한다. 이하의 알고리즘에서 사용된 문자들과 함수들은 모두 이상 설명한 그것들과 동일하므로, 추가적인 설명의 부가 없이도 통상의 기술자가 용이하게 이하의 알고리즘을 이해할 것으로 본다.

Algorithm 1 Binary Unsupervised Representation learning (BURN)

```

1: function BURN( $\mathcal{D}, t, \zeta, h_\zeta, g_\theta, k_\phi, l_\phi$ )
2:    $\theta, \phi \leftarrow \text{Pretrain}(\mathcal{D}, t, \zeta, h_\zeta, g_\theta, k_\phi, l_\phi, \text{STAGE1})$ 
3:    $W \leftarrow \{\zeta\} \cup \{\theta, \phi\}$ 
4:    $\theta, \phi \leftarrow \text{Pretrain}(\mathcal{D}, t, W, h_\zeta, g_\theta, k_\phi, l_\phi, \text{STAGE2})$ 
5:   return  $k_\phi$ 
6: end function
7: function PRETRAIN( $\mathcal{D}, t, W, h_\zeta, g_\theta, k_\phi, l_\phi, F$ )
8:   if F is STAGE1 then
9:      $k_\phi, l_\phi \leftarrow \text{Binarize Activations}$ 
10:     $h_\zeta \leftarrow W$  ▷ Load pretrained weights
11:   else
12:     $k_\phi, l_\phi \leftarrow \text{Binarize Activations and Weights}$ 
13:     $h_\zeta, g_\theta, k_\phi, l_\phi \leftarrow W$  ▷ Load pretrained weights
14:   end if
15:    $x = \text{RandomSelect}(\mathcal{D})$  ▷ Sample  $x \sim \mathcal{D}$ 
16:    $v_1, v_2 = h_\zeta(x), k_\phi(x)$  ▷ Feature vectors  $v_1, v_2$ 
17:    $p_1, p_2 = g_\theta(v_1), l_\phi(v_2)$  ▷ Softmax Probabilities  $p_1, p_2$ 
18:    $\mathcal{L}_{\zeta, \theta, \phi} = \text{AugmentedLoss}(v_1, v_2, p_1, p_2, t)$ 
19:    $\theta \leftarrow \text{Optimizer}(\nabla_\theta \mathcal{L}_{\zeta, \theta, \phi}, \eta)$  ▷ Update  $\theta$ 
20:    $\phi \leftarrow \text{Optimizer}(\nabla_\phi \mathcal{L}_{\zeta, \theta, \phi}, \eta)$  ▷ Update  $\phi$ 
21:   return  $\theta, \phi$ 
22: end function
23: function AUGMENTEDLOSS( $v_1, v_2, p_1, p_2, t$ )
24:    $\mathcal{L}_{KL} = \mathcal{D}_{KL}(p_2 \| p_1)$  ▷ KL Divergence
25:    $\mathcal{L}_{FS} = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|_2 \cdot \|v_2\|_2}$  ▷ Cosine Distance
26:    $\lambda(t) = \lambda_T - (\lambda_T - \lambda_0) \cdot (\cos(\pi t/T) + 1)/2$  ▷ Eq. 3
27:    $\mathcal{L} = (1 - \lambda(t)) \cdot \mathcal{L}_{KL} + \lambda(t) \cdot \mathcal{L}_{aug}$  ▷ Eq. 4
28:   return  $\mathcal{L}$ 
29: end function

```

[0098]

[0099] 이상 설명된 본 발명에 따른 실시예들은 다양한 컴퓨터 구성요소를 통하여 수행될 수 있는 프로그램 명령어의 형태로 구현되어 컴퓨터 판독 가능한 기록 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능한 기록 매체는 프로그램 명령어, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 컴퓨터 판독 가능한 기록 매체에 기록되는 프로그램 명령어는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능한 기록 매체의 예에는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 ROM, RAM, 플래시 메모리 등과 같은 프로그램 명령어를 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령어의 예에는, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함된다. 상기 하드웨어 장치는 본 발명에 따른 처리를 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

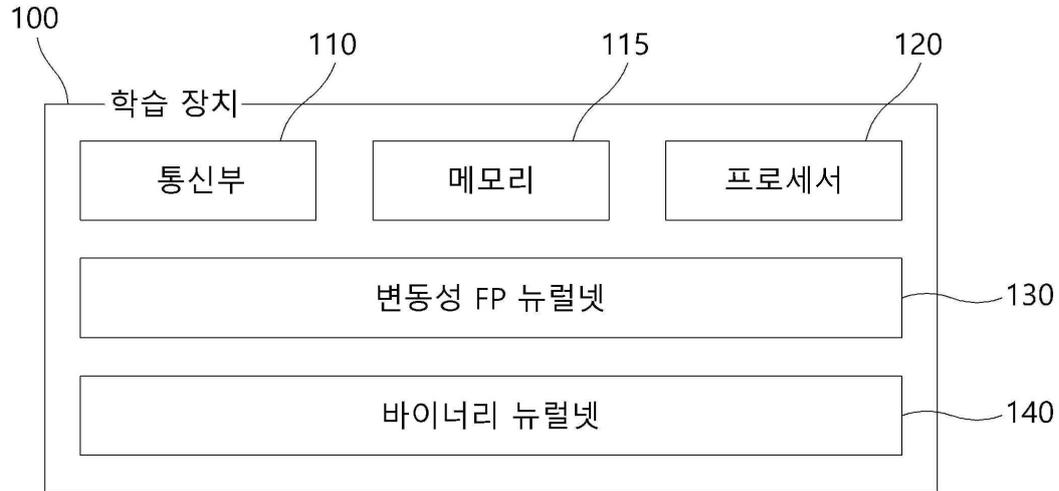
[0100] 이상에서 본 발명이 구체적인 구성요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나, 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명이 상기 실시예들에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형을 꾀할 수 있다.

[0101]

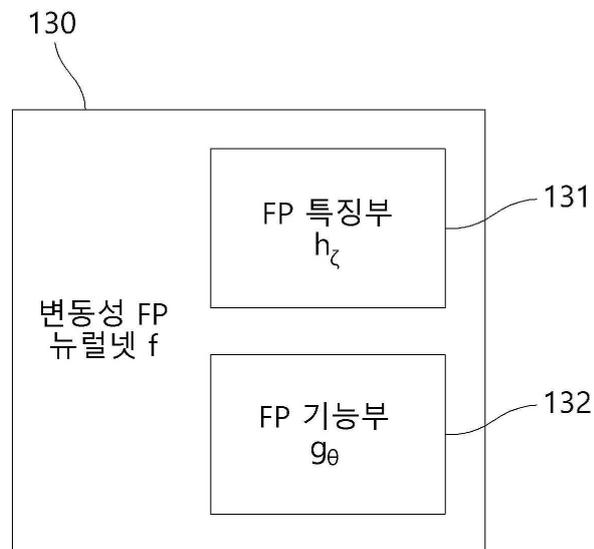
따라서, 본 발명의 사상은 상기 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위 뿐만 아니라 이 특허청구범위와 균등하게 또는 등가적으로 변형된 모든 것들은 본 발명의 사상의 범주에 속한다고 할 것이다.

도면

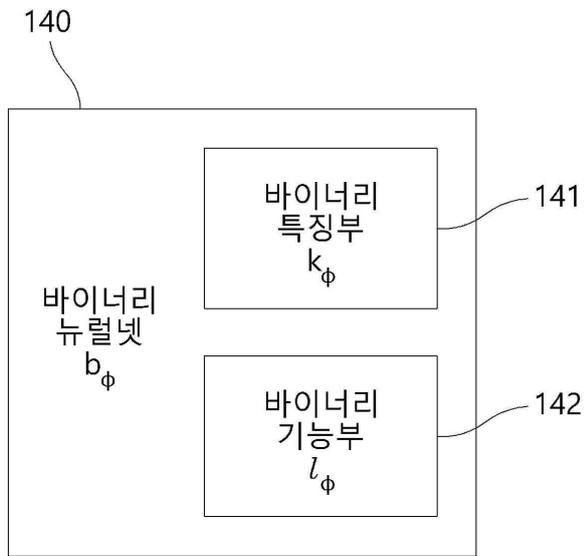
도면1



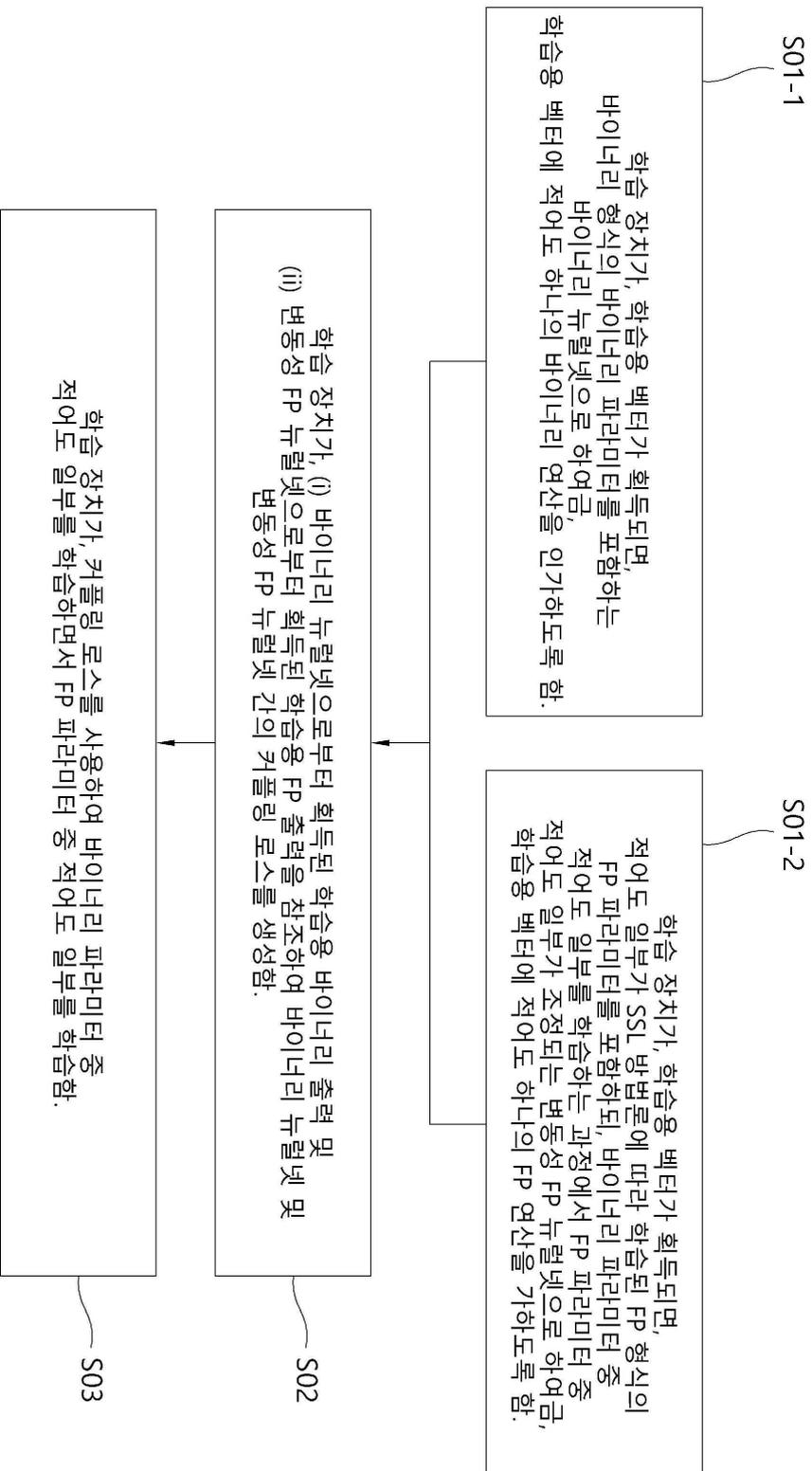
도면2a



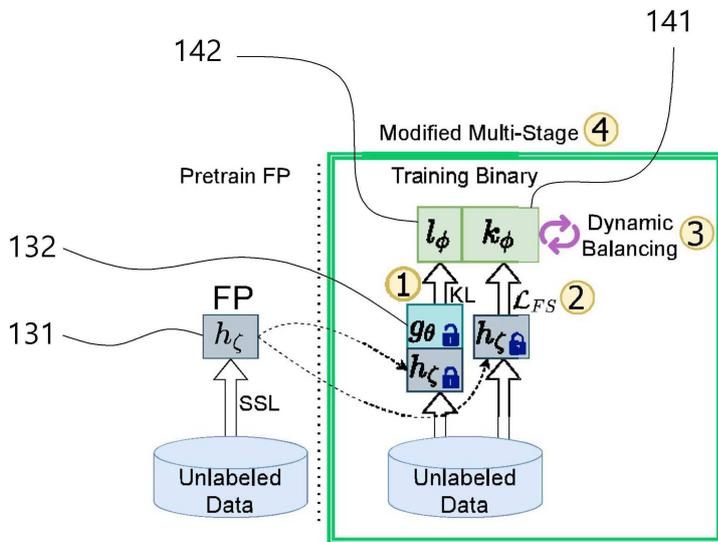
도면2b



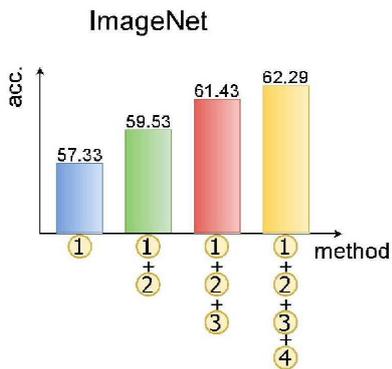
도면3



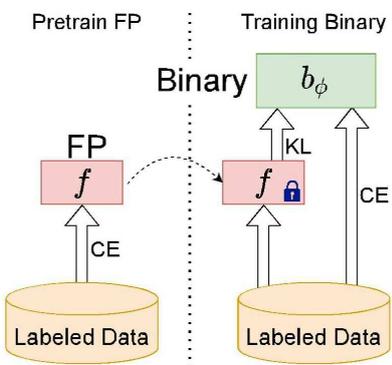
도면4



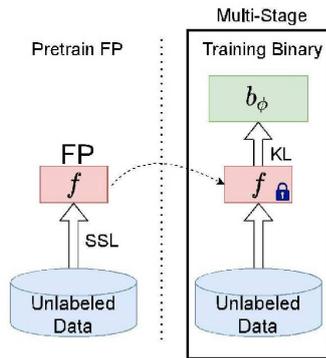
도면5



도면6a



도면6b



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 11

【변경전】

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 방법에 있어서,

(a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 단계

를 포함하되,

멀티스테이지 학습 프로세스를 적용하는 경우,

상기 (a) 단계 이전에,

(a0) 상기 학습 장치가, 상기 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 단계;

를 더 포함하고,

(c) 상기 학습 장치가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(d) 상기 학습 장치가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 단계

를 더 포함하는 것을 특징으로 하는 방법.

【변경후】

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 방법에 있어서,

(a) 학습 장치가, 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(b) 상기 학습 장치가, (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 단계

를 포함하되,

멀티스테이지 학습 프로세스를 적용하는 경우,

상기 (a) 단계 이전에,

(a0) 상기 학습 장치가, 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 단계;

를 더 포함하고,

(c) 상기 학습 장치가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 단계; 및

(d) 상기 학습 장치가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 단계

를 더 포함하는 것을 특징으로 하는 방법.

【직권보정 2】

【보정항목】 청구범위

【보정세부항목】 청구항 23

【변경전】

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 장치에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식인 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (II) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를

수행하되,

멀티스테이지 학습 프로세스를 적용하는 경우,

상기 (I) 프로세스 이전에,

(10) 상기 프로세서가, 상기 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 프로세스;

를 더 수행하고,

(III) 상기 프로세서가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (IV) 상기 프로세서가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 프로세스를 더 수행하는 것을 특징으로 하는 장치.

【변경후】

변동성 플로팅 포인트(Floating Point, FP) 뉴럴넷을 이동 표적으로 사용하여 바이너리 뉴럴넷을 학습하는 자기 지도학습(Self-Supervised Learning, SSL) 장치에 있어서,

인스트럭션들을 저장하는 하나 이상의 메모리; 및

상기 인스트럭션들을 수행하도록 설정된 하나 이상의 프로세서를 포함하되, 상기 프로세서는, (I) 학습용 벡터가 획득되면, (i) 적어도 일부가 바이너리 형식의 바이너리 파라미터를 포함하는 상기 바이너리 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 다른 일부가 SSL 방법론에 따라 학습된 FP 형식의 FP 파라미터를 포함하되, 상기 바이너리 파라미터 중 적어도 일부를 학습하는 과정에서 상기 FP 파라미터 중 적어도 일부가 조정되는 상기 변동성 FP 뉴럴넷으로 하여금, 상기 학습용 벡터에 적어도 하나의 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (II) (i) 상기 바이너리 뉴럴넷으로부터 획득된 학습용 바이너리 출력 및 (ii) 상기 변동성 FP 뉴럴넷으로부터 획득된 학습용 FP 출력을 참조하여 상기 바이너리 뉴럴넷 및 상기 변동성 FP 뉴럴넷 간의 커플링 로스를 생성한 후, 상기 커플링 로스를 사용하여 상기 바이너리 파라미터 중 적어도 일부를 학습하면서 상기 FP 파라미터 중 적어도 일부를 학습하는 프로세스를 수행하되,

멀티스테이지 학습 프로세스를 적용하는 경우,

상기 (I) 프로세스 이전에,

(10) 상기 프로세서가, 바이너리화된 액티베이션 값 및 바이너리화되지 않은 웨이트 값을 포함하는 상기 바이너리 파라미터를 획득하는 프로세스;

를 더 수행하고,

(III) 상기 프로세서가, (i) 상기 커플링 로스를 사용하여 학습된 상기 바이너리 뉴럴넷의 상기 웨이트 값을 바이너리화함으로써 획득된 조정 바이너리 뉴럴넷으로 하여금, 제2 학습용 벡터에 적어도 하나의 조정 바이너리 연산을 인가하도록 하는 프로세스 및 (ii) 상기 변동성 FP 뉴럴넷의 상기 FP 파라미터 중 적어도 일부가 상기 커플링 로스를 사용하여 학습된 상태인 조정 변동성 FP 뉴럴넷으로 하여금, 상기 제2 학습용 벡터에 적어도 하나의 조정 FP 연산을 가하도록 하는 프로세스를 수행하는 프로세스; 및 (IV) 상기 프로세서가, (i) 상기 조정 바이너리 뉴럴넷으로부터 획득된 제2 학습용 바이너리 출력 및 (ii) 상기 조정 변동성 FP 뉴럴넷으로부터 획득된 제2 학습용 FP 출력을 참조하여 제2 커플링 로스를 생성한 후, 상기 제2 커플링 로스를 사용하여 상기 조정 바이너리 뉴럴넷의 조정 바이너리 파라미터 중 적어도 일부를 학습함으로써 멀티스테이지 학습 프로세스를 완료하는 프로세스를 더 수행하는 것을 특징으로 하는 장치.