(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0079328
(43) 공개일자 2025년06월04일

(51) 국제특허분류(Int. Cl.)
G06N 3/0985 (2023.01) *G06N 3/04* (2023.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 3/0985 (2023.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2023-0166387
(22) 출원일자 2023년11월27일
심사청구일자 2023년11월27일

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
이슬기
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
김보성
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
(74) 대리인
특허법인 수

전체 청구항 수 : 총 27 항

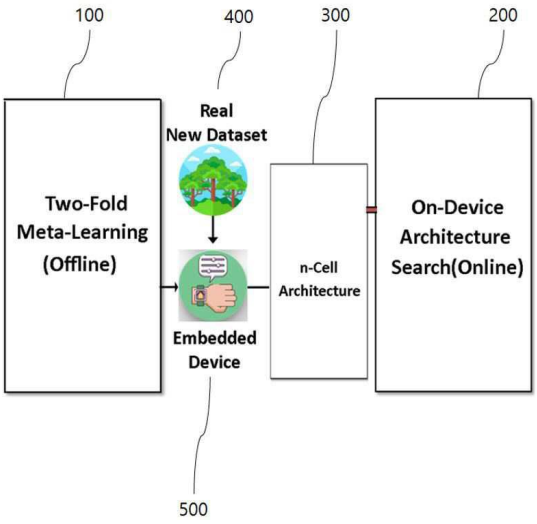
(54) 발명의 명칭 **임베디드 디바이스 학습 시스템 및 그 방법**

(57) 요 약

본 발명은 가능한 연산들의 가중치를 연속적으로 최적화하여 가장 알맞은 모델 구조를 찾고, 연산 혹은 입력값의 일부만 사용하는 학습을 토대로 요구 메모리량을 축소할 수 있는 임베디드 디바이스 학습 시스템 및 그 방법에 관한 것이다.

본 발명의 임베디드 디바이스 학습 시스템은 임베디드 디바이스에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습부 및 투폴드 메타학습부의 결과를 토대로 임베디드 디바이스 내에 탑재되어 새로운 실제 데이터의 학습을 수행하는 디바이스 내 아키텍쳐 탐색부로 이루어진다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/084 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193983
과제번호 2020-0-01336-004
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 인공지능대학원지원(울산과학기술원)
기 여 율 50.00/100
과제수행기관명 울산과학기술원
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711199491
과제번호 00277383
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 야생에서 학습하기: 효율적인 온-디바이스 트레이닝을 위해 무엇을, 언제, 어떻게 학습해야 하는가
기 여 율 50.00/100
과제수행기관명 울산과학기술원
연구기간 2023.09.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

임베디드 디바이스에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습부; 및

상기 투폴드 메타학습부의 결과를 토대로 상기 임베디드 디바이스 내에 탑재되어 새로운 실제 데이터의 학습을 수행하는 디바이스 내 아키텍쳐 탐색부;를 포함하는 임베디드 디바이스 학습 시스템.

청구항 2

제 1항에 있어서,

상기 디바이스 내 아키텍쳐 탐색부는,

학습을 수행하기 전에 N셀 아키텍쳐부에서 학습 모듈을 N개의 셀로 분해하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 3

제 2항에 있어서,

상기 투폴드 메타학습부는,

메타타스크 학습 수행을 위해 적어도 두 개 이상의 데이터셋을 구비하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 4

제 3항에 있어서,

상기 투폴드 메타학습부는,

상기 데이터셋을 토대로 메타타스크 학습을 수행하여 적어도 두 개 이상의 셀모듈을 구비하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 5

제 4항에 있어서,

상기 투폴드 메타학습부는,

상기 셀모듈을 토대로 메타셀 학습을 수행하여 하나로 병합된 메타셀을 생성하여 상기 임베디드 디바이스에 탑재하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 6

제 5항에 있어서,

상기 N셀 아키텍쳐부는,

상기 메타셀을 분해하여 상기 투폴드 메타학습부의 상기 셀모듈과 동일한 셀모듈을 복원하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 7

제 6항에 있어서,

상기 디바이스 내 아키텍쳐 탐색부는,

상기 N셀 아키텍쳐부에서 분해한 셀 중 어느 하나를 선택하는 셀모듈;을 포함하는 것을 특징으로 하는 임베디드

디바이스 학습 시스템.

청구항 8

제 7항에 있어서,

상기 디바이스 내 아키텍처 탐색부는,

상기 셀모듈의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색모듈;을 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 9

제 8항에 있어서,

상기 디바이스 내 아키텍처 탐색부는,

상기 셀모듈의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지페어 탐색모듈;을 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 10

제 9항에 있어서,

상기 디바이스 내 아키텍처 탐색부는,

단계별 역전파를 수행하는 단계적 역전파모듈;을 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 11

제 10항에 있어서,

상기 셀모듈은,

적어도 두 개 이상의 노드를 구비하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 12

제 11항에 있어서,

상기 기대값 기반 오퍼레이션 탐색모듈은,

적어도 두 개 이상의 상기 노드 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 13

제 12항에 있어서,

상기 기대값 기반 에지페어 탐색모듈은,

적어도 두 개 이상의 상기 노드의 조합에 대한 가중치 합의 기대값을 기반으로 상기 노드의 조합인 엣지쌍 탐색을 토대로 업데이트를 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 14

제 13항에 있어서,

상기 단계적 역전파모듈은,

역전파를 수행 시 하나의 상기 셀모듈 씩 업데이트 되도록 순서대로 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 시스템.

청구항 15

투폴드 메타학습부에서 임베디드 디바이스에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습단계; 및

디바이스 내 아키텍처 탐색부에서 상기 투폴드 메타학습부의 결과를 토대로 상기 임베디드 디바이스 내에 탑재되어 새로운 실제 데이터의 학습을 수행하는 디바이스 내 아키텍처 탐색단계;를 포함하는 임베디드 디바이스 학습 방법.

청구항 16

제 15항에 있어서,

상기 디바이스 내 아키텍처 탐색단계는,

학습을 수행하기 전에 N셀 아키텍처부에서 학습 모듈을 N개의 셀로 분해하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 17

제 16항에 있어서,

상기 투폴드 메타학습단계에서는,

상기 투폴드 메타학습부 내 적어도 두 개 이상의 데이터셋에 대해 두 개 이상의 셀모듈이 구비되도록 메타타스크 학습을 수행하는 메타타스크 학습단계를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 18

제 17항에 있어서,

상기 투폴드 메타학습단계에서는,

상기 셀모듈을 토대로 상기 임베디드 디바이스에 탑재할 수 있는 하나로 병합된 메타셀이 생성되도록 메타셀 학습을 수행하는 메타셀 학습단계를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 19

제 18항에 있어서,

상기 디바이스 내 아키텍처 탐색단계에서 상기 N셀 아키텍처부는 상기 메타셀을 분해하여 상기 투폴드 메타학습부의 상기 셀모듈과 동일한 N개의 셀모듈을 복원하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 20

제 19항에 있어서,

상기 디바이스 내 아키텍처 탐색단계는,

상기 셀모듈에서 상기 N셀 아키텍처부에서 분해한 셀 중 어느 하나를 선택하는 셀모듈 선택단계;를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 21

제 20항에 있어서,

상기 디바이스 내 아키텍처 탐색단계는,

기대값 기반 오퍼레이션 탐색모듈에서 상기 셀모듈의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색단계;를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 22

제 21항에 있어서,

상기 디바이스 내 아키텍처 탐색단계는,

기대값 기반 에지페어 탐색모듈에서 상기 셀모듈의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지페어 탐색단계;를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 23

제 22항에 있어서,

상기 디바이스 내 아키텍처 탐색단계는,

단계적 역전파모듈에서 단계별 역전파를 수행하는 단계적 역전파단계;를 포함하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 24

제 23항에 있어서,

상기 셀모듈 선택단계에서 셀모듈은 적어도 두 개 이상의 노드를 구비하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 25

제 24항에 있어서,

상기 기대값 기반 오퍼레이션 탐색단계에서는,

적어도 두 개 이상의 상기 노드 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 26

제 25항에 있어서,

상기 기대값 기반 에지페어 탐색단계에서는,

적어도 두 개 이상의 상기 노드의 조합에 대한 가중치 합의 기대값을 기반으로 상기 노드의 조합인 엣지쌍 탐색을 토대로 업데이트를 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

청구항 27

제 26항에 있어서,

상기 단계적 역전파단계에서는,

역전파를 수행 시 하나의 상기 셀모듈 씩 업데이트 되도록 순서대로 수행하는 것을 특징으로 하는 임베디드 디바이스 학습 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 임베디드 디바이스 학습 시스템 및 그 방법에 관한 것으로, 상세하게는, 임베디드 디바이스에서 인공지능 학습이 가능하도록 메모리량을 최소화하는 것에 관한 것이다.

[0002] 또한, 본 발명은 처리 속도가 낮은 임베디드 디바이스에 인공지능 학습이 수행 가능하도록 학습 처리량을 최소화할 수 있는 임베디드 디바이스 학습 시스템 및 그 방법에 관한 것이다.

배 경 기 술

[0003] 메타러닝(meta learning)은 배우는 방법을 배우는 작업으로서, 학습 당시와는 다른 실제의 데이터나, 새로이 맞이하게 되는 데이터에 더 쉽게 적응하는 것을 목표로 한다.

[0004] 그 중 기울기(gradient) 기반 메타러닝의 분야는, 서로 다른 데이터에 해당하는 모델의 서로 다른 여러 최적점(학습 후의 상태)에 공통적으로 가장 빠르게 도달할 수 있는 시작점을 찾는 것을 목표로 하는 알고리즘 분야이다.

[0005] 신경구조탐색(NAS; Neural Architecture Search)은 직접 설계하기에는 단순 반복 노동인 인공지능 모델 설계를

위하여 자동으로 최적의 모델 구조를 탐색하는 알고리즘으로서, 어디에 어떤 연산을 넣을지, 얼마나 반복할지, 어디와 어떻게 연결할지, 연산의 세부 설정은 무엇을 사용할지 결정하여, 최종 아키텍처(모델 구조)를 결정하는 것이다.

[0006] 그러나, 이러한 아키텍처를 만들고, 훈련하고, 시험하는 과정이 공통적으로 들어가 계산량, 메모리 요구량, 및 완료까지의 소요 시간이 오래 걸리는 것이 단점이다.

[0007] 이러한 메타러닝과 신경구조탐색을 응용하여 학습 속도를 개선하는 연구가 진행되어 왔다.

[0008] 그 일례로, 미국 공개특허공보 제2017-106114호에서는 메타 훈련에서 사용하는 신경구조탐색(NAS)을 사용하여 인공 신경세포를 자동으로 학습하기 위해 적응형 제어기를 사용함으로써, 몇 개의 샷으로 인공지능 네트워크를 구성하는 방법에 관해 개시하고 있다.

[0009] 그러나, 이 경우에도 어떤 연산을 어디에 사용할지 이산적(discrete)으로 선택하고 시험하는 반면, 가능한 연산들을 연속적(continuous)으로 최적화하지 못하는 문제점이 있다.

[0010] 또한, 임베디드 디바이스와 같이 메모리가 적고 속도가 낮은 영역에서는 사용이 불가한 문제점이 있다.

선행기술문헌

특허문헌

[0011] (특허문헌 0001) 미국 공개특허공보 제2017-106114호 (2020.11.29)

발명의 내용

해결하려는 과제

[0012] 본 발명의 목적은, 가능한 연산들의 가중치를 연속적(continuous)으로 최적화하여 가장 알맞은 모델 구조를 찾는 임베디드 디바이스 학습 시스템 및 그 방법을 제공하는 것이다.

[0013] 본 발명은 연산 혹은 입력값의 일부만 사용하는 학습을 토대로 요구 메모리량을 축소할 수 있는 임베디드 디바이스 학습 시스템 및 그 방법을 제공하는데 또 다른 목적이 있다.

과제의 해결 수단

[0014] 본 발명에 따른 임베디드 디바이스 학습 시스템은 임베디드 디바이스에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습부 및 투폴드 메타학습부의 결과를 토대로 임베디드 디바이스 내에 탑재되어 새로운 실제 데이터의 학습을 수행하는 디바이스 내 아키텍처 탐색부를 포함할 수 있다.

[0015] 여기서, 디바이스 내 아키텍처 탐색부는 학습을 수행하기 전에 N셀 아키텍처부에서 학습 모듈을 N개의 셀로 분해할 수 있다.

[0016] 또한, 투폴드 메타학습부는 메타타스크 학습 수행을 위해 적어도 두 개 이상의 데이터셋을 구비할 수 있다.

[0017] 여기서, 투폴드 메타학습부는 데이터셋을 토대로 메타타스크 학습을 수행하여 적어도 두 개 이상의 셀모듈을 구비할 수 있다.

[0018] 또한, 투폴드 메타학습부는 셀모듈을 토대로 메타셀 학습을 수행하여 하나로 병합된 메타셀을 생성하여 임베디드 디바이스에 탑재할 수 있다.

[0019] 여기서, N셀 아키텍처부는 메타셀을 분해하여 투폴드 메타학습부의 셀모듈과 동일한 셀모듈을 복원할 수 있다.

[0020] 또한, 디바이스 내 아키텍처 탐색부는 N셀 아키텍처부에서 분해한 셀 중 어느 하나를 선택하는 셀모듈을 포함할 수 있다.

[0021] 여기서, 디바이스 내 아키텍처 탐색부는 셀모듈의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색모듈을 포함할 수 있다.

[0022] 또한, 디바이스 내 아키텍처 탐색부는 셀모듈의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지

페어 탐색모듈을 포함할 수 있다.

[0023] 여기서, 디바이스 내 아키텍쳐 탐색부는 단계별 역전파를 수행하는 단계적 역전파모듈을 포함할 수 있다.

[0024] 또한, 셀모듈은 적어도 두 개 이상의 노드를 구비할 수 있다.

[0025] 여기서, 기대값 기반 오퍼레이션 탐색모듈은 적어도 두 개 이상의 노드 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행할 수 있다.

[0026] 또한, 기대값 기반 에지페어 탐색모듈은 적어도 두 개 이상의 노드의 조합에 대한 가중치 합의 기대값을 기반으로 노드의 조합인 엣지쌍 탐색을 토대로 업데이트를 수행할 수 있다.

[0027] 한편, 단계적 역전파모듈은 역전파를 수행 시 하나의 셀모듈 씩 업데이트 되도록 순서대로 수행할 수 있다.

[0028] 본 발명의 다른 실시예에 따른 임베디드 디바이스 학습 방법은 투폴드 메타학습부에서 임베디드 디바이스에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습단계 및 디바이스 내 아키텍쳐 탐색부에서 투폴드 메타학습부의 결과를 토대로 임베디드 디바이스 내에 탑재되어 새로운 실제 데이터의 학습을 수행하는 디바이스 내 아키텍쳐 탐색단계를 포함할 수 있다.

[0029] 여기서, 디바이스 내 아키텍쳐 탐색단계는 학습을 수행하기 전에 N셀 아키텍쳐부에서 학습 모듈을 N개의 셀로 분해할 수 있다.

[0030] 또한, 투폴드 메타학습단계에서는 투폴드 메타학습부 내 적어도 두 개 이상의 데이터셋에 대해 두 개 이상의 셀모듈이 구비되도록 메타타스크 학습을 수행하는 메타타스크 학습단계를 포함할 수 있다.

[0031] 여기서, 투폴드 메타학습단계에서는 셀모듈을 토대로 임베디드 디바이스에 탑재할 수 있는 하나로 병합된 메타셀이 생성되도록 메타셀 학습을 수행하는 메타셀 학습단계를 포함할 수 있다.

[0032] 또한, 디바이스 내 아키텍쳐 탐색단계에서 N셀 아키텍쳐부는 메타셀을 분해하여 투폴드 메타학습부의 셀모듈과 동일한 N개의 셀모듈을 복원할 수 있다.

[0033] 여기서, 디바이스 내 아키텍쳐 탐색단계는 셀모듈에서 N셀 아키텍쳐부에서 분해한 셀 중 어느 하나를 선택하는 셀모듈 선택단계를 포함할 수 있다.

[0034] 또한, 디바이스 내 아키텍쳐 탐색단계는 기대값 기반 오퍼레이션 탐색모듈에서 셀모듈의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색단계를 포함할 수 있다.

[0035] 여기서, 디바이스 내 아키텍쳐 탐색단계는 기대값 기반 에지페어 탐색모듈에서 셀모듈의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지페어 탐색단계를 포함할 수 있다.

[0036] 또한, 디바이스 내 아키텍쳐 탐색단계는 단계적 역전파모듈에서 단계별 역전파를 수행하는 단계적 역전파단계를 포함할 수 있다.

[0037] 여기서, 셀모듈 선택단계에서 셀모듈은 적어도 두 개 이상의 노드를 구비할 수 있다.

[0038] 또한, 기대값 기반 오퍼레이션 탐색단계에서는 적어도 두 개 이상의 노드 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행할 수 있다.

[0039] 여기서, 기대값 기반 에지페어 탐색단계에서는 적어도 두 개 이상의 노드의 조합에 대한 가중치 합의 기대값을 기반으로 노드의 조합인 엣지쌍 탐색을 토대로 업데이트를 수행할 수 있다.

[0040] 또한, 단계적 역전파단계에서는 역전파를 수행 시 하나의 셀모듈 씩 업데이트 되도록 순서대로 수행할 수 있다.

발명의 효과

[0041] 본 발명에 의한 임베디드 디바이스 학습 시스템 및 그 방법은 가능한 연산들의 가중치를 연속적(continuous)으로 최적화하여 가장 알맞은 모델 구조를 찾는 장점이 있다.

[0042] 또한, 본 발명에 의한 임베디드 디바이스 학습 시스템 및 그 방법은 연산 혹은 입력값의 일부만 사용하는 학습을 토대로 요구 메모리량을 축소할 수 있는 장점이 있다.

도면의 간단한 설명

[0043] 도 1은 본 발명의 일 실시예에 따른 임베디드 디바이스 학습 시스템을 나타낸 개략구성도이다.

도 2는 도 1의 투폴드 메타학습부를 상세히 나타낸 도면이다.

도 3은 도 1의 N셀 아키텍쳐부를 상세히 나타낸 도면이다.

도 4는 도 1의 디바이스 내 아키텍쳐 탐색부를 상세히 나타낸 도면이다.

도 5는 도 4의 셀모듈을 상세히 나타낸 도면이다.

도 6은 도 4의 기대값 기반 오퍼레이션 탐색모듈을 상세히 나타낸 도면이다.

도 7은 도 4의 기대값 기반 에지페어 탐색모듈을 상세히 나타낸 도면이다.

도 8은 도 4의 단계적 역전파모듈을 상세히 나타낸 도면이다.

도 9는 본 발명의 일 실시예에 따른 임베디드 디바이스 학습 방법을 나타낸 순서도이다.

도 10은 도 9의 투폴드 메타학습단계를 상세히 나타낸 순서도이다.

도 11은 도 9의 디바이스 내 아키텍쳐 탐색단계를 상세히 나타낸 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0044] 이하, 본 발명의 실시를 위한 구체적인 실시예를 첨부된 도면들을 참조하여 설명한다.

[0045] 본 발명을 설명함에 있어서 제 1, 제 2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되지 않을 수 있다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다.

[0046] 어떤 구성요소가 다른 구성요소에 연결되어 있다거나 접속되어 있다고 언급되는 경우는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해될 수 있다.

[0047] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다.

[0048] 본 명세서에서, 포함하다 또는 구비하다 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것으로서, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해될 수 있다.

[0049] 또한, 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0050] 이하, 첨부된 도면을 참조하여 본 발명에 따른 임베디드 디바이스 학습 시스템 및 그 방법에 대해 상세히 설명한다.

[0052] 도 1은 본 발명의 일 실시예에 따른 임베디드 디바이스 학습 시스템을 나타낸 개략구성도이며, 도 2 내지 도 8은 도 1을 상세히 설명하기 위한 세부 도면이다.

[0053] 이하, 도 1 내지 도 8을 참조하여 본 발명에 따른 임베디드 디바이스 학습 시스템을 설명한다.

[0054] 먼저, 도 1을 참조하면, 본 발명의 임베디드 디바이스 학습 시스템은 임베디드 디바이스(Embedded device)(500)에 탑재할 목적으로 메타 학습(Meta-learning)을 수행하는 투폴드 메타학습부(Two-Fold Meta-Learning)(100) 및 투폴드 메타학습부(100)의 결과를 토대로 임베디드 디바이스(500) 내에 탑재되어 새로운 실제 데이터(400)의 학습을 수행하는 디바이스 내 아키텍쳐 탐색부(On-Device Architecture Search)(200)로 이루어진다.

[0055] 즉, 본 발명에 의한 임베디드 디바이스 학습 시스템은 투폴드 메타학습부(100)에서 대량의 데이터로 학습을 수행하는 MetaNAS(Meta Neural Architecture Search)에서 시작하여, Jetson Nano (2GB~4GB)와 같이 메모리 여유량이 적은 장치인 임베디드 디바이스(500)에서 직접 구동 가능한 NAS(Neural Architecture Search) 솔루션인

디바이스 내 아키텍쳐 탐색부(200)를 생성하는 것으로서, 다음과 같은 순서로 진행된다.

[0056] 투폴드 메타학습부(100)에서는 거대한 하나의 데이터 셋(Data Set)에서 샘플링을 통해 만들어낸 작은 여러 데이터셋에 모델을 학습하되, 여러 데이터셋에 대해 메타러닝(1차)을 수행한 후, 동시에 여러 셀에 대해 하나의 셀을 다시 메타러닝(2차)을 수행하여 셀 하나만 생성하는 것이며, 도 2에서 상세히 설명한다.

[0057] 또한, 디바이스 내 아키텍쳐 탐색부(200)는 학습을 수행하기 전에 N셀 아키텍쳐부(n-Cell Architecture)(300)에서 학습 모듈을 N개의 셀로 분해하는데, 투폴드 메타학습부(100)에서 생성된 셀 하나를 임베디드 디바이스(500)에 배치 후 학습 및 활용 시 원래의 크기인 N개의 셀로 분해하여 원래의 모델로 구성하는 것이며, 도 3에서 상세히 설명한다.

[0058] 또한, 임베디드 디바이스(500) 내에서는 임베디드 디바이스(500)에서 학습 구동 가능한 사양에 맞추어, 디바이스 내 아키텍쳐 탐색부(200)에서 내부 연산 뭉치 단위 최적화(1차)와 내부 입력 페어 최적화(2차) 및 단계적 역전파(3차)를 통해 메모리 요구량을 최소화하며, 도 4 내지 도 8에서 상세히 설명한다.

[0059] 이와 같은 본 발명에 의한 임베디드 디바이스 학습 시스템은 외부와의 연결이 원활하지 않거나 간헐적인 공간에서 사용되는 모델의 학습에 활용할 수 있어, 적은 메모리 요구량으로 시간을 소비하여 지금의 환경에 적합한 모델의 구조를 찾아줄 수 있는 장점이 있다.

[0060] 특히, 기존의 On-Device Training 방법론들과 같이 모델의 학습만을 목표로 하는 것이 아닌, 모델의 구조 자체를 변형시킬 수 있도록 하여 모델 자체의 유연성을 극대화할 수 있다.

[0061] 예를 들어, 지중 전력 터널과 같은 지하 70m의 외부와의 인터넷 연결이 차단된 환경과 같은 곳에서는, 외부에서 내부의 환경 변화 (크랙, 시간에 따른 변화, 서로 다른 터널의 구조) 등에 대해 대응하여 새로운 모델을 배포하기 쉽지 않다.

[0062] 또한, 오지에 배치된 카메라나, 인터넷이 연결되지 않고 내부망만이 기동하는 군사용 시설 등에 배치된 영상 및 이미지 인식 모델 등의 경우도 마찬가지이다. 혹시, 연결이 가능하더라도, 새로운 데이터를 원래의 메인 서버에 옮기고 해당 데이터로 학습하여 다시 새로운 모델을 필요한 곳에 배치하는 과정은 높은 비용과 보안 문제를 일으킬 수 있다.

[0063] 따라서, 이상과 같은 환경에서는 기존의 학습된 모델을 그대로 사용하거나 해당 학습 지점부터 추가로 학습하는 것 보다, 본 발명에 의한 임베디드 디바이스 학습 시스템에 의해 지금의 새로운 환경에 대해 알맞은 모델을 새로이 만들어내는 것이 더욱 효과적일 수 있다.

[0065] 도 2는 도 1의 투폴드 메타학습부(100)를 상세히 나타낸 도면이다.

[0066] 도 2에서 볼 수 있는 바와 같이, 투폴드 메타학습부(100)는 메타타스크 학습(Meta Task Learning) 수행을 위해 적어도 두 개 이상의 데이터셋(110)을 구비하고, 데이터셋(110)을 토대로 메타타스크 학습을 수행하여 적어도 두 개 이상의 셀모듈(Cell module)(120)을 구비하며, 셀모듈(120)을 토대로 메타셀 학습(Meta-Cell Learning)을 수행하여 하나로 병합된 메타셀(Meta Cell)(130)을 생성하여 임베디드 디바이스(500)에 탑재한다.

[0067] 기존 메타러닝 연구에서는 여러 데이터 셋에 적응하는 하나의 모델을 찾는 것이나, 본 발명에 의한 투폴드 메타학습부(100)는 여러 위치에 적응하는 하나의 메타셀(130)을 찾는 문제로 해석하여, 여러 데이터셋에 빠르게 적응할 수 있는 모델을 만들 수 있는 하나의 셀인 메타셀(130)을 찾도록 하는 것이 특징이다.

[0068] 원래의 전체 모델의 학습 능력을 메타셀(130)로 압축했기에 발생할 수 있는 학습 속도의 저하는, 계산해 낸 별도의 계수를 기울기(gradient)에 곱해주는 방법을 나타내는 아래 수학식 1과 같이 학습을 가속화함으로써 보상해줄 수 있다.

[0069] 수학식 1은 s개의 샘플링된 데이터셋을 학습하는 동안 가중치의 총 변화량, 즉 기울기의 총 합의 크기를 구하는 식이다.

[0070] [수학식1]

$$\mathcal{G}_{T_i}^{\alpha} \triangleq \lambda_{\alpha} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial \alpha_{T_i}} = \alpha_{meta} - \alpha_{T_i} \quad , \quad \alpha_{T_i} = \alpha_{meta} - \lambda_{\alpha} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial \alpha_{T_i}}$$

$$\mathcal{G}_{T_i}^{\beta} \triangleq \lambda_{\beta} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial \beta_{T_i}} = \beta_{meta} - \beta_{T_i} \quad , \quad \beta_{T_i} = \beta_{meta} - \lambda_{\beta} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial \beta_{T_i}}$$

$$\mathcal{G}_{T_i}^{w} \triangleq \lambda_{w} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial w_{T_i}} = w_{meta} - w_{T_i} \, , \, w_{T_i} = w_{meta} - \lambda_{w} \sum_{j=1}^{s} \frac{\partial \mathcal{L}_{T_{i,j}}}{\partial w_{T_i}}$$

[0071]

[0073] 수학식 1을 통해 구해지는 G값은 두 가지 방법의 메타러닝에 맞게 활용된다.

[0074] 첫째로, 메타타스크 학습을 수행하여 데이터셋(110)을 토대로 두 개 이상의 셀모듈(120)을 구비하는 방법으로서, 각각의 태스크(샘플링된 데이터셋)에 대해, 해당하는 G값을 소프트맥스(Softmax)함수를 통해 확률로서 나타내어 모델의 기울기에 곱해줄 계수로 사용하고, 이를 통해 “더 중요”하다고 여겨지는 데이터셋(110)의 영향력이 더 크게 할 수 있도록 타스크 계수(Task Coefficient)를 아래 수학식 2로 나타낼 수 있다.

[0075] [수학식2]

$$\mathcal{T}_{T_i}^{\alpha}, \mathcal{T}_{T_i}^{\beta}, \mathcal{T}_{T_i}^{w} = \frac{\exp(\mathcal{G}_{T_i}^{\alpha})}{\sum_{j=1}^{d} \exp(\mathcal{G}_{T_j}^{\alpha})}, \frac{\exp(\mathcal{G}_{T_i}^{\beta})}{\sum_{j=1}^{d} \exp(\mathcal{G}_{T_j}^{\beta})}, \frac{\exp(\mathcal{G}_{T_i}^{w})}{\sum_{j=1}^{d} \exp(\mathcal{G}_{T_j}^{w})}$$

[0076]

[0078] 둘째로, 메타셀 학습을 수행하여 셀모듈(120)로부터 하나로 병합된 메타셀(130)을 생성하는 방법으로서, 각각의 셀모듈(120)이 데이터셋(110)을 통틀어 얼마나 변했는지 합산하고, 메타타스크 학습과는 다르게 데이터셋(110) 별이 아닌 셀모듈(120) 별로 합산하여 소프트맥스함수를 통해 확률로 나타내고, 이를 통해 “더 중요”하다고 여겨지는 셀모듈(120)의 변화를 더 크게 가져갈 수 있도록 셀 계수(Cell Coefficient)를 아래 수학식 3으로 나타낼 수 있다.

[0079] [수학식3]

$$C_{c_l}^{\alpha}, C_{c_l}^{\beta}, C_{c_l}^{w} = \frac{\exp(\mathcal{G}_{c_l}^{\alpha})}{\sum_{j=1}^{n} \exp(\mathcal{G}_{c_j}^{\alpha})}, \frac{\exp(\mathcal{G}_{c_l}^{\beta})}{\sum_{j=1}^{n} \exp(\mathcal{G}_{c_j}^{\beta})}, \frac{\exp(\mathcal{G}_{c_l}^{w})}{\sum_{j=1}^{n} \exp(\mathcal{G}_{c_j}^{w})}$$

[0080]

[0082] 이후 임베디드 디바이스(500) 내에서 원래의 N개의 개수를 가진 셀모듈(120)로 전개되어 사용할 수 있게 되는데, 도 3에서 상세히 설명한다.

[0084] 도 3은 도 1의 N셀 아키텍처부(300)를 상세히 나타낸 도면이다.

[0085] 도 3에서 알 수 있는 바와 같이, N셀 아키텍처부(300)는 메타셀(130)을 분해하여 투폴드 메타학습부(100)의 셀모듈(120)과 동일한 셀모듈(310)을 복원한다.

[0086] 여기서, 메타셀(130)은 수학식1 내지 수학식 3의 결과로 나타낸 것으로 이에 대한 역순에 의해 투폴드 메타학습부(100)의 셀모듈(120)과 갯수가 동일한 셀모듈(310)을 복원할 수 있다.

[0087] 복원된 셀모듈(310)은 각각 새로운 실제 데이터(400)를 입력으로 학습을 수행하고, 하나의 셀모듈(310) 마다 디바이스 내 아키텍쳐 탐색부(200)의 구조로 학습을 수행할 수 있으며, 도 4 내지 도 8에서 상세히 설명한다.

[0089] 도 4는 도 1의 디바이스 내 아키텍쳐 탐색부(200)를 상세히 나타낸 도면이다.

[0090] 도 4에서 볼 수 있는 바와 같이, 디바이스 내 아키텍쳐 탐색부(200)는 N셀 아키텍쳐부(300)에서 분해한 셀 중 어느 하나를 선택하는 셀모듈(210), 셀모듈(210)의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색모듈(Expectation-based Operation Search)(220), 셀모듈(210)의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지페어 탐색모듈(Expecation-based Edge Pair Search)(230), 및 단계별 역전파를 수행하는 단계적 역전파모듈(Step-By-Step Backpropagation)(240)로 이루어져, 셀모듈(210)을 최적화 및 단순화 시킬 수 있으며, 도 5 내지 도 8에서 상세히 설명한다.

[0092] 도 5는 도 4의 셀모듈(210)을 상세히 나타낸 도면이다.

[0093] 도 5에서 알 수 있는 바와 같이, 셀모듈(210)은 적어도 두 개 이상의 노드(Node)(211)를 구비한다.

[0094] 즉, 셀모듈(210) 내 다수의 노드(211)에 대해 가장 중요한 노드(211) 간 가중치 및 가장 중요한 노드(211)의 페어(pair)를 업데이트하여 단순화할 수 있다.

[0095] 여기서, 가장 중요한 노드(211) 간 가중치의 업데이트는 기대값 기반 오퍼레이션 탐색모듈(220)에서 수행하는 것으로 도 6에서 상세히 설명한다. 또한, 가장 중요한 노드(211)의 페어의 업데이트는 기대값 기반 에지페어 탐색모듈(230)에서 수행하는 것으로 도 7에서 상세히 설명한다.

[0097] 도 6은 도 4의 기대값 기반 오퍼레이션 탐색모듈(220)을 상세히 나타낸 도면이다.

[0098] 도 6에서 볼 수 있는 바와 같이, 기대값 기반 오퍼레이션 탐색모듈(220)은 적어도 두 개 이상의 노드(211) 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행한다.

[0099] 노드(211) 간의 연결은 다수개로 구성될 수 있으며, 연결된 값을 알파값(α) 이라 한다.

[0100] 이때, 노드(211) 간 연결된 개수만큼 예정된 학습을 모두 끝냈을 때, 각 연산의 알파값이 최종적으로 가지게 될 값의 기댓값을 계산하고, 해당 기댓값이 높은 연산만을 골라 정해진 개수의 연산만을 학습할 수 있다.

[0101] 한 위치에서 사용 가능한 모든 연산값이 전처리(Forward) 단계에서 사용되나, 역전파(Backward) 단계에서는 기댓값이 낮은 연산들은 모두 업데이트를 하지 않고 가장 기대값이 높은 연산만 업데이트 한다.

[0102] 따라서, 연산의 업데이트 개수를 줄임으로써, 저장 메모리를 절약할 수 있으며, 연산 속도를 낮출 수 있는 장점이 있다.

[0104] 도 7은 도 4의 기대값 기반 에지페어 탐색모듈(230)을 상세히 나타낸 도면이다.

[0105] 도 7에서 알 수 있는 바와 같이, 기대값 기반 에지페어 탐색모듈(230)은 적어도 두 개 이상의 노드(211)의 조합에 대한 가중치 합의 기대값을 기반으로 노드(211)의 조합인 엣지쌍(edge pair) 탐색을 토대로 업데이트를 수행한다.

[0106] 노드(211) 간의 엣지쌍(페어)이 가지는 가중치를 베타(β) 라 하는데, 이 베타 값 또한, 최종적으로 가지게 될 값의 기댓값을 계산해 중요도가 높을 페어의 가중치만을 학습하도록 할 수 있다.

[0107] 모든 엣지쌍(페어)이 사용되나, 가중치가 업데이트 되는 것은 기댓값이 가장 높은 엣지쌍(페어)만 해당된다.

[0108] 따라서, 엣지쌍(페어)의 업데이트 개수를 줄임으로써, 저장 메모리를 절약할 수 있으며, 요구되는 연산 속도를 낮출 수 있는 장점이 있다.

[0110] 도 8은 도 4의 단계적 역전파모듈(240)을 상세히 나타낸 도면이다.

[0111] 도 8에서 볼 수 있는 바와 같이, 단계적 역전파모듈(240)은 역전파를 수행 시 하나의 셀모듈(310) 씩 업데이트 되도록 순서대로 수행한다.

[0112] 도 6 및 도 7에서 기 언급한대로 기대값 기반 오퍼레이션 탐색모듈(220) 및 기대값 기반 에지페어 탐색모듈(230)에 의해 저장 메모리를 절약하고 연산 속도를 낮출 수 있는데, 도 8에서는 추가적으로 메모리를 더욱 절약하기 위해, N개의 셀모듈(210)이 반복되는 구성인 구조에 맞추어 단계적 역전파모듈(240)을 실행할 수 있다.

[0113] 셀모듈(210)의 개수만큼의 페이즈(phase)를 반복하며, 해당하는 페이즈의 숫자와 동일한 순서의 셀모듈(210) 1개에 대해서만 업데이트 한다.

[0114] 이로써, 이론상 N개의 셀모듈(210)을 가진 모델은 1/N의 메모리로 학습을 진행할 수 있게 되어 저장 메모리를 절약할 수 있으며, 연산 속도를 낮출 수 있는 장점이 있다.

[0116] 도 9는 본 발명의 일 실시예에 따른 임베디드 디바이스 학습 방법을 나타낸 순서도이며, 도 10 및 도 11은 도 9를 상세히 설명하기 위한 세부 순서도이다.

[0117] 이하, 도 9 내지 도 11을 참조하여 본 발명에 따른 임베디드 디바이스 학습 시스템 방법을 설명한다.

[0118] 먼저, 도 9를 참조하면, 본 발명의 임베디드 디바이스 학습 방법은 투폴드 메타학습부(100)에서 임베디드 디바이스(500)에 탑재할 목적으로 메타 학습을 수행하는 투폴드 메타학습단계(S100) 및 디바이스 내 아키텍처 탐색부(200)에서 투폴드 메타학습부(100)의 결과를 토대로 임베디드 디바이스(500) 내에 탑재되어 새로운 실제 데이터(400)의 학습을 수행하는 디바이스 내 아키텍처 탐색단계(S200)로 이루어진다.

[0119] 즉, 본 발명에 의한 임베디드 디바이스 학습 방법은 투폴드 메타학습부(100)에서 대량의 데이터로 학습을 수행하는 MetaNAS(Meta Neural Architecture Search)에서 시작하여, Jetson Nano (2GB~4GB)와 같이 메모리 여유량이 적은 장치인 임베디드 디바이스(500)에서 직접 구동 가능한 NAS(Neural Architecture Search) 솔루션인 디바이스 내 아키텍처 탐색부(200)를 생성하는 것으로서, 다음과 같은 순서로 진행된다.

[0120] 투폴드 메타학습단계(S100)에서는 거대한 하나의 데이터 셋에서 샘플링을 통해 만들어낸 작은 여러 데이터셋에 모델을 학습하되, 여러 데이터셋에 대해 메타러닝(1차)을 수행한 후, 동시에 여러 셀에 대해 하나의 셀을 다시 메타러닝(2차)을 수행하여 셀 하나만 생성하는 것으로 도 10에서 상세히 설명한다.

[0121] 한편, 디바이스 내 아키텍처 탐색단계(S200)는 학습을 수행하기 전에 N셀 아키텍처부(300)에서 학습 모듈을 N개의 셀로 분해한다. 즉, 메타셀(130)을 분해하여 투폴드 메타학습부(100)의 셀모듈(120)과 동일한 N개의 셀모듈(310)을 복원한다.

[0122] 여기서, 메타셀(130)은 상술한 수학식1 내지 수학식 3의 조합의 결과로 나타낸 것으로 이에 대한 역순에 의해 투폴드 메타학습부(100)의 셀모듈(120)과 개수가 동일한 셀모듈(310)을 복원할 수 있다.

[0123] 복원된 셀모듈(310)은 각각 새로운 실제 데이터(400)를 입력으로 학습을 수행하고, 하나의 셀모듈(310) 마다 디바이스 내 아키텍처 탐색부(200)의 구조로 학습을 수행할 수 있다.

[0124] 또한, 디바이스 내 아키텍처 탐색단계(S200)에서는 임베디드 디바이스(500)에서 학습 구동 가능한 사양에 맞출 있도록, 디바이스 내 아키텍처 탐색부(200)에서 내부 연산 뭉치 단위 최적화(1차)와 내부 입력 페어 최적화(2차) 및 단계적 역전파(3차)를 통해 메모리 요구량을 최소화하는 것으로서, 도 11에서 상세히 설명한다.

[0125] 이와 같은 본 발명에 의한 임베디드 디바이스 학습 방법은 외부와의 연결이 원활하지 않거나 간헐적인 공간에서 사용되는 모델의 학습에 활용할 수 있어 적은 메모리 요구량으로 시간을 소비하여 지금의 환경에 적합한 모델의 구조를 찾아줄 수 있는 장점이 있다.

[0127] 도 10은 도 9의 투폴드 메타학습단계(S100)를 상세히 나타낸 순서도이다.

[0128] 도 10에서 볼 수 있는 바와 같이, 투폴드 메타학습단계(S100)는 투폴드 메타학습부(100) 내 적어도 두 개 이상의 데이터셋(110)에 대해 두 개 이상의 셀모듈(120)이 구비되도록 메타타스크 학습을 수행하는 메타타스크 학습단계(S110)를 포함하고, 또한 셀모듈(120)을 토대로 임베디드 디바이스(500)에 탑재할 수 있는 하나로 병합된 메타셀(130)이 생성되도록 메타셀 학습을 수행하는 메타셀 학습단계(S120)를 포함한다.

[0129] 기존 메타러닝 연구에서는 여러 데이터 셋에 적응하는 하나의 모델을 찾는 것이나, 본 발명에 의한 투폴드 메타 학습단계(S100)는 여러 위치에 적응하는 하나의 메타셀(130)을 찾는 문제로 해석하여, 여러 데이터셋에 빠르게 적응할 수 있는 모델을 만들 수 있는 하나의 셀인 메타셀(130)을 찾도록 하는 것이 특징이다.

[0130] 원래의 전체 모델의 학습 능력을 메타셀(130)로 압축했기에 발생할 수 있는 학습 속도의 저하는, 계산해 낸 별도의 계수를 기울기(gradient)에 곱해주는 방법을 나타내는 상술한 수학식 1과 같이 학습을 가속화함으로써 보상해줄 수 있다.

[0131] 첫째로, 메타타스크 학습단계(S110)는 데이터셋(110)을 토대로 두 개 이상의 셀모듈(120)을 구비하는 방법으로서, 각각의 태스크(샘플링된 데이터셋)에 대해, 해당하는 G값을 소프트맥스(Softmax)함수를 통해 확률로서 나타내어 모델의 기울기에 곱해줄 계수로 사용하고, 이를 통해 “더 중요”하다고 여겨지는 데이터셋(110)의 영향력이 더 크게 할 수 있도록 타스크 계수(Task Coefficient)를 상술한 수학식 2로 나타낼 수 있다.

[0132] 둘째로, 메타셀 학습단계(S120)는 셀모듈(120)로부터 하나로 병합된 메타셀(130)을 생성하는 방법으로서, 각각의 셀모듈(120)이 데이터셋(110)을 통틀어 얼마나 변했는지 합산하고, 메타타스크 학습과는 다르게 데이터셋(110) 별이 아닌 셀모듈(120) 별로 합산하여 소프트맥스함수를 통해 확률로 나타내고, 이를 통해 “더 중요”하다고 여겨지는 셀모듈(120)의 변화를 더 크게 가져갈 수 있도록 셀 계수(Cell Coefficient)를 상술한 수학식 3으로 나타낼 수 있다.

[0133] 이후 임베디드 디바이스(500) 내 N셀 아키텍쳐부(300)에서 원래의 N개의 개수를 가진 셀모듈(120)로 전개되어 사용할 수 있게 된다.

[0135] 도 11은 도 9의 디바이스 내 아키텍쳐 탐색단계(S200)를 상세히 나타낸 순서도이다.

[0136] 도 11에서 알 수 있는 바와 같이, 디바이스 내 아키텍쳐 탐색단계(S200)는 셀모듈(210)에서 N셀 아키텍쳐부(300)에서 분해한 셀 중 어느 하나를 선택하는 셀모듈 선택단계(S210), 기대값 기반 오퍼레이션 탐색모듈(220)에서 셀모듈(210)의 출력에 대해 기대값 기반 오퍼레이션을 탐색하는 기대값 기반 오퍼레이션 탐색단계(S220), 기대값 기반 에지페어 탐색모듈(230)에서 셀모듈(210)의 출력에 대해 기대값 기반 에지쌍을 탐색하는 기대값 기반 에지페어 탐색단계(S230), 및 단계적 역전파모듈(240)에서 단계별 역전파를 수행하는 단계적 역전파단계(S240)로 이루어진다.

[0137] 여기서, 셀모듈 선택단계(S210)는 셀모듈(210) 내 적어도 두 개 이상의 노드(211)를 구비한다.

[0138] 즉, 디바이스 내 아키텍쳐 탐색단계(S200)에서는 셀모듈(210) 내 다수의 노드(211)에 대해 가장 중요한 노드(211) 간 가중치 및 가장 중요한 노드(211)의 페어를 업데이트하여 단순화할 수 있는데, 여기서, 가장 중요한 노드(211) 간 가중치의 업데이트는 기대값 기반 오퍼레이션 탐색단계(S220)에서 수행하고, 가장 중요한 노드(211)의 페어의 업데이트는 기대값 기반 에지페어 탐색단계(S230)에서 수행한다.

[0139] 도 6에서 이미 설명한 바와 같이, 본 발명의 기대값 기반 오퍼레이션 탐색단계(S220)에서는 기대값 기반 오퍼레이션 탐색모듈(220)에서 적어도 두 개 이상의 노드(211) 간 가중치의 기대값을 기반으로 오퍼레이션 탐색을 토대로 업데이트를 수행한다. 이때, 노드(211) 간의 연결은 다수개로 구성될 수 있으며, 연결된 값을 알파값이라 한다.

[0140] 또한, 노드(211) 간 연결된 개수만큼 예정된 학습을 모두 끝냈을 때, 각 연산의 알파값이 최종적으로 가지게 될 값의 기댓값을 계산하고, 해당 기댓값이 높은 연산만을 골라 정해진 개수의 연산만을 학습할 수 있다.

[0141] 한 위치에서 사용 가능한 모든 연산값이 전처리(Forward) 단계에서 사용되나, 역전파(Backward) 단계에서는 기댓값이 낮은 연산들은 모두 업데이트를 하지 않고 가장 기대값이 높은 연산만 업데이트 한다.

[0142] 이를 통해, 연산의 업데이트 개수를 줄임으로써, 저장 메모리를 절약할 수 있으며, 연산 속도를 낮출 수 있는 장점이 있다.

[0143] 또한, 도 7에서 이미 설명한 바와 같이, 본 발명의 기대값 기반 에지페어 탐색단계(S230)에서는 기대값 기반 에지페어 탐색모듈(230)에서 적어도 두 개 이상의 노드(211)의 조합에 대한 가중치 합의 기대값을 기반으로 노드(211)의 조합인 엣지쌍 탐색을 토대로 업데이트를 수행한다.

[0144] 이때, 노드(211) 간의 엣지쌍 (페어)이 가지는 가중치를 베타라 하는데, 이 베타 값 또한, 최종적으로 가지게

될 값의 기댓값을 계산해 중요도가 높을 페어의 가중치만을 학습하도록 할 수 있다.

[0145] 모든 엣지쌍(페어)이 사용되나, 가중치가 업데이트 되는 것은 기댓값이 가장 높은 엣지쌍(페어)만 해당된다.

[0146] 따라서, 엣지쌍(페어)의 업데이트 개수를 줄임으로써, 저장 메모리를 절약할 수 있으며, 요구되는 연산 속도를 낮출 수 있는 장점이 있다.

[0147] 한편, 도 8에서 이미 설명한 바와 같이, 본 발명의 단계적 역전파단계(S240)에서는 단계적 역전파모듈(240)에서 역전파를 수행 시 하나의 셀모듈(310) 씩 업데이트 되도록 순서대로 수행한다.

[0148] 즉, 기대값 기반 오퍼레이션 탐색단계(S220) 및 기대값 기반 에지페어 탐색단계(S230)에 의해 저장 메모리를 절약하고 요구되는 연산 속도를 낮출 수 있는데, 단계적 역전파단계(S240)에서는 추가적으로 메모리 사용을 더욱 절약하기 위해, N개의 셀모듈(210)이 반복되는 구성인 구조에 맞추어 단계적 역전파모듈(240)을 실행할 수 있다.

[0149] 셀모듈(210)의 개수만큼의 페이즈(phase)를 반복하며, 해당하는 페이즈의 숫자와 동일한 순서의 셀모듈(210) 1개에 대해서만 업데이트 한다.

[0150] 이로써, 이론상 N개의 셀모듈(210)을 가진 모델은 1/N의 메모리로 학습을 진행할 수 있게 되어 저장 메모리를 절약할 수 있으며, 연산 속도를 낮출 수 있는 장점이 있다.

[0152] 이상과 같이 본 발명에 따른 임베디드 디바이스 학습 시스템 및 그 방법은 가능한 연산들의 가중치를 연속적(continuous)으로 최적화하여 가장 알맞은 모델 구조를 찾는 장점이 있으며, 연산 혹은 입력값의 일부만 사용하는 학습을 토대로 요구 메모리량을 축소할 수 있는 장점이 있다.

[0154] 본 발명의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 소프트웨어로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 발명의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식들로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 발명의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

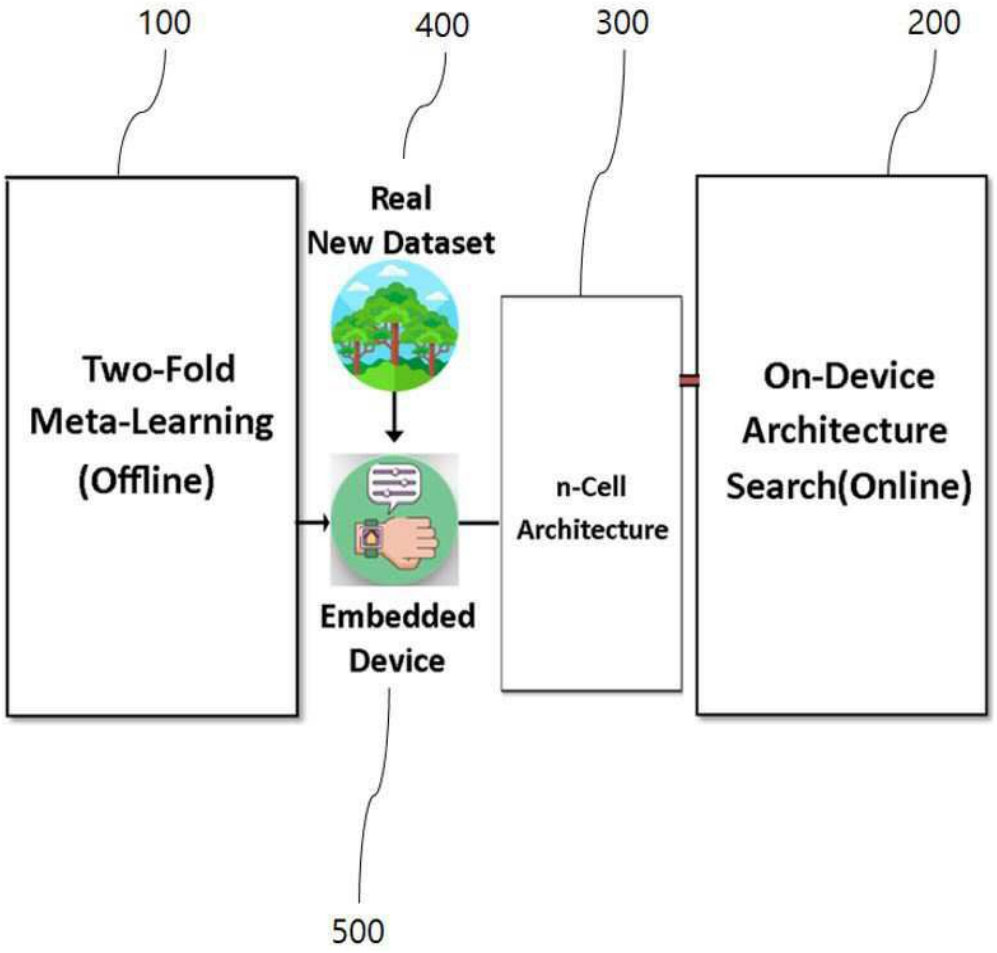
[0155] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조 물품(article)으로 구현될 수 있다. 용어 제조 물품은 임의의 컴퓨터-판독가능 저장장치로부터 액세스 가능한 컴퓨터 프로그램, 캐리어, 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트립, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되는 것은 아니다. 또한, 여기서 제시되는 다양한 저장 매체는 정보를 저장하기 위한 하나 이상의 장치 및/또는 다른 기계-판독가능한 매체를 포함한다.

[0156] 제시된 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조는 예시적인 접근들의 일례임을 이해하도록 한다. 설계 우선순위들에 기반하여, 본 발명의 범위 내에서 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조가 재배열될 수 있다는 것을 이해하도록 한다. 첨부된 방법 청구항들은 샘플 순서로 다양한 단계들의 엘리먼트들을 제공하지만 제시된 특정한 순서 또는 계층 구조에 한정되는 것을 의미하지는 않는다.

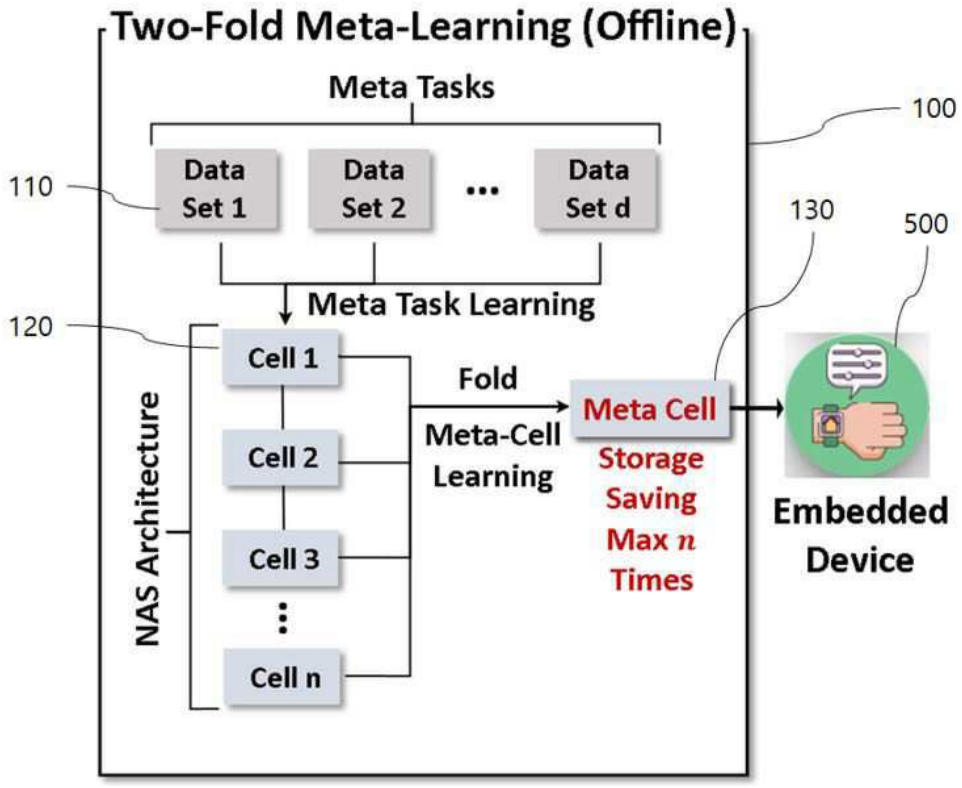
[0157] 제시된 실시예들에 대한 설명은 임의의 본 발명의 기술 분야에서 통상의 지식을 가진 자가 본 발명을 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 발명의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 발명의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 발명은 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

도면

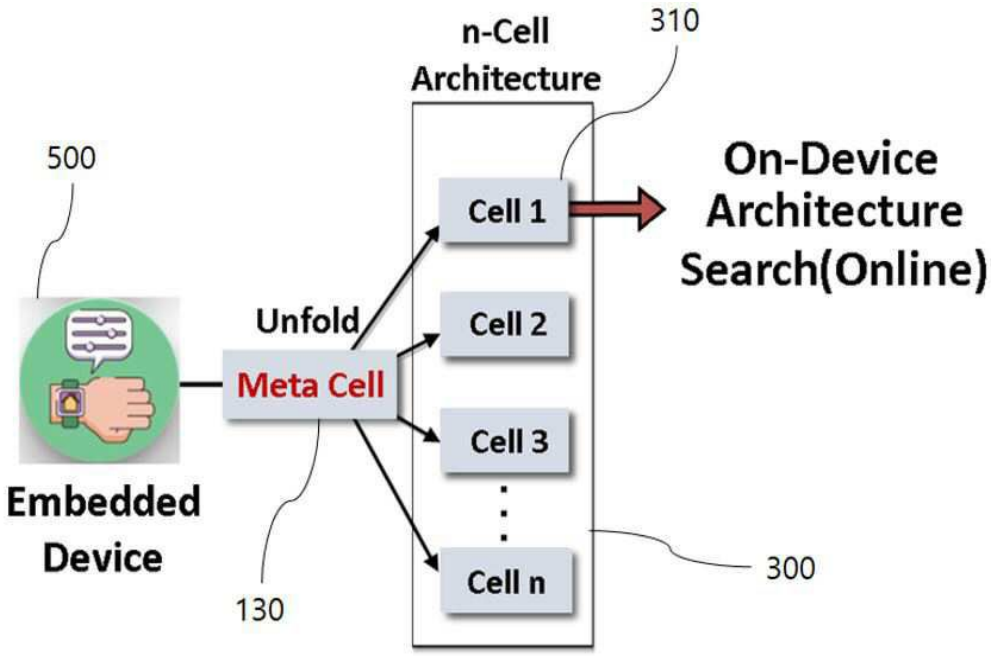
도면1


100
400
300
200
Real
New Dataset
Two-Fold
Meta-Learning
(Offline)
n-Cell
Architecture
On-Device
Architecture
Search(Online)
Embedded
Device
500

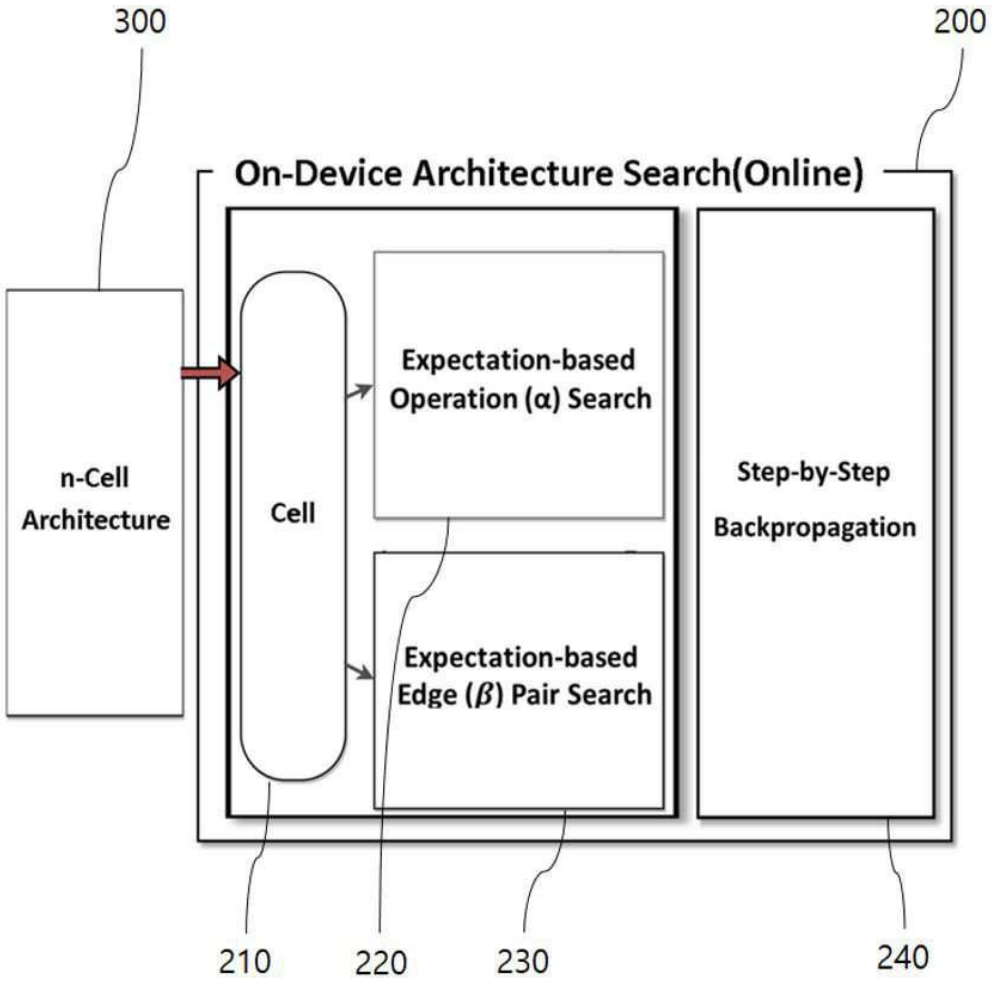
도면2


Two-Fold Meta-Learning (Offline)
Meta Tasks
100
110
Data Set 1
Data Set 2
...
Data Set d
130
500
Meta Task Learning
120
Cell 1
Cell 2
Cell 3
Cell n
NAS Architecture
Fold
Meta-Cell Learning
Meta Cell
Storage Saving Max n Times
Embedded Device

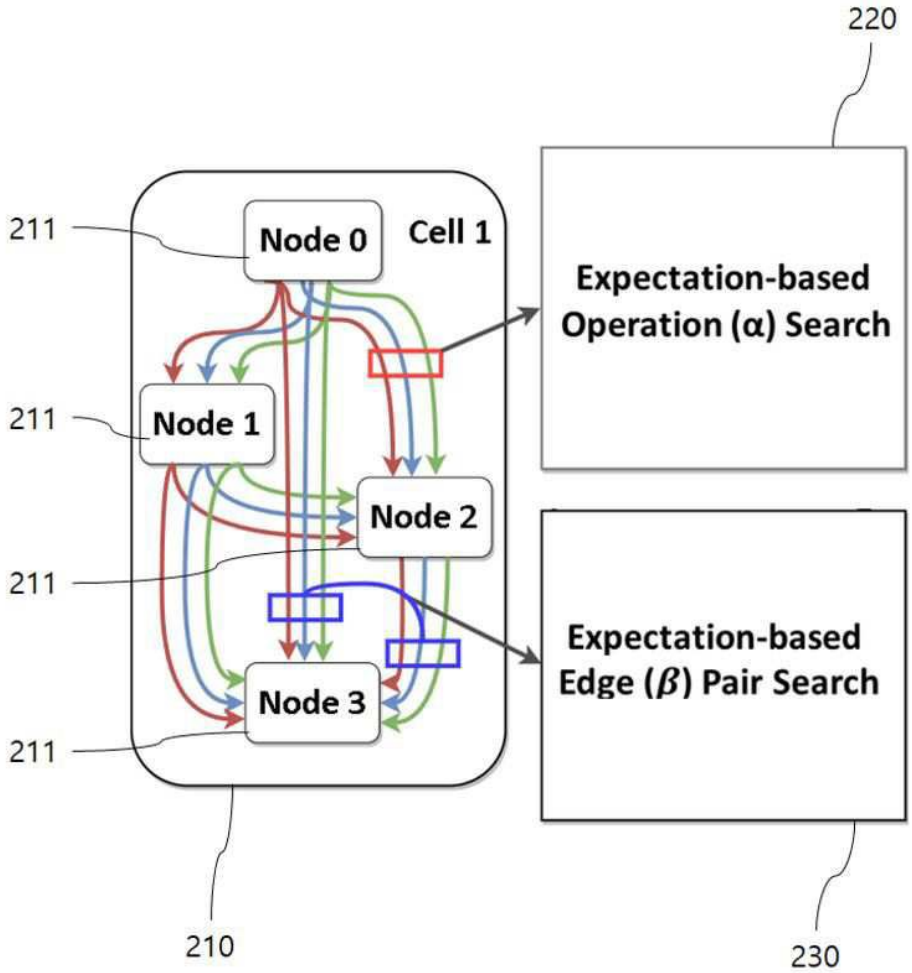
도면3


n-Cell Architecture
310
500
Cell 1
On-Device Architecture Search(Online)
Unfold
Meta Cell
Cell 2
Cell 3
Cell n
Embedded Device
130
300

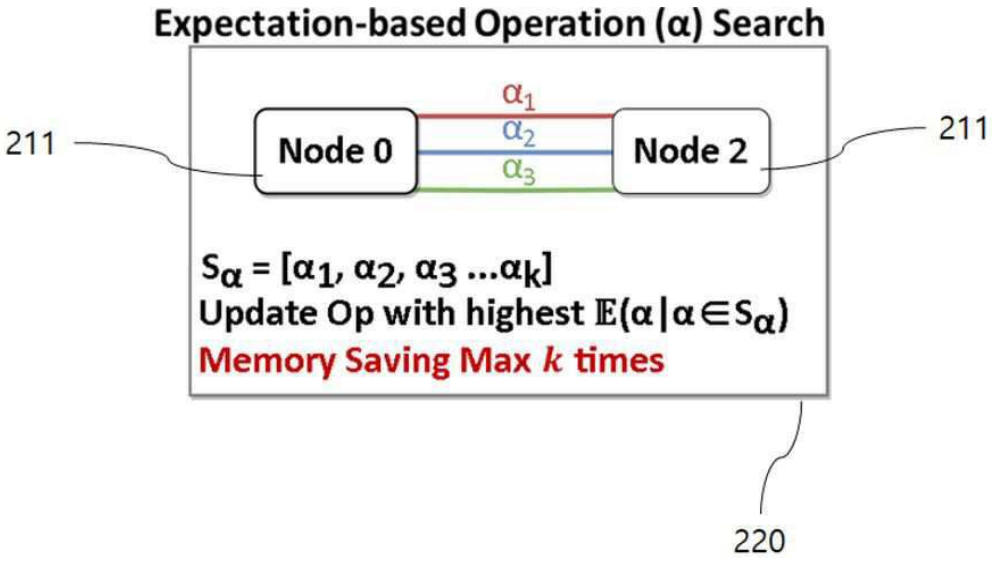
도면4


300
200
On-Device Architecture Search(Online)
n-Cell Architecture
Cell
Expectation-based Operation (α) Search
Expectation-based Edge (β) Pair Search
Step-by-Step Backpropagation
210
220
230
240

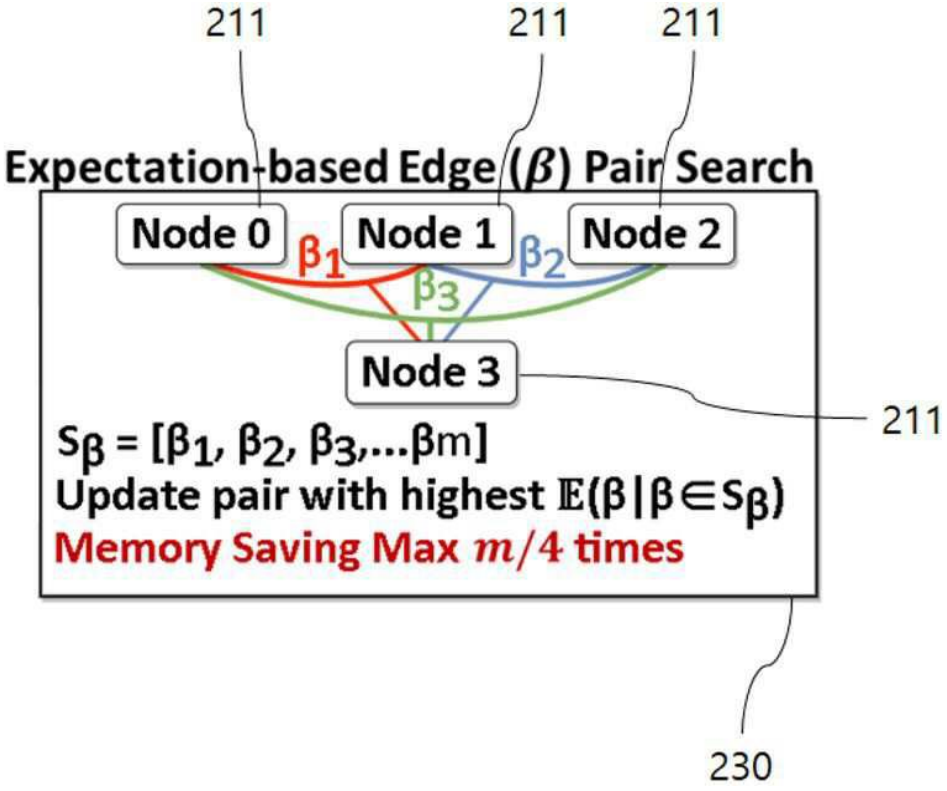
도면5


220
211
Node 0
Cell 1
211
Node 1
Node 2
211
Node 3
211
210
Expectation-based
Operation (α) Search
Expectation-based
Edge (β) Pair Search
230

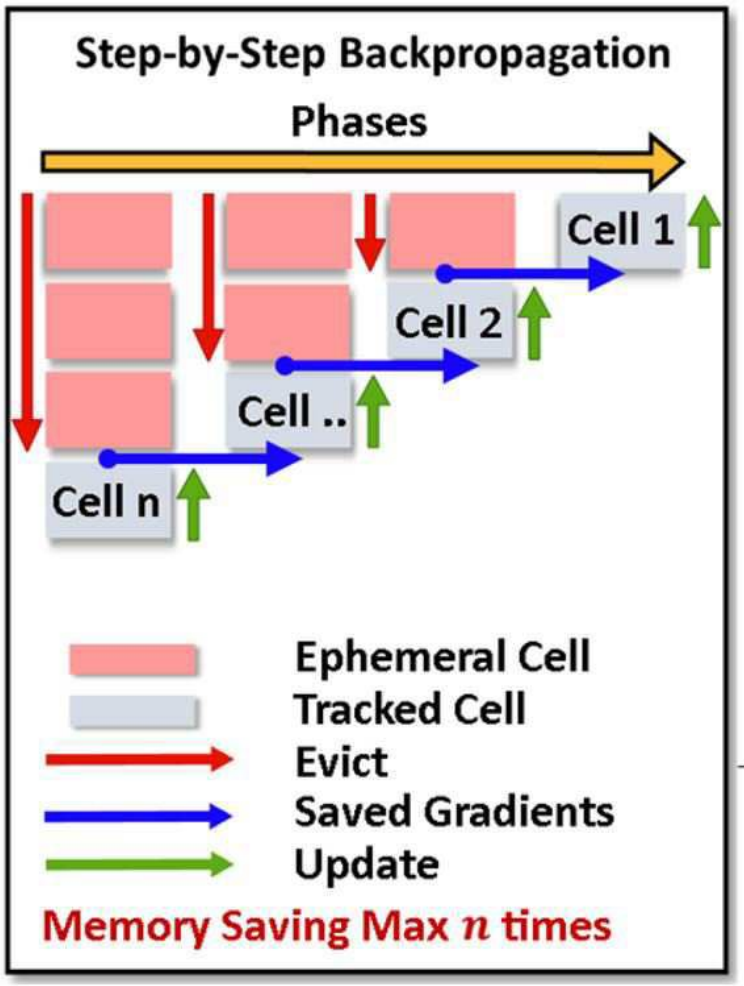
도면6


Expectation-based Operation (α) Search
211
Node 0
α1
α2
α3
Node 2
211
Sα = [α1, α2, α3 ...αk]
Update Op with highest 𝔼(α|α∈Sα)
Memory Saving Max k times
220

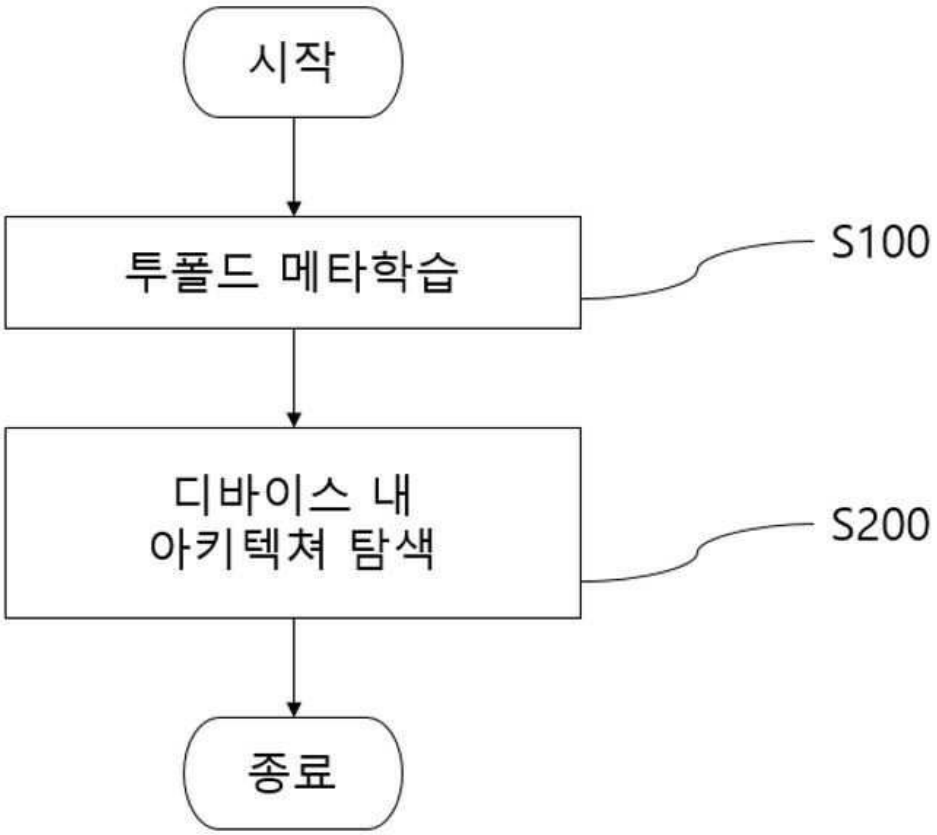
도면7


211
211
211
Expectation-based Edge (β) Pair Search
Node 0
β1
Node 1
β2
Node 2
β3
Node 3
211
Sβ = [β1, β2, β3,...βm]
Update pair with highest 𝔼(β|β∈Sβ)
Memory Saving Max m/4 times
230

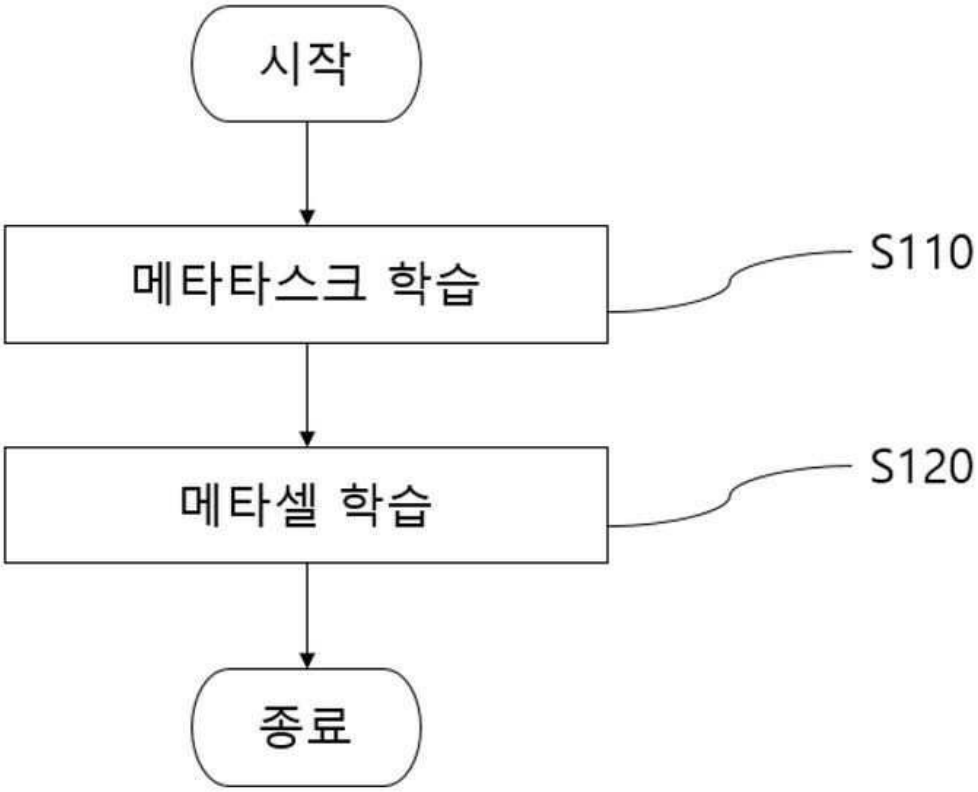
도면8


Step-by-Step Backpropagation
Phases
Cell 1
Cell 2
Cell ..
Cell n
Ephemeral Cell
Tracked Cell
Evict
Saved Gradients
Update
Memory Saving Max n times
240

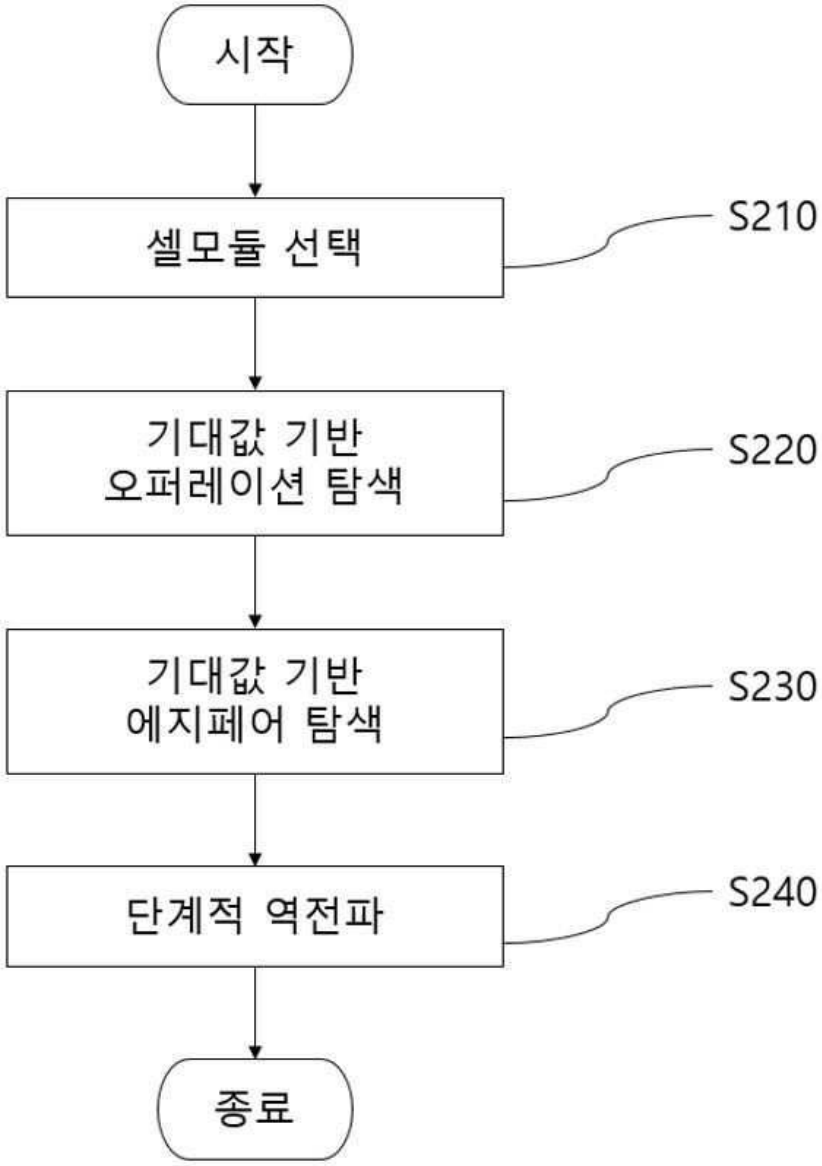
도면9


시작
투폴드 메타학습
S100
디바이스 내
아키텍쳐 탐색
S200
종료

도면10


시작
메타타스크 학습
S110
메타셀 학습
S120
종료

도면11


시작
셀모듈 선택
S210
기대값 기반
오퍼레이션 탐색
S220
기대값 기반
에지페어 탐색
S230
단계적 역전파
S240
종료