(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년01월15일
(11) 등록번호 10-2624299
(24) 등록일자 2024년01월09일

(51) 국제특허분류(Int. Cl.)
G06N 3/098 (2023.01) *G06N 3/045* (2023.01)
G06N 3/0985 (2023.01)
(52) CPC특허분류
G06N 3/098 (2023.01)
G06N 3/045 (2023.01)
(21) 출원번호 10-2023-0022972
(22) 출원일자 2023년02월21일
심사청구일자 2023년02월21일
(30) 우선권주장
17/944,650 2022년09월14일 미국(US)
(56) 선행기술조사문헌
KR1020220103247 A*
Sannara EK et al., A Federated Learning Aggregation Algorithm for Pervasive Computing: Evaluation and Comparison, 2021 IEEE International Conference on Pervasive Computing and Communications, 1-10page*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
디어젠 주식회사
대전광역시 서구 청사로 136, 18층 R1846호 (월평동)
(72) 발명자
김예찬
경기도 성남시 분당구 미금로 246, 603동 901호(금곡동, 청솔마을주공6단지아파트)
신봉근
경상북도 포항시 북구 대안길 56, 124동 801호(용흥동, 우방타운)
(74) 대리인
이대호, 박건홍

전체 청구항 수 : 총 11 항 심사관 : 권계민

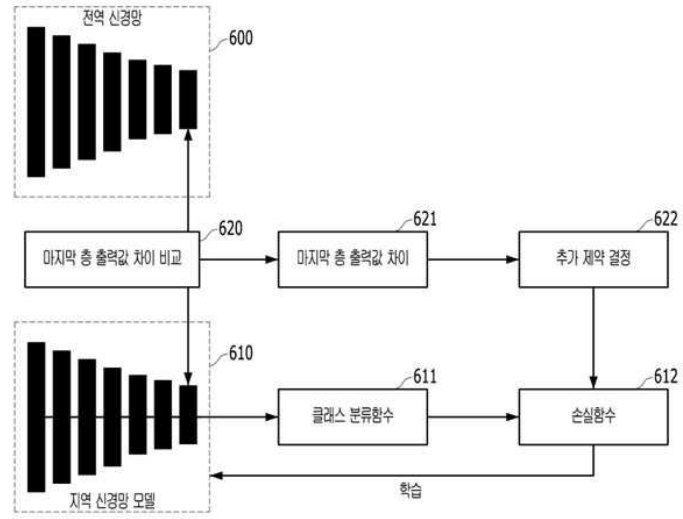
(54) 발명의 명칭 **연합 학습을 위한 지역 신경망 모델 학습 방법**

(57) 요 약

본 개시는 학습 데이터가 서로 상이한 이종환경을 고려하여, 연합 학습 기반의 지역 신경망 모델을 학습시키는 것을 해결과제로 한다.

전술한 바와 같은 과제를 실현하기 위한 본 개시의 실시예에 따라, 적어도 하나의 컴퓨팅 장치에 의해 수행되는, 연합 학습(federated learning) 기반의 지역(local) 신경망 모델을 학습시키기 위한 학습 방법으로, 전역(global) 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계; 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약(regularization)을 결정하는 단계; 및 상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키는 단계를 포함할 수 있다.

대 표 도 - 도6



(52) CPC특허분류
G06N 3/0985 (2023.01)

명 세 서

청구범위

청구항 1

적어도 하나의 컴퓨팅 장치에 의해 수행되는, 연합 학습(federated learning) 기반의 지역(local) 신경망 모델을 학습시키기 위한 학습 방법으로서,

전역(global) 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계;

상기 산출된 차이 및 제약의 강도를 조절하기 위한 하이퍼-파라미터에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약(regularization)을 결정하는 단계; 및

상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키는 단계

를 포함하고,

상기 하이퍼-파라미터는 상기 연합 학습에 관련된 데이터의 이종환경 분포에 비례하도록 결정되는,

학습 방법.

청구항 2

제 1 항에 있어서,

상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는,

상기 전역 신경망 모델의 각 층의 출력값들과 상기 지역 신경망 모델의 대응되는 각 층의 출력값들 사이의 차이들을 산출하는 단계; 및

상기 산출된 차이들 중 가장 큰 차이를 식별하는 단계;

를 포함하는,

학습 방법.

청구항 3

제 1 항에 있어서,

상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는,

상기 전역 신경망 모델의 마지막 층의 출력값과 상기 지역 신경망 모델의 마지막 층의 출력값 사이의 차이를 산출하는 단계

를 포함하는,

학습 방법.

청구항 4

제 1 항에 있어서,

상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는,

상기 전역 신경망 모델의 미리 결정된 층의 출력값과 상기 지역 신경망 모델의 미리 결정된 층의 출력값 사이의 차이를 산출하는 단계

를 포함하는,

학습 방법.

청구항 5

제 3 항에 있어서,

상기 전역 신경망 모델의 마지막 층의 출력값과 상기 지역 신경망 모델의 마지막 층의 출력값 사이의 차이를 산출하는 단계는,

상기 전역 신경망 모델의 마지막 층의 출력값에 표준화된 클래스 분류함수를 적용하고 로그를 취한 값과 상기 지역 신경망 모델의 마지막 층의 출력값에 상기 표준화된 클래스 분류함수를 적용하고 로그를 취한 값 사이의 차이를 산출하는 단계

를 포함하는,

학습 방법.

청구항 6

제 5 항에 있어서,

상기 표준화된 클래스 분류함수는, 입력 벡터와 마지막 층의 클래스를 나타내는 벡터 사이의 사잇각을 활용하여 출력값을 산출하는,

학습 방법.

청구항 7

제 1 항에 있어서,

상기 산출된 차이 및 제약의 강도를 조절하기 위한 하이퍼-파라미터에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하는 단계는,

상기 산출된 차이(D)의 음의 값(-D)을 획득하는 단계; 및

상기 산출된 차이(D)의 음의 값(-D)을 활용하여 상기 추가 제약을 결정하는 단계

를 포함하는,

학습 방법.

청구항 8

제 7 항에 있어서,

상기 추가 제약은,

상기 지역 신경망 모델의 마지막층 출력값에 표준화된 클래스 분류함수를 적용하고 로그를 취한 값; 및

상기 산출된 차이(D)의 음의 값(-D)에 표준화된 클래스 분류함수를 적용한 값에 기초하여 결정되는,

학습 방법.

청구항 9

삭제

청구항 10

삭제

청구항 11

제 1 항에 있어서,

상기 손실함수는, 크로스 엔트로피 손실값 및 상기 추가 제약에 기초한 값을 포함하고,

상기 지역 신경망 모델은, 상기 손실함수에 기초하여, 지역 데이터를 학습함과 동시에 상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 감소시키는 방향으로 학습되는,

학습 방법.

청구항 12

연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 장치로서,

하나 이상의 코어를 포함하는 프로세서; 및

메모리;

를 포함하고,

상기 프로세서는,

전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하고;

상기 산출된 차이 및 제약의 강도를 조절하기 위한 하이퍼-파라미터에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하고; 그리고

상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키도록 구성되고,

상기 하이퍼-파라미터는 상기 연합 학습에 관련된 데이터의 이종환경 분포에 비례하도록 결정되는,

장치.

청구항 13

컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 동작들을 수행하게 하고, 상기 동작들은:

전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 동작;

상기 산출된 차이 및 제약의 강도를 조절하기 위한 하이퍼-파라미터에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하는 동작; 그리고

상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키는 동작,

을 포함하고,

상기 하이퍼-파라미터는 상기 연합 학습에 관련된 데이터의 이종환경 분포에 비례하도록 결정되는,

컴퓨터 판독가능한 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 개시는 연합 학습(federate learning)을 위한 신경망 모델 학습 방법에 관한 것으로써, 구체적으로는, 연합

학습을 위한 지역 신경망 모델 학습 방법에 관한 것이다.

배 경 기 술

[0003] 연합 학습이란, 서로 다른 데이터를 가진 여러 참여자가 서로 데이터를 공유하지 않고 하나의 신경망을 학습하는 방법이다. 예를 들어, 연합 학습은, 서버로부터 수신된 신경망 모델을 복수의 기기에 저장된 학습 데이터와 프로세서 자원을 사용하여 개별 학습시킨 뒤에, 복수의 기기에서 개별 학습된 신경망 모델들을 서버로 전송하여 취합하는 학습 방법일 수 있다. 이러한 연합 학습은, 다수의 인원이 개인의 기기를 사용해 하나의 신경망 모델 학습에 참여하는 것이 가능해진다. 이러한 점으로 인해 학습이 각각의 기기에서 이루어질 수 있으므로, 폭 넓은 학습 데이터 확보에 용이하고, 학습 데이터가 서버로 노출되지 않아 개인정보보호 및 보안 정보에 대한 리스크가 적은 장점이 있다.

[0004] 예를 들어, 환자의 개인정보를 다루는 의료 분야에서는 환자의 진료정보를 기초로 신경망 모델을 학습시키고자 하더라도, 환자의 개인정보를 보호하면서 신경망 모델을 학습시키는데 어려움이 따를 수 있다. 따라서, 개인의 데이터를 외부에 노출하지 않고 하나의 신경망 모델을 학습시킬 수 있는 연합 학습이 주목받고 있다.

[0005] 그러나, 참여자 간 학습 데이터의 분포가 크게 다른 이종환경에서 연합 학습을 수행할 경우, 오히려 학습 성능 저하가 발생할 수 있다는 문제점이 있다. 이때, 상기 문제점의 원인은, 서버에 포함된 전역 신경망의 학습 방향성과 참여자가 학습시킨 지역 신경망의 학습 방향성의 차이(예컨대, 드리프트(drift))로 인한 것일 수 있다.

[0006] 즉, 지역 신경망과 전역 신경망 사이의 과도한 최적화 차이로 인한 문제를 해결할 새로운 방법이 필요하다.

[0007] 본 발명에서 언급하는 연합 학습은 "Tian Li 외 3인. Federated Learning: Challenges, Methods, and Future Directions. IEEE. IEEE Signal Processing Magazine. 2020년 5월. 제37권. 제5호. pp.50~60 (2020.05.06)"으로 부터 이해될 수 있다. 그러나, 본 개시의 적용가능한 범위는 이에 한정되지 않고, 이와 유사한 개량형, 또는 파생형 학습 방법에도 적용될 수 있다.

발명의 내용

해결하려는 과제

[0009] 본 개시는 학습 데이터가 서로 상이한 이종환경을 고려하여, 연합 학습 기반의 지역 신경망 모델을 학습시키는 것을 해결과제로 한다.

[0010] 본 개시의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 개시의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0012] 전술한 바와 같은 과제를 실현하기 위한 본 개시의 실시예에 따라, 적어도 하나의 컴퓨팅 장치에 의해 수행되는, 연합 학습(federated learning) 기반의 지역(local) 신경망 모델을 학습시키기 위한 학습 방법으로, 전역(global) 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계; 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약(regularization)을 결정하는 단계; 및 상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키는 단계를 포함할 수 있다.

[0013] 대안적 실시예에서, 상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는, 상기 전역 신경망 모델 각 층의 출력값들과 상기 지역 신경망 모델의 대응되는 각 층의 출력값들 사이의 차이들을 산출하는 단계; 및 상기 산출된 차이들 중 가장 큰 차이를 식별하는 단계를 포함할 수 있다.

[0014] 대안적 실시예에서, 상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는, 상기 전역 신경망 모델의 마지막 층의 출력값과 상기 지역 신경망 모델의 마지막 층의 출력값 사이의 차이를 산출하는 단계를 포함할 수 있다.

[0015] 대안적 실시예에서, 상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 단계는, 상기 전역 신경망 모델의 미리 결정된 층의 출력값과 상기 지역 신경망 모델의 미리 결정된 층의 출력값 사이의 차이를 산출하는 단계를 포함할 수 있다.

[0016] 대안적 실시예에서, 상기 전역 신경망 모델의 마지막 층의 출력값과 상기 지역 신경망 모델의 마지막 층의 출력값 사이의 차이를 산출하는 단계는, 상기 전역 신경망 모델의 마지막 층의 출력값에 표준화된 클래스 분류함수를 적용하고 로그를 취한 값과 상기 지역 신경망 모델의 마지막 층의 출력값에 상기 표준화된 클래스 분류함수를 적용하고 로그를 취한 값 사이의 차이를 산출하는 단계를 포함할 수 있다.

[0017] 대안적 실시예에서, 상기 표준화된 클래스 분류함수는, 입력 벡터와 마지막 층의 클래스를 나타내는 벡터 사이의 사잇각을 활용하여 출력값을 산출할 수 있다.

[0018] 대안적 실시예에서, 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하는 단계는, 상기 산출된 차이(D)의 음의 값(-D)을 획득하는 단계 및 상기 산출된 차이(D)의 음의 값(-D)을 활용하여 상기 추가 제약을 결정하는 단계를 포함할 수 있다.

[0019] 대안적 실시예에서, 상기 추가 제약은, 상기 지역 신경망 모델의 마지막층 출력값에 표준화된 클래스 분류함수를 적용하고 로그를 취한 값; 및 상기 산출된 차이(D)의 음의 값(-D)에 표준화된 클래스 분류함수를 적용한 값에 기초하여 결정될 수 있다.

[0020] 대안적 실시예에서, 상기 추가 제약은, 제약의 강도를 조절하기 위한 하이퍼-파라미터에 추가로 기초하여 결정될 수 있다.

[0021] 대안적 실시예에서, 상기 하이퍼-파라미터는 상기 연합 학습에 관련된 데이터의 이종환경 분포에 비례하도록 결정될 수 있다.

[0022] 대안적 실시예에서, 상기 손실함수는, 크로스 엔트로피 손실값 및 상기 추가 제약에 기초한 값을 포함하고, 상기 지역 신경망 모델은, 상기 손실함수에 기초하여, 지역 데이터를 학습함과 동시에 상기 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 감소시키는 방향으로 학습될 수 있다.

[0023] 전술한 바와 같은 과제를 실현하기 위한 본 개시의 실시예에 따라, 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 장치로서, 하나 이상의 코어를 포함하는 프로세서; 및 메모리; 를 포함하고, 상기 프로세서는, 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하고; 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하고; 그리고 상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키도록 구성될 수 있다.

[0024] 전술한 바와 같은 과제를 실현하기 위한 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 동작들을 수행하게 하고, 상기 동작들은: 전역 신경망 모델과 상기 지역 신경망 모델 사이의 차이를 산출하는 동작; 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하는 동작; 그리고 상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시키는 동작을 포함할 수 있다.

발명의 효과

[0026] 본 개시는, 연합 학습과 관련하여, 지역 신경망 모델과 전역 신경망 모델의 차이로 인해 발생할 수 있는 성능 저하 문제를 방지할 수 있다.

도면의 간단한 설명

[0028] 도 1은 본 개시의 일 실시예에 따른 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 학습 컴퓨팅 장치의 블록 구성도이다.

도 2는 본 개시의 일 실시예에 따라 네트워크 함수를 나타낸 개략도이다.

도 3은 본 개시의 일 실시예를 설명하기 앞서, 연합 학습을 설명하기 위한 개략도이다.

도 4는 본 개시의 일 실시예에 따라 연합 학습 기반의 지역 신경망 모델을 학습시키는 방법을 나타낸 순서도이다.

도 5는 본 개시의 일 실시예에 따라 전역 신경망과 지역 신경망의 각 층 출력값 차이를 기초로 지역 신경망 모델을 학습시키는 방법을 나타낸 순서도이다.

도 6은 본 개시의 일 실시예에 따라 전역 신경망과 지역 신경망의 마지막 층 출력값 차이를 기초로 지역 신경망

모델을 학습시키는 방법을 나타낸 순서도이다.

도 7은 본 개시의 실시예에 따른 방법인 추가 제약을 사용하여 테스트한 결과를 나타낸 그래프이다.

도 8은 본 개시의 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 간략하고 일반적인 개략도이다.

발명을 실시하기 위한 구체적인 내용

[0029] 다양한 실시예들이 이제 도면을 참조하여 설명된다. 본 명세서에서, 다양한 설명들이 본 개시의 이해를 제공하기 위해서 제시된다. 그러나, 이러한 실시예들은 이러한 구체적인 설명 없이도 실행될 수 있음이 명백하다.

[0030] 본 명세서에서 사용되는 용어 "컴포넌트", "모듈", "시스템" 등은 컴퓨터-관련 엔티티, 하드웨어, 펌웨어, 소프트웨어, 소프트웨어 및 하드웨어의 조합, 또는 소프트웨어의 실행을 지칭한다. 예를 들어, 컴포넌트는 프로세서 상에서 실행되는 처리과정(procedure), 프로세서, 객체, 실행 스레드, 프로그램, 및/또는 컴퓨터일 수 있지만, 이들로 제한되는 것은 아니다. 예를 들어, 컴퓨팅 장치에서 실행되는 애플리케이션 및 컴퓨팅 장치 모두 컴포넌트일 수 있다. 하나 이상의 컴포넌트는 프로세서 및/또는 실행 스레드 내에 상주할 수 있다. 일 컴포넌트는 하나의 컴퓨터 내에 로컬화 될 수 있다. 일 컴포넌트는 2개 이상의 컴퓨터들 사이에 분배될 수 있다. 또한, 이러한 컴포넌트들은 그 내부에 저장된 다양한 데이터 구조들을 갖는 다양한 컴퓨터 판독가능한 매체로부터 실행할 수 있다. 컴포넌트들은 예를 들어 하나 이상의 데이터 패킷들을 갖는 신호(예를 들면, 로컬 시스템, 분산 시스템에서 다른 컴포넌트와 상호작용하는 하나의 컴포넌트로부터의 데이터 및/또는 신호를 통해 다른 시스템과 인터넷과 같은 네트워크를 통해 전송되는 데이터)에 따라 로컬 및/또는 원격 처리들을 통해 통신할 수 있다.

[0031] 더불어, 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, "X는 A 또는 B를 이용한다"는 자연적인 내포적 치환 중 하나를 의미하는 것으로 의도된다. 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있다. 또한, 본 명세서에 사용된 "및/또는"이라는 용어는 열거된 관련 아이템들 중 하나 이상의 아이템의 가능한 모든 조합을 지칭하고 포함하는 것으로 이해되어야 한다.

[0032] 또한, "포함한다" 및/또는 "포함하는"이라는 용어는, 해당 특징 및/또는 구성요소가 존재함을 의미하는 것으로 이해되어야 한다. 다만, "포함한다" 및/또는 "포함하는"이라는 용어는, 하나 이상의 다른 특징, 구성요소 및/또는 이들의 그룹의 존재 또는 추가를 배제하지 않는 것으로 이해되어야 한다. 또한, 달리 특정되지 않거나 단수 형태를 지시하는 것으로 문맥상 명확하지 않은 경우에, 본 명세서와 청구범위에서 단수는 일반적으로 "하나 또는 그 이상"을 의미하는 것으로 해석되어야 한다.

[0033] 그리고, "A 또는 B 중 적어도 하나"이라는 용어는, "A만을 포함하는 경우", "B 만을 포함하는 경우" "A와 B의 구성으로 조합된 경우"를 의미하는 것으로 해석되어야 한다.

[0034] 통상의 기술자들은 추가적으로 여기서 개시된 실시예들과 관련되어 설명된 다양한 예시적 논리적 블록들, 구성들, 모듈들, 회로들, 수단들, 로직들, 및 알고리즘 단계들이 전자 하드웨어, 컴퓨터 소프트웨어, 또는 양쪽 모두의 조합들로 구현될 수 있음을 인식해야 한다. 하드웨어 및 소프트웨어의 상호교환성을 명백하게 예시하기 위해, 다양한 예시적 컴포넌트들, 블록들, 구성들, 수단들, 로직들, 모듈들, 회로들, 및 단계들은 그들의 기능성 측면에서 일반적으로 위에서 설명되었다. 그러한 기능성이 하드웨어로 또는 소프트웨어로서 구현되는지 여부는 전반적인 시스템에 부과된 특정 어플리케이션(application) 및 설계 제한들에 달려 있다. 숙련된 기술자들은 각각의 특정 어플리케이션들을 위해 다양한 방법들로 설명된 기능성을 구현할 수 있다. 다만, 그러한 구현의 결정들이 본 개시내용의 영역을 벗어나게 하는 것으로 해석되어서는 안 된다.

[0035] 제시된 실시예들에 대한 설명은 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이다. 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예 들로 한정되는 것이 아니다. 본 개시는 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

[0036] 본 개시에서 네트워크 함수와 인공 신경망, 및 뉴럴 네트워크(neural network)는 상호교환 가능하게 사용될 수 있고, 본 개시에서 사용되는 신경망 모델은 상기 언급한 네트워크 함수와 관련한 것들을 포함하는 모델로 이해될 수 있다.

[0038] 이하에서 살필 본 개시에 따른 연합 학습 방법은, 지역 신경망 모델과 전역 신경망 모델 사이의 차이(예컨대, 드리프트)로 인해 발생할 수 있는 성능 저하 문제를 방지할 수 있다.

[0039] 예를 들어, 본 개시는, 전역 신경망 모델과 지역 신경망 모델 사이의 차이를 산출하고 상기 차이에 기초하여 지역 신경망 모델의 학습을 위한 추가 제약(regularization)을 결정할 수 있다. 또한, 본 개시는, 상기 추가 제약을 기초로 손실함수에서 발생하는 손실값을 조절하여, 지역 신경망 모델이 학습하는 학습 데이터의 방향성이 전역 신경망 모델의 방향성과 상이할 경우, 해당 지역 신경망 모델의 학습이 보정되도록 할 수 있다.

[0040] 따라서, 본 개시는, 클라이언트-드리프트(client-drift), 치명적 망각(catastrophic forgetting) 등으로 인한 성능 저하를 방지할 수 있다. 또한, 본 개시는, 불특정 다수가 학습에 참여할 수 있는 연합 학습을 함에 있어, 특정 환경에서 악의적으로 혹은 의도하진 않았지만 학습목표와 전혀 상관없는 학습 데이터가 포함된 데이터셋을 기초로 지역 신경망 모델이 학습될 때의 문제를 해결할 수 있으며, 이러한 기술적 효과 이외에도 다양한 기술적 효과들을 구현할 수 있다. 한편, 상기 지역 신경망 모델과 전역 신경망 모델 사이의 차이(예컨대, 드리프트)는 지역 신경망 모델의 학습 데이터 분포에 따라 정도만 다를 뿐, 일반적으로 연합 학습시에 항상 발생하므로, 본 개시는 연합 학습과 관련된 다양한 실시예에서 구현될 수 있다.

[0042] 도 1은 본 개시의 일 실시예에 따른 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 학습 컴퓨팅 장치의 블록 구성도이다.

[0043] 도 1에 도시된 컴퓨팅 장치(100)의 구성은 간략화 하여 나타낸 예시일 뿐이다. 본 개시의 일 실시예에서 컴퓨팅 장치(100)는 컴퓨팅 장치(100)의 컴퓨팅 환경을 수행하기 위한 다른 구성들이 포함될 수 있고, 개시된 구성들 중 일부만이 컴퓨팅 장치(100)를 구성할 수도 있다.

[0044] 컴퓨팅 장치(100)는 프로세서(110), 메모리(130), 네트워크부(150)를 포함할 수 있다.

[0045] 프로세서(110)는 하나 이상의 코어로 구성될 수 있으며, 컴퓨팅 장치의 중앙 처리 장치(CPU: central processing unit), 범용 그래픽 처리 장치 (GPGPU: general purpose graphics processing unit), 텐서 처리 장치(TPU: tensor processing unit) 등의 데이터 분석, 딥러닝을 위한 프로세서를 포함할 수 있다. 프로세서(110)는 메모리(130)에 저장된 컴퓨터 프로그램을 판독하여 본 개시의 일 실시예에 따른 기계 학습을 위한 데이터 처리를 수행할 수 있다. 본 개시의 일실시예에 따라 프로세서(110)는 신경망의 학습을 위한 연산을 수행할 수 있다. 프로세서(110)는 딥러닝(DL: deep learning)에서 학습을 위한 입력 데이터의 처리, 입력 데이터에서의 피처 추출, 오차 계산, 역전파(backpropagation)를 이용한 신경망의 가중치 업데이트 등의 신경망의 학습을 위한 계산을 수행할 수 있다. 프로세서(110)의 CPU, GPGPU, 및 TPU 중 적어도 하나가 네트워크 함수의 학습을 처리할 수 있다. 예를 들어, CPU 와 GPGPU가 함께 네트워크 함수의 학습, 네트워크 함수를 이용한 데이터 분류를 처리할 수 있다. 또한, 본 개시의 일 실시예에서 복수의 컴퓨팅 장치의 프로세서를 함께 사용하여 네트워크 함수의 학습, 네트워크 함수를 이용한 데이터 분류를 처리할 수 있다. 또한, 본 개시의 일 실시예에 따른 컴퓨팅 장치에서 수행되는 컴퓨터 프로그램은 CPU, GPGPU 또는 TPU 실행가능 프로그램일 수 있다.

[0046] 본 개시의 일 실시예에 따르면, 메모리(130)는 프로세서(110)가 생성하거나 결정한 임의의 형태의 정보 및 네트워크부(150)가 수신한 임의의 형태의 정보를 저장할 수 있다.

[0047] 본 개시의 일 실시예에 따르면, 메모리(130)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory), PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 컴퓨팅 장치(100)는 인터넷(internet) 상에서 상기 메모리(130)의 저장 기능을 수행하는 웹 스토리지(web storage) 혹은 클라우드(cloud)와 관련되어 동작할 수도 있다. 전술한 메모리에 대한 기재는 예시일 뿐, 본 개시는 이에 제한되지 않는다.

[0048] 본 개시의 일 실시예에 따른 네트워크부(150)는 공중전화 교환망(PSTN: Public Switched Telephone Network), xDSL(x Digital Subscriber Line), RADSL(Rate Adaptive DSL), MDSL(Multi Rate DSL), VDSL(Very High Speed DSL), UADSL(Universal Asymmetric DSL), HDSL(High Bit Rate DSL) 및 근거리 통신망(LAN) 등과 같은 다양한 유선 통신 시스템들을 사용할 수 있다.

[0049] 또한, 본 명세서에서 제시되는 네트워크부(150)는 CDMA(Code Division Multi Access), TDMA(Time Division

Multi Access), FDMA(Frequency Division Multi Access), OFDMA(Orthogonal Frequency Division Multi Access), SC-FDMA(Single Carrier-FDMA) 및 다른 시스템들과 같은 다양한 무선 통신 시스템들을 사용할 수 있다.

[0050] 본 개시에서 네트워크부(150)는 유선 및 무선 등과 같은 그 통신 양태를 가리지 않고 구성될 수 있으며, 단거리 통신망(PAN: Personal Area Network), 근거리 통신망(WAN: Wide Area Network) 등 다양한 통신망으로 구성될 수 있다. 또한, 상기 네트워크는 공지의 월드와이드웹(WWW: World Wide Web)일 수 있으며, 적외선(IrDA: Infrared Data Association) 또는 블루투스(Bluetooth)와 같이 단거리 통신에 이용되는 무선 전송 기술을 이용할 수도 있다. 본 명세서에서 설명된 기술들은 위에서 언급된 다른 네트워크들 에서도 사용될 수 있다.

[0052] 도 2는 본 개시의 일 실시예에 따라 네트워크 함수를 나타낸 개략도이다.

[0053] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 동일한 의미로 사용될 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다. 신경망들을 구성하는 노드(또는 뉴런)들은 하나 이상의 링크에 의해 상호 연결될 수 있다.

[0054] 신경망 내에서, 링크를 통해 연결된 하나 이상의 노드들은 상대적으로 입력 노드 및 출력 노드의 관계를 형성할 수 있다. 입력 노드 및 출력 노드의 개념은 상대적인 것으로서, 하나의 노드에 대하여 출력 노드 관계에 있는 임의의 노드는 다른 노드와의 관계에서 입력 노드 관계에 있을 수 있으며, 그 역도 성립할 수 있다. 상술한 바와 같이, 입력 노드 대 출력 노드 관계는 링크를 중심으로 생성될 수 있다. 하나의 입력 노드에 하나 이상의 출력 노드가 링크를 통해 연결될 수 있으며, 그 역도 성립할 수 있다.

[0055] 하나의 링크를 통해 연결된 입력 노드 및 출력 노드 관계에서, 출력 노드의 데이터는 입력 노드에 입력된 데이터에 기초하여 그 값이 결정될 수 있다. 여기서 입력 노드와 출력 노드를 상호 연결하는 링크는 가중치(weight)를 가질 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변 될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드 값을 결정할 수 있다.

[0056] 상술한 바와 같이, 신경망은 하나 이상의 노드들이 하나 이상의 링크를 통해 상호 연결되어 신경망 내에서 입력 노드 및 출력 노드 관계를 형성한다. 신경망 내에서 노드들과 링크들의 개수 및 노드들과 링크들 사이의 연관관계, 링크들 각각에 부여된 가중치의 값에 따라, 신경망의 특성이 결정될 수 있다. 예를 들어, 동일한 개수의 노드 및 링크들이 존재하고, 링크들의 가중치 값이 상이한 두 신경망이 존재하는 경우, 두 개의 신경망들은 서로 상이한 것으로 인식될 수 있다.

[0057] 신경망은 하나 이상의 노드들의 집합으로 구성될 수 있다. 신경망을 구성하는 노드들의 부분 집합은 레이어(layer)를 구성할 수 있다. 신경망을 구성하는 노드들 중 일부는, 최초 입력 노드로부터의 거리들에 기초하여, 하나의 레이어(layer)를 구성할 수 있다. 예를 들어, 최초 입력 노드로부터 거리가 n인 노드들의 집합은, n 레이어를 구성할 수 있다. 최초 입력 노드로부터 거리는, 최초 입력 노드로부터 해당 노드까지 도달하기 위해 거쳐야 하는 링크들의 최소 개수에 의해 정의될 수 있다. 그러나, 이러한 레이어의 정의는 설명을 위한 임의적인 것으로서, 신경망 내에서 레이어의 차수는 상술한 것과 상이한 방법으로 정의될 수 있다. 예를 들어, 노드들의 레이어는 최종 출력 노드로부터 거리에 의해 정의될 수도 있다.

[0058] 최초 입력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서 링크를 거치지 않고 데이터가 직접 입력되는 하나 이상의 노드들을 의미할 수 있다. 또는, 신경망 네트워크 내에서, 링크를 기준으로 한 노드 간의 관계에 있어서, 링크로 연결된 다른 입력 노드들을 가지지 않는 노드들을 의미할 수 있다. 이와 유사하게, 최종 출력 노드는 신경망 내의 노드들 중 다른 노드들과의 관계에서, 출력 노드를 가지지 않는 하나 이상의 노드들을 의미할 수 있다. 또한, 히든 노드는 최초 입력 노드 및 최후 출력 노드가 아닌 신경망을 구성하는 노드들을 의미할 수 있다.

[0059] 딥 뉴럴 네트워크(DNN: deep neural network, 심층신경망)는 입력 레이어와 출력 레이어 외에 복수의 히든 레이어를 포함하는 신경망을 의미할 수 있다. 딥 뉴럴 네트워크를 이용하면 데이터의 잠재적인 구조(latent structures)를 파악할 수 있다. 즉, 사진, 글, 비디오, 음성, 음악의 잠재적인 구조(예를 들어, 어떤 물체가 사진에 있는지, 글의 내용과 감정이 무엇인지, 음성의 내용과 감정이 무엇인지 등)를 파악할 수 있다. 딥 뉴럴 네트워크는 컨볼루션 뉴럴 네트워크(CNN: convolutional neural network), 리커런트 뉴럴 네트워크(RNN:

recurrent neural network), 오토 인코더(auto encoder), GAN(Generative Adversarial Networks), 제한 볼츠만 머신(RBM: restricted boltzmann machine), 심층 신뢰 네트워크(DBN: deep belief network), Q 네트워크, U 네트워크, 샴 네트워크, 적대적 생성 네트워크(GAN: Generative Adversarial Network) 등을 포함할 수 있다. 전술한 딥 뉴럴 네트워크의 기재는 예시일 뿐이며 본 개시는 이에 제한되지 않는다.

[0060] 뉴럴 네트워크는 교사 학습(supervised learning), 비교사 학습(unsupervised learning), 반교사학습(semi supervised learning), 또는 강화학습(reinforcement learning) 중 적어도 하나의 방식으로 학습될 수 있다. 뉴럴 네트워크의 학습은 뉴럴 네트워크가 특정한 동작을 수행하기 위한 지식을 뉴럴 네트워크에 적용하는 과정일 수 있다.

[0061] 뉴럴 네트워크는 출력의 오류를 최소화하는 방향으로 학습될 수 있다. 뉴럴 네트워크의 학습에서 반복적으로 학습 데이터를 뉴럴 네트워크에 입력시키고 학습 데이터에 대한 뉴럴 네트워크의 출력과 타겟의 에러를 계산하고, 에러를 줄이기 위한 방향으로 뉴럴 네트워크의 에러를 뉴럴 네트워크의 출력 레이어에서부터 입력 레이어 방향으로 역전파(backpropagation)하여 뉴럴 네트워크의 각 노드의 가중치를 업데이트 하는 과정이다. 교사 학습의 경우 각각의 학습 데이터에 정답이 라벨링 되어있는 학습 데이터를 사용하며(즉, 라벨링된 학습 데이터), 비교사 학습의 경우는 각각의 학습 데이터에 정답이 라벨링되어 있지 않을 수 있다. 즉, 예를 들어 데이터 분류에 관한 교사 학습의 경우의 학습 데이터는 학습 데이터 각각에 카테고리가 라벨링 된 데이터일 수 있다. 라벨링된 학습 데이터가 뉴럴 네트워크에 입력되고, 뉴럴 네트워크의 출력(카테고리)과 학습 데이터의 라벨을 비교함으로써 오류(error)가 계산될 수 있다. 다른 예로, 데이터 분류에 관한 비교사 학습의 경우 입력인 학습 데이터가 뉴럴 네트워크 출력과 비교됨으로써 오류가 계산될 수 있다. 계산된 오류는 뉴럴 네트워크에서 역방향(즉, 출력 레이어에서 입력 레이어 방향)으로 역전파 되며, 역전파에 따라 뉴럴 네트워크의 각 레이어의 각 노드들의 연결 가중치가 업데이트 될 수 있다. 업데이트 되는 각 노드의 연결 가중치는 학습률(learning rate)에 따라 변화량이 결정될 수 있다. 입력 데이터에 대한 뉴럴 네트워크의 계산과 에러의 역전파는 학습 사이클(epoch)을 구성할 수 있다. 학습률은 뉴럴 네트워크의 학습 사이클의 반복 횟수에 따라 상이하게 적용될 수 있다. 예를 들어, 뉴럴 네트워크의 학습 초기에는 높은 학습률을 사용하여 뉴럴 네트워크가 빠르게 일정 수준의 성능을 확보하도록 하여 효율성을 높이고, 학습 후기에는 낮은 학습률을 사용하여 정확도를 높일 수 있다.

[0062] 뉴럴 네트워크의 학습에서 일반적으로 학습 데이터는 실제 데이터(즉, 학습된 뉴럴 네트워크를 이용하여 처리하고자 하는 데이터)의 부분집합일 수 있으며, 따라서, 학습 데이터에 대한 오류는 감소하나 실제 데이터에 대해서는 오류가 증가하는 학습 사이클이 존재할 수 있다. 과적합(overfitting)은 이와 같이 학습 데이터에 과하게 학습하여 실제 데이터에 대한 오류가 증가하는 현상이다. 예를 들어, 노란색 고양이를 보여 고양이를 학습한 뉴럴 네트워크가 노란색 이외의 고양이를 보고는 고양이임을 인식하지 못하는 현상이 과적합의 일종일 수 있다. 과적합은 머신러닝 알고리즘의 오류를 증가시키는 원인으로 작용할 수 있다. 이러한 과적합을 막기 위하여 다양한 최적화 방법이 사용될 수 있다. 과적합을 막기 위해서는 학습 데이터를 증가시키거나, 추가 제약(regularization), 학습의 과정에서 네트워크의 노드 일부를 비활성화하는 드롭아웃(dropout), 배치 정규화 레이어(batch normalization layer)의 활용 등의 방법이 적용될 수 있다.

[0063] 본 개시에서의 신경망 모델 사이의 차이는, 신경망 모델이 포함하는 노드가 학습된 파라미터의 차이를 의미할 수 있다. 이때, 상기 파라미터는 가중치, 또는 유사한 목적의 변수를 의미할 수 있고, 신경망 모델의 각 층을 비교하는 것은, 각 층에 포함된 노드들의 출력값을 기초로 클래스 분류함수를 사용하여 출력된 값을 비교하는 것을 포함할 수 있다.

[0064] 본 개시의 일 실시예에 따라 데이터 구조를 저장한 컴퓨터 판독가능 매체가 개시된다.

[0065] 데이터 구조는 데이터에 효율적인 접근 및 수정을 가능하게 하는 데이터의 조직, 관리, 저장을 의미할 수 있다. 데이터 구조는 특정 문제(예를 들어, 최단 시간으로 데이터 검색, 데이터 저장, 데이터 수정) 해결을 위한 데이터의 조직을 의미할 수 있다. 데이터 구조는 특정한 데이터 처리 기능을 지원하도록 설계된, 데이터 요소들 간의 물리적이거나 논리적인 관계로 정의될 수도 있다. 데이터 요소들 간의 논리적인 관계는 사용자 정의 데이터 요소들 간의 연결관계를 포함할 수 있다. 데이터 요소들 간의 물리적인 관계는 컴퓨터 판독가능 저장매체(예를 들어, 영구 저장 장치)에 물리적으로 저장되어 있는 데이터 요소들 간의 실제 관계를 포함할 수 있다. 데이터 구조는 구체적으로 데이터의 집합, 데이터 간의 관계, 데이터에 적용할 수 있는 함수 또는 명령어를 포함할 수 있다. 효과적으로 설계된 데이터 구조를 통해 컴퓨팅 장치는 컴퓨팅 장치의 자원을 최소한으로 사용하면서 연산을 수행할 수 있다. 구체적으로 컴퓨팅 장치는 효과적으로 설계된 데이터 구조를 통해 연산, 읽기, 삽입, 삭제, 비교, 교환, 검색의 효율성을 높일 수 있다.

[0066] 데이터 구조는 데이터 구조의 형태에 따라 선형 데이터 구조와 비선형 데이터 구조로 구분될 수 있다. 선형 데이터 구조는 하나의 데이터 뒤에 하나의 데이터만이 연결되는 구조일 수 있다. 선형 데이터 구조는 리스트(List), 스택(Stack), 큐(Queue), 데크(Deque)를 포함할 수 있다. 리스트는 내부적으로 순서가 존재하는 일련의 데이터 집합을 의미할 수 있다. 리스트는 연결 리스트(Linked List)를 포함할 수 있다. 연결 리스트는 각각의 데이터가 포인터를 가지고 한 줄로 연결되어 있는 방식으로 데이터가 연결된 데이터 구조일 수 있다. 연결 리스트에서 포인터는 다음이나 이전 데이터와의 연결 정보를 포함할 수 있다. 연결 리스트는 형태에 따라 단일 연결 리스트, 이중 연결 리스트, 원형 연결 리스트로 표현될 수 있다. 스택은 제한적으로 데이터에 접근할 수 있는 데이터 나열 구조일 수 있다. 스택은 데이터 구조의 한 쪽 끝에서만 데이터를 처리(예를 들어, 삽입 또는 삭제)할 수 있는 선형 데이터 구조일 수 있다. 스택에 저장된 데이터는 늦게 들어갈수록 빨리 나오는 데이터 구조(LIFO-Last in First Out)일 수 있다. 큐는 제한적으로 데이터에 접근할 수 있는 데이터 나열 구조로서, 스택과 달리 늦게 저장된 데이터일수록 늦게 나오는 데이터 구조(FIFO-First in First Out)일 수 있다. 데크는 데이터 구조의 양 쪽 끝에서 데이터를 처리할 수 있는 데이터 구조일 수 있다.

[0067] 비선형 데이터 구조는 하나의 데이터 뒤에 복수개의 데이터들이 연결되는 구조일 수 있다. 비선형 데이터 구조는 그래프(Graph) 데이터 구조를 포함할 수 있다. 그래프 데이터 구조는 정점(Vertex)과 간선(Edge)으로 정의될 수 있으며 간선은 서로 다른 두개의 정점을 연결하는 선을 포함할 수 있다. 그래프 데이터 구조 트리(Tree) 데이터 구조를 포함할 수 있다. 트리 데이터 구조는 트리에 포함된 복수개의 정점들 중에서 서로 다른 두개의 정점을 연결시키는 경로가 하나인 데이터 구조일 수 있다. 즉 그래프 데이터 구조에서 루프(loop)를 형성하지 않는 데이터 구조일 수 있다.

[0068] 본 명세서에 걸쳐, 연산 모델, 신경망, 네트워크 함수, 뉴럴 네트워크(neural network)는 동일한 의미로 사용될 수 있다. 이하에서는 신경망으로 통일하여 기술한다. 데이터 구조는 신경망을 포함할 수 있다. 그리고 신경망을 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망을 포함한 데이터 구조는 신경망에 의한 처리를 위하여 전처리된 데이터, 신경망에 입력되는 데이터, 신경망의 가중치, 신경망의 하이퍼 파라미터, 신경망으로부터 획득한 데이터, 신경망의 각 노드 또는 레이어와 연관된 활성 함수, 신경망의 학습을 위한 손실함수 등을 포함할 수 있다. 신경망을 포함한 데이터 구조는 상기 개시된 구성들 중 임의의 구성 요소들을 포함할 수 있다. 즉 신경망을 포함한 데이터 구조는 신경망에 의한 처리를 위하여 전처리된 데이터, 신경망에 입력되는 데이터, 신경망의 가중치, 신경망의 하이퍼 파라미터, 신경망으로부터 획득한 데이터, 신경망의 각 노드 또는 레이어와 연관된 활성 함수, 신경망의 학습을 위한 손실함수 등 전부 또는 이들의 임의의 조합을 포함하여 구성될 수 있다. 전술한 구성들 이외에도, 신경망을 포함한 데이터 구조는 신경망의 특성을 결정하는 임의의 다른 정보를 포함할 수 있다. 또한, 데이터 구조는 신경망의 연산 과정에 사용되거나 발생되는 모든 형태의 데이터를 포함할 수 있으며 전술한 사항에 제한되는 것은 아니다. 컴퓨터 판독가능 매체는 컴퓨터 판독가능 기록 매체 및/또는 컴퓨터 판독가능 전송 매체를 포함할 수 있다. 신경망은 일반적으로 노드라 지칭될 수 있는 상호 연결된 계산 단위들의 집합으로 구성될 수 있다. 이러한 노드들은 뉴런(neuron)들로 지칭될 수도 있다. 신경망은 적어도 하나 이상의 노드들을 포함하여 구성된다.

[0069] 데이터 구조는 신경망에 입력되는 데이터를 포함할 수 있다. 신경망에 입력되는 데이터를 포함하는 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망에 입력되는 데이터는 신경망 학습 과정에서 입력되는 학습 데이터 및/또는 학습이 완료된 신경망에 입력되는 입력 데이터를 포함할 수 있다. 신경망에 입력되는 데이터는 전처리(pre-processing)를 거친 데이터 및/또는 전처리 대상이 되는 데이터를 포함할 수 있다. 전처리는 데이터를 신경망에 입력시키기 위한 데이터 처리 과정을 포함할 수 있다. 따라서 데이터 구조는 전처리 대상이 되는 데이터 및 전처리로 발생되는 데이터를 포함할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0070] 데이터 구조는 신경망의 가중치를 포함할 수 있다. (본 명세서에서 가중치, 파라미터는 동일한 의미로 사용될 수 있다.) 그리고 신경망의 가중치를 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 신경망은 복수개의 가중치들을 포함할 수 있다. 가중치는 가변적일 수 있으며, 신경망이 원하는 기능을 수행하기 위해, 사용자 또는 알고리즘에 의해 가변 될 수 있다. 예를 들어, 하나의 출력 노드에 하나 이상의 입력 노드가 각각의 링크에 의해 상호 연결된 경우, 출력 노드는 상기 출력 노드와 연결된 입력 노드들에 입력된 값들 및 각각의 입력 노드들에 대응하는 링크에 설정된 가중치에 기초하여 출력 노드에서 출력되는 데이터 값을 결정할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0071] 제한이 아닌 예로서, 가중치는 신경망 학습 과정에서 가변되는 가중치 및/또는 신경망 학습이 완료된 가중치를 포함할 수 있다. 신경망 학습 과정에서 가변되는 가중치는 학습 사이클이 시작되는 시점의 가중치 및/또는 학습

사이클 동안 가변되는 가중치를 포함할 수 있다. 신경망 학습이 완료된 가중치는 학습 사이클이 완료된 가중치를 포함할 수 있다. 따라서 신경망의 가중치를 포함한 데이터 구조는 신경망 학습 과정에서 가변되는 가중치 및/또는 신경망 학습이 완료된 가중치를 포함한 데이터 구조를 포함할 수 있다. 그러므로 상술한 가중치 및/또는 각 가중치의 조합은 신경망의 가중치를 포함한 데이터 구조에 포함되는 것으로 한다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0072] 신경망의 가중치를 포함한 데이터 구조는 직렬화(serialization) 과정을 거친 후 컴퓨터 판독가능 저장 매체(예를 들어, 메모리, 하드 디스크)에 저장될 수 있다. 직렬화는 데이터 구조를 동일하거나 다른 컴퓨팅 장치에 저장하고 나중에 다시 재구성하여 사용할 수 있는 형태로 변환하는 과정일 수 있다. 컴퓨팅 장치는 데이터 구조를 직렬화 하여 네트워크를 통해 데이터를 송수신할 수 있다. 직렬화된 신경망의 가중치를 포함한 데이터 구조는 역직렬화(deserialization)를 통해 동일한 컴퓨팅 장치 또는 다른 컴퓨팅 장치에서 재구성될 수 있다. 신경망의 가중치를 포함한 데이터 구조는 직렬화에 한정되는 것은 아니다. 나아가 신경망의 가중치를 포함한 데이터 구조는 컴퓨팅 장치의 자원을 최소한으로 사용하면서 연산의 효율을 높이기 위한 데이터 구조(예를 들어, 비선형 데이터 구조에서 B-Tree, Trie, m-way search tree, AVL tree, Red-Black Tree)를 포함할 수 있다. 전술한 사항은 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0073] 데이터 구조는 신경망의 하이퍼 파라미터(Hyper-parameter)를 포함할 수 있다. 그리고 신경망의 하이퍼 파라미터를 포함한 데이터 구조는 컴퓨터 판독가능 매체에 저장될 수 있다. 하이퍼 파라미터는 사용자에 의해 가변되는 변수일 수 있다. 하이퍼 파라미터는 예를 들어, 학습률(learning rate), 비용 함수(cost function), 학습 사이클 반복 횟수, 가중치 초기화(Weight initialization)(예를 들어, 가중치 초기화 대상이 되는 가중치 값의 범위 설정), Hidden Unit 개수(예를 들어, 히든 레이어의 개수, 히든 레이어의 노드 수)를 포함할 수 있다. 전술한 데이터 구조는 예시일 뿐 본 개시는 이에 제한되지 않는다.

[0075] 도 3은 본 개시의 일 실시예를 설명하기 앞서, 연합 학습을 설명하기 위한 개략도이다.

[0076] 본 개시의 실시예들을 설명하기 위해 필요한 연합 학습의 개념들을 도 3을 참조하여 설명한다. 아래에서 살펴볼 도 3을 참조한 설명은 연합 학습의 예시이다. 상기 예시는 본 개시의 실시예들을 설명하기 위한 것일 뿐, 본 개시를 한정하지는 않는다.

[0077] 서버(300)가 포함하고 있는 전역 신경망 모델을 대상으로 연합 학습을 수행함에 있어, 서버(300)는 적어도 하나의 학습 참여자(310)들에게 전역 신경망을 각각 할당할 수 있다. 이어서, 학습 참여자(310)들은 개별적으로 가지고 있는 학습 데이터와 프로세서를 기초로 할당 받은 전역 신경망 모델인 지역 신경망 모델을 학습시킬 수 있다. 최종적으로, 서버(300)는 상기 학습된 지역 신경망 모델(330)들을 집계하여 전역 신경망 모델에 반영할 수 있다.

[0078] 이때, 연합 학습은 거시적으로 '지역 학습 단계'와, '전역 집계 단계'로 분류될 수 있다. 구체적으로, 각각의 학습 참여자(310)가 서버(300)에 포함된 전역 신경망 모델을 할당 받아, 학습시키는 '지역 학습 단계' 그리고, 상기 각각의 학습 참여자(310)가 학습시킨, 학습된 지역 신경망 모델(330)들을 상기 서버(300)가 집계하여 전역 신경망 모델을 학습시키는 것을 '전역 집계 단계'로 분류될 수 있다. - 이때, 상기 학습 참여자(310)는 자연인을 지칭하는 것이 아닌 서로 상이할 수 있는 프로세서 및 지역 학습 데이터를 포함하는 컴퓨팅 장치로 이해될 수 있다. 또한, 상기 서버는 서버용 컴퓨터에 한정되지 않고, 프로세서를 포함하는 컴퓨팅 장치로 이해될 수 있다.

[0079] 위의 예시와 같이 연합 학습을 수행할 경우, 학습 참여자(310)의 학습 데이터가 서버(300)에 노출되지 않는다. 이는 개인정보 보호 차원에서 장점이 될 수 있지만, 학습 참여자(310)가 지역 신경망 모델을 연합 학습 성능에 악영향을 끼치는 학습 데이터로 학습시키더라도 서버(300)가 통제하는데 어려움이 따를 수 있다.

[0080] 한편, 실제 연합 학습 환경에서 학습 참여자(310) 간의 프로세서 및 학습 데이터의 분포가 상이할 수 있고, 이러한 상이한 분포를 발생시키는 환경을 본 개시에서는 이종환경(heterogenous environment)이라고 칭한다.

[0081] 서버(300)가 이종환경에서 학습된 지역 신경망 모델(330)들을 사용하여 전역 집계 단계를 수행할 경우, 전역 신경망 모델의 학습에 성능이 떨어질 수 있다. 예를 들어, 이종환경에서 학습된 지역 신경망 모델(330)들이 전역 신경망 모델의 수렴을 방해하는 문제가 발생될 수 있다. 또한, 지역 최적화 단계에서도, 이종환경에 의해 최적화가 방해되는 문제가 발생될 수 있다. 예컨대, 이종환경에서 지역 최적화를 수행하는 과정에서, 이전의 전역 집계 단계에서 집계한 파라미터를 망각하는 치명적 망각 현상(catastrophic)과 같은 성능 저하 현상이 발생할 수 있다. 한편, 이와 같은 이종환경에서의 성능 저하의 주된 요인 중 하나로, 이종환경에서의 학습에 유발되는

전역 신경망 모델과 지역 신경망 모델 사이의 차이(예컨대, 드리프트)가 있을 수 있다.

[0082] 요컨대, 연합 학습과 관련하여, 지역 최적화로 인해 전역 신경망 모델과 지역 신경망 모델 사이의 큰 차이가 유발될 경우 학습 성능이 떨어질 수 있으므로, 성능 개선을 위하여 상기 차이를 감소시키는 것이 필요하다.

[0084] 지금까지, 도 3을 기초로 본 개시의 실시예들을 설명하기 위한 연합 학습의 예시와 그 외의 것을 다루었다. 아래는 도 4 내지 6을 기초로 본 개시의 과제를 해결하기 위한 실시예들을 살핀다.

[0085] 도 4를 참조하여 연합 학습 기반의 지역 신경망 모델을 학습시키는 방법에 관한 본 개시의 일 실시예가 개시된다. 도 4를 참조하면 본 개시의 프로세서는, 연합 학습 기반의 지역 신경망 모델을 학습시키기 위한 목적을 달성하기 위해 복수의 단계들을 수행할 수 있다.

[0086] 예를 들어, 프로세서(110)는 전역 신경망 모델과 지역 신경망 모델 사이의 차이를 산출할 수 있다(S401). 여기서 차이는, 예를 들어, "가장 최근에 전역 집계된 전역 신경망 모델" 및 "이러한 전역 신경망 모델에 지역 데이터를 기반으로 추가적인 학습이 적용되어 생성된 지역 신경망 모델"의 차이일 수 있다. 앞서 설명했듯이, 신경망 모델과 지역 신경망 모델 사이의 차이가 클수록 연합 학습의 성능이 감소될 수 있으므로, 이러한 차이에 기초한 상관관계가 연합 학습의 추가 제약을 결정하기 위해 활용될 수 있다.

[0087] 또한, 프로세서(110)는 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정할 수 있다(S402). 이때, 추가 제약은, 상기 지역 신경망 모델의 학습에 직접적인 영향을 주는 손실값을 보정하기 위한 것일 수 있다.

[0088] 또한, 프로세서(110)는 상기 결정된 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시킬 수 있다(S403). 이때, 프로세서(110)는 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시킴으로써, 추가 제약을 부여하지 않을 때와 비교하여 상기 전역 신경망 모델과의 차이를 좁힐 수 있다. 다시 말해, 프로세서(110)는, 추가 제약을 포함하는 손실함수에 기초하여 상기 지역 신경망 모델을 학습시킴으로써, 상기 지역 신경망 모델이 전체 데이터 분포에 대한 최적화 범위를 벗어난 상태로 지역 데이터에 대해서만 과도하게 최적화되는 것을 방지할 수 있다.

[0090] 한편, 지역 신경망 모델과 전역 신경망 모델의 차이는, 신경망 모델의 모든 층을 대상으로 발생할 수 있다. 이때, 차이가 가장 큰 층을 기초로 추가 제약을 결정하는 것이 가장 변별력이 있으므로, 차이가 가장 큰 층을 식별할 필요가 있다.

[0091] 상기 필요성에 따른 일 실시예가 도 5를 참조하여 개시된다. 도 5를 참조하면, 프로세서(110)는 전역 신경망 모델의 각 층의 출력값과 상기 지역 신경망 모델의 대응되는 각 층의 출력값 사이의 차이들을 산출(520)할 수 있다. 이때, 상기 전역 신경망 모델(500)과 상기 지역 신경망 모델(510)이 각각 포함하는 층과 각 층의 노드의 개수는 상기 지역 신경망 모델(510)이 상기 전역 신경망 모델(500)로 할당 받기 때문에 동일한 크기와 구조를 가진다고 가정한다. 또한, 프로세서(110)는, 다양한 방식에 기초하여, 각 층의 출력값들 사이의 차이들을 산출할 수 있다. 예를 들어, 프로세서(110)는, 각 층의 출력값들 사이에 직접적인 감산 연산을 적용하여 차이들을 산출할 수 있다. 또한, 프로세서(110)는, 신경망 모델의 추론과 관련된 최적화의 관점에서 차이들을 명확하게 평가하기 위해, "상기 전역 신경망 모델(500)의 각 층의 출력값에 표준화된 클래스 분류함수를 적용하고 로그를 취한 값"과 "상기 지역 신경망 모델(510)의 각 층의 출력값에 표준화된 클래스 분류함수(511)를 적용하고 로그를 취한 값" 사이의 차이를 연산할 수도 있다.

[0092] 다음으로, 프로세서(110)는 각 층들 사이에서 산출된 차이들 중 가장 큰 차이를 식별할 수 있다. 또한, 프로세서(110)는, 가장 큰 차이를 갖는 것으로 식별된 층의 차이 (521)를 기초로 추가 제약을 결정(522)을 할 수 있다. 마지막으로, 프로세서(110)는 상기 결정된 추가 제약이 반영된 손실함수(512)를 사용하여 상기 지역 신경망 모델(510)을 학습시킬 수 있다.

[0093] 앞서 전역 신경망 모델과 지역 신경망 모델 사이의 차이가 가장 크게 발생하는 층에 관한 실험을 진행했을 때, 신경망 모델의 마지막 층에서 가장 차이가 크게 발생되는 것이 확인되었다. 따라서, 전역 신경망 모델의 마지막 층과 지역 신경망 모델의 마지막 층 사이의 차이를 기초로 손실함수에 추가적인 제약을 부여하는 것이, 치명적 망각, 클라이언트 드리프트 등의 문제점을 해결하고 연합 학습의 성능을 개선시키는 데에 더욱 효과적일 수 있다.

[0094] 이러한 점을 고려한 일 실시예가 도 6을 참조하여 개시된다. 한편, 이러한 실시예는 신경망 모델의 모든 층을 고려하지 않아도 되므로, 비교적 프로세서 자원을 적게 소비하는 효과 역시 추가적으로 구현할 수 있다. 도 6를

참조하면, 프로세서(110)는 전역 신경망 모델의 마지막 층과 지역 신경망 모델의 마지막 층 사이의 차이를 비교(620)할 수 있다. 또한, 프로세서(110)는 상기 마지막 층들 사이의 비교(620)를 기초로, 상기 마지막 층들의 출력값들 사이의 차이(621)를 산출할 수 있으며, 산출된 차이에 기초하여 손실함수의 보정을 위한 추가 제약을 결정(622)할 수 있다. 또한, 프로세서(110)는 상기 결정된 추가 제약을 손실함수(612)에 부가하여, 수정된 손실함수를 생성할 수 있다. 마지막으로, 프로세서(110)는 수정된 손실함수를 사용하여 상기 지역 신경망 모델(610)을 학습시킬 수 있다.

[0095] 한편, 추가 제약을 결정하기 위한 층을 정하는 실시예는 이에 한정되지 않고, 다양한 실시예들이 구현될 수 있다. 예를 들어, 프로세서(110)는, 전역 신경망 모델의 "미리 결정된 층"과 지역 신경망 모델의 "미리 결정된 층" 사이의 차이에 기초하여 손실함수에 추가적인 제약을 부여할 수도 있다. 여기서 "미리 결정된 층"은, 사전 지식, 실험, 시뮬레이션 등에 의해서, 전역 신경망 모델과 지역 신경망 모델 사이의 차이가 클 것으로 알려져 있거나 예상되는 층일 수 있다.

[0097] 앞서, 전역 신경망과 지역 신경망의 차이를 획득하기 위한 층을 정하고, 상기 층에 대한 전역 신경망과 지역 신경망의 차이를 연산하여 추가 제약을 결정하는 본 개시의 실시예들을 개시하였다. 아래에서는, 이기종 환경에서 연합학습을 수행함에 있어, 본 개시의 실시예의 방법으로 추가 제약을 사용한 연합학습과 그렇지 않은 연합 학습을 테스트하고 상기 테스트의 결과를 도 7과 표 1을 기초로 설명한다.

[0098] 상기 테스트에서의 비교 기준이 되는 연합학습은 이기종 환경에서 10개의 지역 신경망 모델의 학습 파라미터를 1개의 전역 신경망 모델에 집계하여 정확도가 측정되었다. 또한, 상기 테스트는 이기종 환경을 구성하기 위해 각 지역 신경망 모델에 소정의 Non-IID 학습 데이터(예컨대, 집중 파라미터 β를 갖는 디리클레 분포(Dirichlet distribution)에 기초하여 생성된 Non-IID 학습 데이터)를 적용하였다.

표 1

[0099]

차이를 획득할 층	본 개시의 실시예에 따른 방법의 정확도 (기준과 비교)	기준의 정확도 (FedAvg)
특징 추출층 (Feature extractor)	68.9 ±0.3(-0.2)	69.1 ±0.6
입력층 (Header)	69.6 ±0.5(+0.5)	
마지막층(Classifier)	70.8 ±0.4(+1.7)	

[0100] 표 1은 상기 테스트의 결과를 기초로 기준(즉, 추가 제약이 없는 방법)과 본 개시의 실시예에 따른 추가 제약을 적용한 방법을 비교한 것이다. (이때, 본 개시의 실시예의 방법으로 차이를 획득하는 층은 특징 추출층(feature extractor), 입력층(Header), 및 마지막층(Classifier)을 각각 사용하였다.)

[0101] 표 1을 참조하면, 특징 추출층의 차이를 기초한 경우 정확도가 기준과 비교하여 약간 떨어졌으나, 입력층의 차이를 기초한 경우와 마지막층의 차이를 기초한 경우는 기준과 비교하여 유의미한 향상이 확인되었다. 특히, 마지막층의 차이를 기초한 경우 기준과 비교하여 가장 높은 정확도를 보였다. (기준과 비교하여 1.7 향상)

[0102] 또한, 도 7을 참조하면, 본 개시의 실시예에 따른 방법인 추가 제약을 적용하였을 경우, (추가 제약이 없는) 기준인 'FedAvg'와 비교하였을 때, 특징 추출층(Feature extractor)의 차이에 기초한 실시예, 입력층(Header)의 차이에 기초한 실시예, 및 마지막층(Classifier)의 차이에 기초한 실시예 모두 모든 지역 학습 단계에서 KL 발산(Kullback-Leibler divergence)이 기준보다 낮게 측정되었고, CKA 유사성(CKA similarity)이 기준보다 높게 측정되었다. 즉, 추가 제약이 반영된 실시예들 모두, "전역 신경망 모델과 지역 신경망 모델의 차이(예컨대, 드리프트)"가 (추가 제약이 없는) 기준보다 완화되는 것을 확인할 수 있었다.

[0103] 앞서, 전역 신경망 모델과 지역 신경망 모델간의 차이를 산출하는 방법에 관한 실시예들이 개시되었다. 아래는 수학식 1 내지 4를 참조하여 앞서 언급한 실시예들에 관한 보다 상세한 실시예들이 개시된다.

[0104] 프로세서(110)에 의해 수행되는, 상기 마지막 층의 출력값들을 비교(620)하여 상기 마지막 층의 출력값들 사이의 차이(621)를 출력하는 방법에 관한 일 실시예가 아래 수학식 1을 참조하여 설명된다.

수학식 1

[0106] $$f_D^{y_i}(x_i) = \log(\sigma(f_P^{y_i}(x_i)) - \log(\sigma(f_G^{y_i}(x_i))$$

[0108] 수학식 1은 차이를 산출하는 단계(S401)에 관한 실시예를 수식으로 나타낸 것으로, 상기 수학식 1을 참조하면, $f_D(\cdot)$는 상기 전역 신경망 모델과 지역 신경망 모델의 차이를 의미하고, $f_P(\cdot)$ 및 $f_G(\cdot)$는 각각 지역 신경망 모델의 출력 및 전역 신경망 모델의 출력을 의미한다. 또한, 이러한 출력들은, 바람직하게는, 신경망 모델의 마지막 층의 출력들일 수 있다. 또한, σ는 표준화된 클래스 분류함수를 의미하고, 상기 클래스 분류함수로 소프트맥스(softmax) 함수를 포함할 수 있으나, 이에 한정되지는 않는다.

[0109] 즉, 상기 프로세서(110)는, "상기 전역 신경망 모델의 마지막 층의 출력값을 기초로 표준화된 클래스 분류함수를 적용함에 이어서 로그를 취한 값"과, "상기 지역 신경망 모델 마지막 층의 출력값에 상기 표준화된 클래스 분류함수를 적용함에 이어서 로그를 취한 값" 사이의 차이를 연산할 수 있다.

[0110] 이때, 프로세서(110)가 연산한 상기 차이는, 현재 지역 신경망 모델이 전역 신경망 모델과 비교하여 학습의 방향성이 얼마나 차이가 나는지에 대한 지표가 될 수 있다.

[0111] 또한, 상기 표준화된 클래스 분류함수는, 기존의 클래스 분류함수에 기초하되, 학습 과정에서 파라미터의 크기를 표준화(normalization)하여 학습된 분류함수에 해당한다. 본 개시는, "전역 신경망 모델과 지역 신경망 모델의 차이(예컨대, 드리프트)"를 산출(또는 추정)하는 과정에서 이러한 표준화된 클래스 분류함수를 활용함으로써, 상기 차이가 신경망층의 파라미터의 크기에 따라 달라질 수 있는 문제를 해결할 수 있으며, 이를 통해 상기 차이가 더욱 정확하게 산출(또는 추정)되도록 할 수 있다.

[0112] 한편, 상기 표준화된 클래스 분류함수는, 입력 벡터와 마지막 층의 클래스를 나타내는 벡터 사이의 사잇각을 활용하여 출력값을 산출하는 형태로 구현될 수 있으며, 이와 관련된 일 실시예가 아래 수학식 2를 참조하여 설명된다.

수학식 2

[0114] $$\sigma(f^{yi}(x_i;\omega)) = \frac{e^{cos\theta_i/\tau}}{\sum_{k=1}^{|Y|} e^{cos\theta_k/\tau}}$$

[0115] 수학식 2는 상기 표준화된 클래스 분류함수($\sigma(\cdot)$)를 표현하는 것으로, θ_i는 입력 벡터와 마지막 층의 i 번째 클래스를 나타내는 벡터의 사잇각을 의미할 수 있다. 이때, 상기 사잇각은, 입력 벡터와 마지막 층의 i 번째 클래스를 나타내는 벡터의 내적을 통해 연산될 수 있다.

[0116] 수학식 2를 참조하면, 상기 표준화된 클래스 분류함수는, 마지막 층의 i 번째 클래스에 관한 표준화된 값을 모든 클래스들(Y개의 클래스들)에 대한 표준화된 값들의 합으로 나누는 것을 포함할 수 있다.

[0118] 앞서 언급한 S402 단계에 관련된 실시예로, 프로세서(110)는 상기 산출된 차이에 기초하여 상기 지역 신경망 모델의 학습을 위한 추가 제약을 결정하는 단계에서, 상기 산출된 차이(D)의 음의 값(-D)을 획득하는 단계; 및 상기 산출된 차이(D)의 음의 값(-D)을 활용하여 상기 추가 제약을 결정하는 단계를 포함하여 수행할 수 있다. 상기 추가 제약은 아래 수학식 3을 참조하여 설명된다.

수학식 3

[0120]
$$R^{yi}(x_i, \omega) = -\sigma\left(-f_D^{yi}(x_i)\right) \cdot \log(\sigma\left(f_P^{yi}(x_i)\right))$$

[0122] 수학식 3은 프로세서(110)가 결정할 수 있는 상기 추가 제약($R(\cdot)$)이 표현된 것이다. (이때, 수학식 3의 'yi'는 i 번째 클래스에 대한 출력을 의미한다.)

[0123] 수학식 3을 참조하면, 상기 추가 제약($R(\cdot)$)은 "상기 산출된 차이($f_D(\cdot)$)의 음의 값($-f_D(\cdot)$)을 기초로, 음수를 취한 표준화된 클래스 분류함수($-\sigma(\cdot)$)를 사용한 값"과, "상기 지역 신경망 모델($f_P(\cdot)$)의 마지막층의 출력값을 기초로, 표준화된 클래스 분류함수($\sigma(\cdot)$)를 사용한 값의 로그값"과의 곱(·)을 포함할 수 있다.

[0124] 한편, 상기 추가 제약($R(\cdot)$)은, 학습을 통한 최소화를 위하여 "음수를 취한 표준화된 클래스 분류함수($-\sigma(\cdot)$)"의 형태를 갖고 있고, 차이의 반대 방향(즉, 전역 신경망 모델에 가까워지는 방향)을 구현하기 위해 상기 차이($f_D(\cdot)$)에 음수를 취한 값($-f_D(\cdot)$)을 사용한다. 이러한 추가 제약($R(\cdot)$)은 상기 차이($f_D(\cdot)$)의 반대 방향으로 최적화되도록 지역 신경망 모델 학습을 유도할 수 있다.

[0126] S402 단계에 관련된 추가적인 일 실시예로, 상기 프로세서(110)는 상기 추가 제약을 결정함에 있어, 제약의 강도를 조절하기 위한 하이퍼-파라미터(λ_R)를 추가로 부가하여 결정할 수 있다. 이때, 상기 하이퍼-파라미터(λ_R)는 상기 연합 학습에 관련된 데이터의 이종환경 분포에 비례하도록 상기 프로세서(110)에 의해 결정될 수 있다. 예를 들어, 상기 하이퍼-파라미터(λ_R)는 지역 신경망 모델의 학습 데이터 분포와 전역 신경망 모델의 학습 데이터 분포의 차이가 클수록 더 크게 결정될 수 있으며, 이를 통해, 지역 신경망 모델의 최적화 성능이 향상될 수 있다. 상기 하이퍼-파라미터(λ_R)를 추가로 부가함으로써, 프로세서(110)는 상기 추가 제약을 고려하여 상기 지역 신경망을 학습하되, 추가 제약의 완급을 조절할 수 있게 된다. 이는, 프로세서(110)가 이종환경 분포를 고려함으로써, 지역 신경망 모델이 전역 신경망 모델과 크게 변화하지 않도록 효과적으로 유도하고, 앞서 설명한 치명적 망각 현상을 감소시키는데 도움이 된다. 또한, 전역 신경망 모델에 집계될 때, 전역 신경망 모델이 최적값에 잘 수렴될 수 있도록 도움이 된다.

[0128] S403 단계에 관련된 실시예로, 상기 프로세서(110)는 상기 지역 신경망 모델($f_P(\cdot)$)를 상기 손실함수에 기초하여 학습시키되, 상기 손실함수는 상기 지역 신경망 모델($f_P(\cdot)$)이 전역 신경망 모델($f_G(\cdot)$)과 상기 지역 신경망 모델($f_P(\cdot)$) 사이의 차이($f_D(\cdot)$)를 감소시키는 방향으로 학습되도록 손실값을 출력하는 특징을 가질 수 있다.

[0129] 상기 일 실시예를 위한 상기 손실함수는 아래 수학식 4를 참조하여 설명된다.

수학식 4

[0131]
$$L(T, \omega) = -\sum_{i=1}^{|T|}\sum_{j=1}^{|Y|} y_{ij} \cdot \log \sigma(f^{y_j}(\mathrm{x_i})) + \lambda_R \cdot R^{y_j}(\mathrm{x_i}; \omega)$$

[0133] 수학식 4는 상기 손실함수($L(\cdot)$)를 표현하는 것으로, 수학식 4를 참조하면, 상기 손실함수($L(\cdot)$)는 모든 학습 데이터(T) 및 상기 지역 신경망 모델의 마지막층의 모든 클래스(Y)와 관련된 크로스 엔트로피(cross-entropy) 손실값, 및 추가 제약에 기초한 값($R(\cdot)$)을 포함할 수 있다. 또한, 상기 추가 제약에 기초한 값($R(\cdot)$)에는 상기

하이퍼-파라미터(λ_R)가 추가로 곱해질 수 있으며, 이를 통해, 이종환경 분포를 고려하여 학습의 완급이 조절될 수 있다.

[0134] 이는 프로세서(110)가 상기 지역 신경망 모델을 학습시킬 때, 지역 신경망과 전역 신경망과의 차이 정도에 따라 학습의 정도를 보정하기 위한 것으로, 지역 신경망 모델과 전역 신경망 모델의 과도한 차이를 방지할 수 있다.

[0136] 도 8은 본 개시의 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 간략하고 일반적인 개략도이다.

[0137] 본 개시가 일반적으로 컴퓨팅 장치에 의해 구현될 수 있는 것으로 전술되었지만, 통상의 기술자라면 본 개시가 하나 이상의 컴퓨터 상에서 실행될 수 있는 컴퓨터 실행가능 명령어 및/또는 기타 프로그램 모듈들과 결합되어 및/또는 하드웨어와 소프트웨어의 조합으로써 구현될 수 있다는 것을 잘 알 것이다.

[0138] 일반적으로, 프로그램 모듈은 특정의 태스크를 수행하거나 특정의 추상 데이터 유형을 구현하는 루틴, 프로그램, 컴포넌트, 데이터 구조, 기타 등등을 포함한다. 또한, 통상의 기술자라면 본 개시의 방법이 단일-프로세서 또는 멀티프로세서 컴퓨터 시스템, 미니컴퓨터, 메인프레임 컴퓨터는 물론 퍼스널 컴퓨터, 핸드헬드(handheld) 컴퓨팅 장치, 마이크로프로세서-기반 또는 프로그램가능 가전 제품, 기타 등등(이들 각각은 하나 이상의 연관된 장치와 연결되어 동작할 수 있음)을 비롯한 다른 컴퓨터 시스템 구성으로 실시될 수 있다는 것을 잘 알 것이다.

[0139] 본 개시의 설명된 실시예들은 어떤 태스크들이 통신 네트워크를 통해 연결되어 있는 원격 처리 장치들에 의해 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 둘 다에 위치할 수 있다.

[0140] 컴퓨터는 통상적으로 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터에 의해 액세스 가능한 매체는 그 어떤 것이든지 컴퓨터 판독가능 매체가 될 수 있고, 이러한 컴퓨터 판독가능 매체는 휘발성 및 비휘발성 매체, 일시적(transitory) 및 비일시적(non-transitory) 매체, 이동식 및 비-이동식 매체를 포함한다. 제한이 아닌 예로서, 컴퓨터 판독가능 매체는 컴퓨터 판독가능 저장 매체 및 컴퓨터 판독가능 전송 매체를 포함할 수 있다. 컴퓨터 판독가능 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보를 저장하는 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성 매체, 일시적 및 비-일시적 매체, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 판독가능 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital video disk) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터에 의해 액세스될 수 있고 원하는 정보를 저장하는 데 사용될 수 있는 임의의 기타 매체를 포함하지만, 이에 한정되지 않는다.

[0141] 컴퓨터 판독가능 전송 매체는 통상적으로 반송파(carrier wave) 또는 기타 전송 메커니즘(transport mechanism)과 같은 피변조 데이터 신호(modulated data signal)에 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터 등을 구현하고 모든 정보 전달 매체를 포함한다. 피변조 데이터 신호라는 용어는 신호 내에 정보를 인코딩하도록 그 신호의 특성들 중 하나 이상을 설정 또는 변경시킨 신호를 의미한다. 제한이 아닌 예로서, 컴퓨터 판독가능 전송 매체는 유선 네트워크 또는 직접 배선 접속(direct-wired connection)과 같은 유선 매체, 그리고 음향, RF, 적외선, 기타 무선 매체와 같은 무선 매체를 포함한다. 상술된 매체들 중 임의의 것의 조합도 역시 컴퓨터 판독가능 전송 매체의 범위 안에 포함되는 것으로 한다.

[0142] 컴퓨터(1102)를 포함하는 본 개시의 여러가지 측면들을 구현하는 예시적인 환경(1100)이 나타내어져 있으며, 컴퓨터(1102)는 처리 장치(1104), 시스템 메모리(1106) 및 시스템 버스(1108)를 포함한다. 시스템 버스(1108)는 시스템 메모리(1106)(이에 한정되지 않음)를 비롯한 시스템 컴포넌트들을 처리 장치(1104)에 연결시킨다. 처리 장치(1104)는 다양한 상용 프로세서들 중 임의의 프로세서일 수 있다. 듀얼 프로세서 및 기타 멀티프로세서 아키텍처도 역시 처리 장치(1104)로서 이용될 수 있다.

[0143] 시스템 버스(1108)는 메모리 버스, 주변장치 버스, 및 다양한 상용 버스 아키텍처 중 임의의 것을 사용하는 로컬 버스에 추가적으로 상호 연결될 수 있는 몇 가지 유형의 버스 구조 중 임의의 것일 수 있다. 시스템 메모리(1106)는 판독 전용 메모리(ROM)(1110) 및 랜덤 액세스 메모리(RAM)(1112)를 포함한다. 기본 입/출력 시스템(BIOS)은 ROM, EPROM, EEPROM 등의 비휘발성 메모리(1110)에 저장되며, 이 BIOS는 시동 중과 같은 때에 컴퓨터(1102) 내의 구성요소들 간에 정보를 전송하는 일을 돕는 기본적인 루틴을 포함한다. RAM(1112)은 데이터를 캐싱하기 위한 정적 RAM 등의 고속 RAM을 포함할 수 있다.

[0144] 컴퓨터(1102)는 내장형 하드 디스크 드라이브(HDD)(1114)(예를 들어, EIDE, SATA)-이 내장형 하드 디스크 드라

이브(1114)는 적당한 섀시(도시 생략) 내에서 외장형 용도로 구성될 수 있음-, 자기 플로피 디스크 드라이브(FDD)(1116)(예를 들어, 이동식 디스켓(1118)으로부터 판독을 하거나 그에 기록을 하기 위한 것임), 및 광 디스크 드라이브(1120)(예를 들어, CD-ROM 디스크(1122)를 판독하거나 DVD 등의 기타 고용량 광 매체로부터 판독을 하거나 그에 기록을 하기 위한 것임)를 포함한다. 하드 디스크 드라이브(1114), 자기 디스크 드라이브(1116) 및 광 디스크 드라이브(1120)는 각각 하드 디스크 드라이브 인터페이스(1124), 자기 디스크 드라이브 인터페이스(1126) 및 광 드라이브 인터페이스(1128)에 의해 시스템 버스(1108)에 연결될 수 있다. 외장형 드라이브 구현을 위한 인터페이스(1124)는 USB(Universal Serial Bus) 및 IEEE 1394 인터페이스 기술 중 적어도 하나 또는 그 둘 다를 포함한다.

[0145] 이들 드라이브 및 그와 연관된 컴퓨터 판독가능 매체는 데이터, 데이터 구조, 컴퓨터 실행가능 명령어, 기타 등등의 비휘발성 저장을 제공한다. 컴퓨터(1102)의 경우, 드라이브 및 매체는 임의의 데이터를 적당한 디지털 형식으로 저장하는 것에 대응한다. 상기에서의 컴퓨터 판독가능 매체에 대한 설명이 HDD, 이동식 자기 디스크, 및 CD 또는 DVD 등의 이동식 광 매체를 언급하고 있지만, 통상의 기술자라면 집 드라이브(zip drive), 자기 카세트, 플래쉬 메모리 카드, 카트리지, 기타 등등의 컴퓨터에 의해 판독가능한 다른 유형의 매체도 역시 예시적인 운영 환경에서 사용될 수 있으며 또 임의의 이러한 매체가 본 개시의 방법들을 수행하기 위한 컴퓨터 실행가능 명령어를 포함할 수 있다는 것을 잘 알 것이다.

[0146] 운영 체제(1130), 하나 이상의 애플리케이션 프로그램(1132), 기타 프로그램 모듈(1134) 및 프로그램 데이터(1136)를 비롯한 다수의 프로그램 모듈이 드라이브 및 RAM(1112)에 저장될 수 있다. 운영 체제, 애플리케이션, 모듈 및/또는 데이터의 전부 또는 그 일부분이 RAM(1112)에 캐싱될 수 있다. 본 개시가 여러가지 상업적으로 이용가능한 운영 체제 또는 운영 체제들의 조합에서 구현될 수 있다는 것을 잘 알 것이다.

[0147] 사용자는 하나 이상의 유선/무선 입력 장치, 예를 들어, 키보드(1138) 및 마우스(1140) 등의 포인팅 장치를 통해 컴퓨터(1102)에 명령 및 정보를 입력할 수 있다. 기타 입력 장치(도시 생략)로는 마이크, IR 리모콘, 조이스틱, 게임 패드, 스타일러스 펜, 터치 스크린, 기타 등등이 있을 수 있다. 이들 및 기타 입력 장치가 종종 시스템 버스(1108)에 연결되어 있는 입력 장치 인터페이스(1142)를 통해 처리 장치(1104)에 연결되지만, 병렬 포트, IEEE 1394 직렬 포트, 게임 포트, USB 포트, IR 인터페이스, 기타 등등의 기타 인터페이스에 의해 연결될 수 있다.

[0148] 모니터(1144) 또는 다른 유형의 디스플레이 장치도 역시 비디오 어댑터(1146) 등의 인터페이스를 통해 시스템 버스(1108)에 연결된다. 모니터(1144)에 부가하여, 컴퓨터는 일반적으로 스피커, 프린터, 기타 등등의 기타 주변 출력 장치(도시 생략)를 포함한다.

[0149] 컴퓨터(1102)는 유선 및/또는 무선 통신을 통한 원격 컴퓨터(들)(1148) 등의 하나 이상의 원격 컴퓨터로의 논리적 연결을 사용하여 네트워크화된 환경에서 동작할 수 있다. 원격 컴퓨터(들)(1148)는 워크스테이션, 컴퓨팅 디바이스 컴퓨터, 라우터, 퍼스널 컴퓨터, 휴대용 컴퓨터, 마이크로프로세서-기반 오락 기기, 피어 장치 또는 기타 통상의 네트워크 노드일 수 있으며, 일반적으로 컴퓨터(1102)에 대해 기술된 구성요소들 중 다수 또는 그 전부를 포함하지만, 간략함을 위해, 메모리 저장 장치(1150)만이 도시되어 있다. 도시되어 있는 논리적 연결은 근거리 통신망(LAN)(1152) 및/또는 더 큰 네트워크, 예를 들어, 원거리 통신망(WAN)(1154)에의 유선/무선 연결을 포함한다. 이러한 LAN 및 WAN 네트워킹 환경은 사무실 및 회사에서 일반적인 것이며, 인트라넷 등의 전사적 컴퓨터 네트워크(enterprise-wide computer network)를 용이하게 해주며, 이들 모두는 전세계 컴퓨터 네트워크, 예를 들어, 인터넷에 연결될 수 있다.

[0150] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 유선 및/또는 무선 통신 네트워크 인터페이스 또는 어댑터(1156)를 통해 로컬 네트워크(1152)에 연결된다. 어댑터(1156)는 LAN(1152)에의 유선 또는 무선 통신을 용이하게 해줄 수 있으며, 이 LAN(1152)은 무선 어댑터(1156)와 통신하기 위해 그에 설치되어 있는 무선 액세스 포인트를 포함하고 있다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 모뎀(1158)을 포함할 수 있거나, WAN(1154) 상의 통신 컴퓨팅 디바이스에 연결되거나, 또는 인터넷을 통하는 등, WAN(1154)을 통해 통신을 설정하는 기타 수단을 갖는다. 내장형 또는 외장형 및 유선 또는 무선 장치일 수 있는 모뎀(1158)은 직렬 포트 인터페이스(1142)를 통해 시스템 버스(1108)에 연결된다. 네트워크화된 환경에서, 컴퓨터(1102)에 대해 설명된 프로그램 모듈들 또는 그의 일부분이 원격 메모리/저장 장치(1150)에 저장될 수 있다. 도시된 네트워크 연결이 예시적인 것이며 컴퓨터들 사이에 통신 링크를 설정하는 기타 수단이 사용될 수 있다는 것을 잘 알 것이다.

[0151] 컴퓨터(1102)는 무선 통신으로 배치되어 동작하는 임의의 무선 장치 또는 개체, 예를 들어, 프린터, 스캐너, 데스크톱 및/또는 휴대용 컴퓨터, PDA(portable data assistant), 통신 위성, 무선 검출가능 태그와 연관된 임의

의 장비 또는 장소, 및 전화와 통신을 하는 동작을 한다. 이것은 적어도 Wi-Fi 및 블루투스 무선 기술을 포함한다. 따라서, 통신은 종래의 네트워크에서와 같이 미리 정의된 구조이거나 단순하게 적어도 2개의 장치 사이의 애드혹 통신(ad hoc communication)일 수 있다.

[0152] Wi-Fi(Wireless Fidelity)는 유선 없이도 인터넷 등으로의 연결을 가능하게 해준다. Wi-Fi는 이러한 장치, 예를 들어, 컴퓨터가 실내에서 및 실외에서, 즉 기지국의 통화권 내의 아무 곳에서나 데이터를 전송 및 수신할 수 있게 해주는 셀 전화와 같은 무선 기술이다. Wi-Fi 네트워크는 안전하고 신뢰성 있으며 고속인 무선 연결을 제공하기 위해 IEEE 802.11(a, b, g, 기타)이라고 하는 무선 기술을 사용한다. 컴퓨터를 서로에, 인터넷에 및 유선 네트워크(IEEE 802.3 또는 이더넷을 사용함)에 연결시키기 위해 Wi-Fi가 사용될 수 있다. Wi-Fi 네트워크는 비인가 2.4 및 5GHz 무선 대역에서, 예를 들어, 11Mbps(802.11a) 또는 54 Mbps(802.11b) 데이터 레이트로 동작하거나, 양 대역(듀얼 대역)을 포함하는 제품에서 동작할 수 있다.

[0153] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 정보 및 신호들이 임의의 다양한 상이한 기술들 및 기법들을 이용하여 표현될 수 있다는 것을 이해할 것이다. 예를 들어, 위의 설명에서 참조될 수 있는 데이터, 지시들, 명령들, 정보, 신호들, 비트들, 심볼들 및 칩들은 전압들, 전류들, 전자기파들, 자기장들 또는 입자들, 광학장들 또는 입자들, 또는 이들의 임의의 결합에 의해 표현될 수 있다.

[0154] 본 개시의 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 개시의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식들로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 개시의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

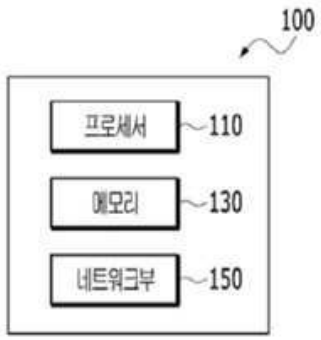
[0155] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조 물품(article)으로 구현될 수 있다. 용어 제조 물품은 임의의 컴퓨터-판독가능 저장장치로부터 액세스 가능한 컴퓨터 프로그램, 캐리어, 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트립, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 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 개시의 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.

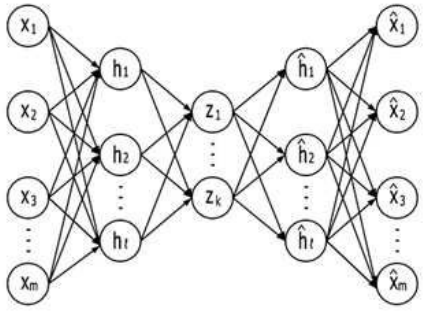
[0157] 제시된 실시예들에 대한 설명은 임의의 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

도면

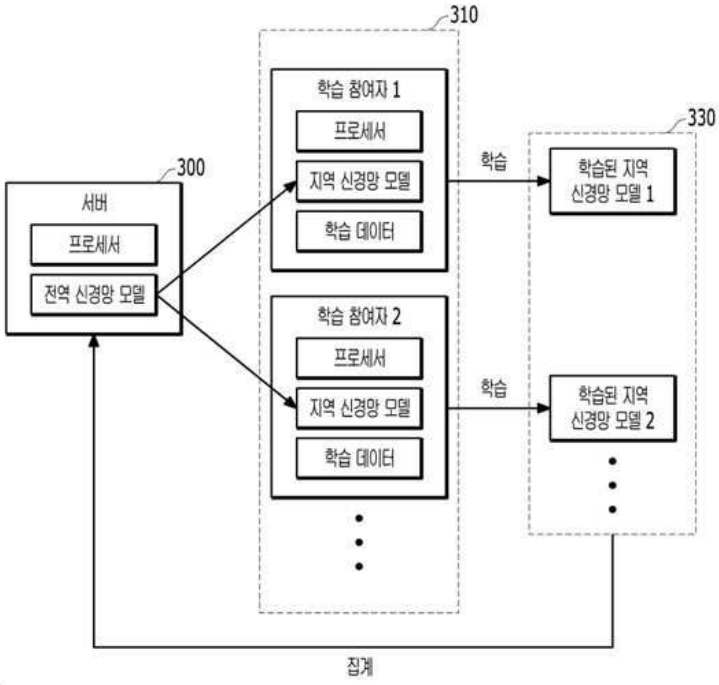
도면1


100
프로세서
110
메모리
130
네트워크부
150

도면2


x1
x2
x3
xm
h1
h2
ht
z1
zk
ĥ1
ĥ2
ĥt
x̂1
x̂2
x̂3
x̂m

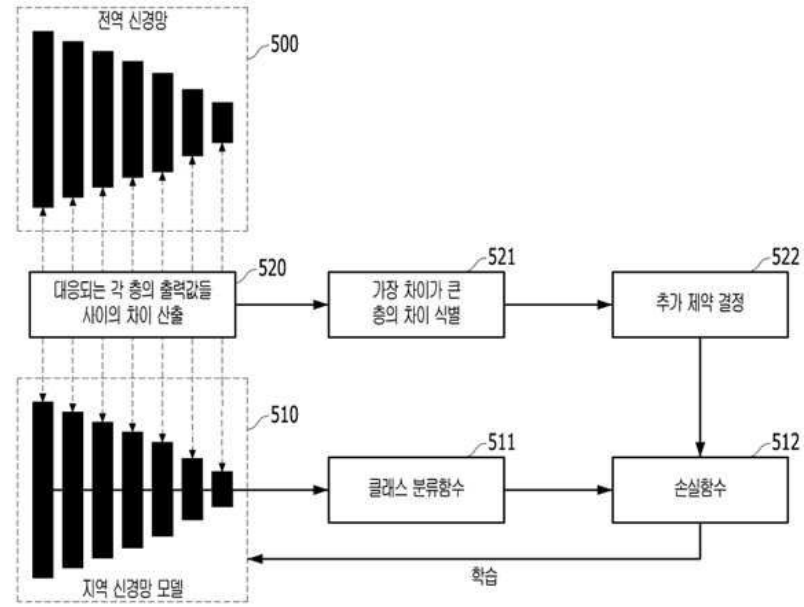
도면3


310
학습 참여자 1
프로세서
지역 신경망 모델
학습 데이터
330
300
서버
프로세서
전역 신경망 모델
학습
학습된 지역
신경망 모델 1
학습 참여자 2
프로세서
지역 신경망 모델
학습 데이터
학습
학습된 지역
신경망 모델 2
집계

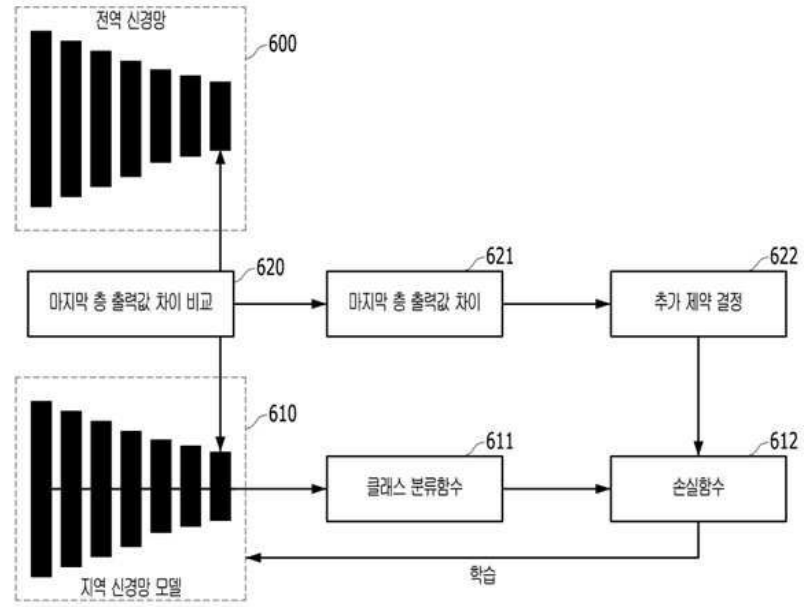
도면4


시작
S401
전역(global) 신경망 모델과 지역 신경망
모델 사이의 차이를 산출하는 단계
S402
상기 산출된 차이에 기초하여 상기 지역
신경망 모델의 학습을 위한 추가
제약(regularization)을 결정하는 단계
S403
상기 결정된 추가 제약을 포함하는 손실함수에
기초하여 상기 지역 신경망 모델을 학습시키는 단계
종료

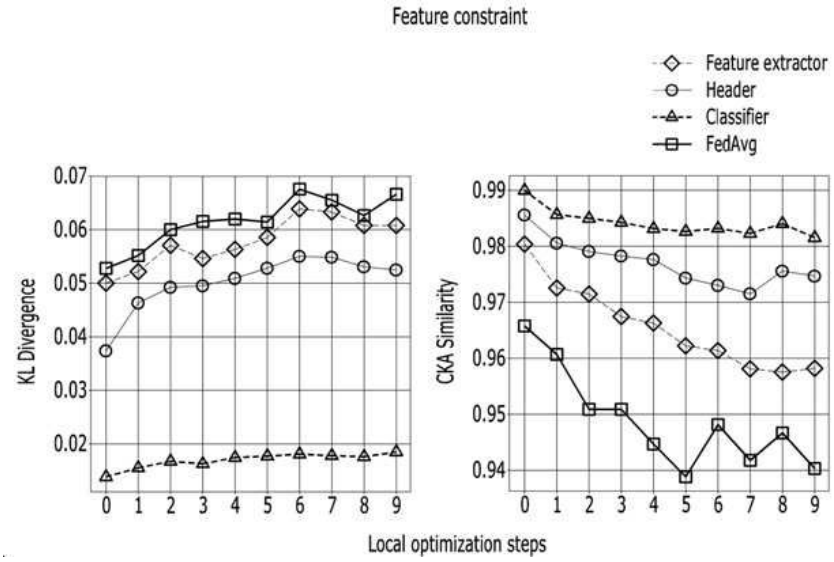
도면5


전역 신경망
500
520
대응되는 각 층의 출력값들 사이의 차이 산출
521
가장 차이가 큰 층의 차이 식별
522
추가 제약 결정
510
511
클래스 분류함수
512
손실함수
학습
지역 신경망 모델

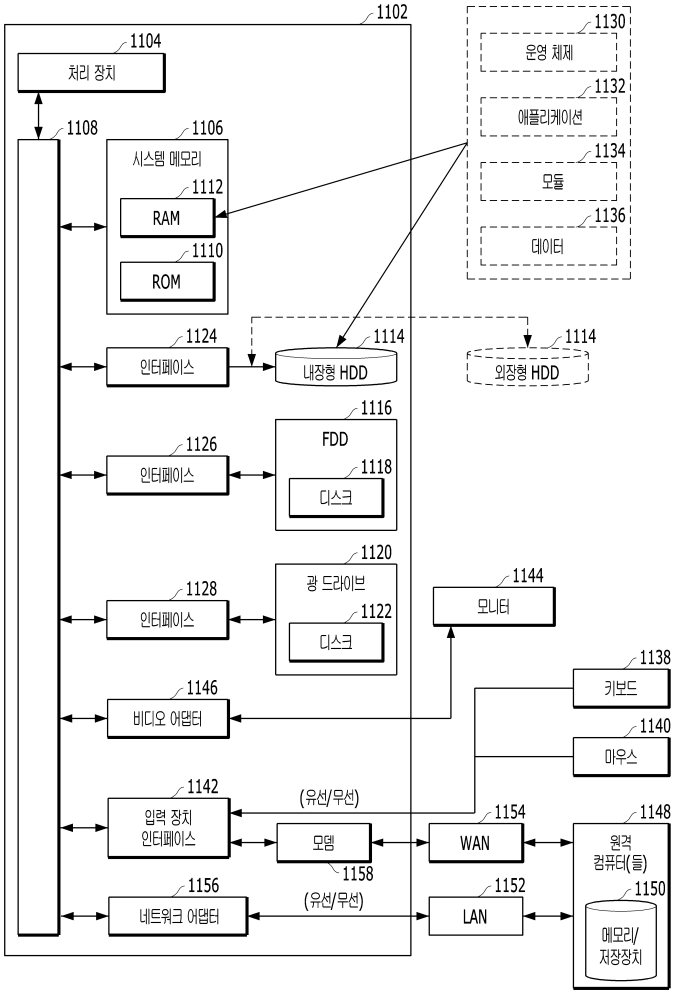
도면6


전역 신경망
600
620
마지막 층 출력값 차이 비교
621
마지막 층 출력값 차이
622
추가 제약 결정
610
611
클래스 분류함수
612
손실함수
학습
지역 신경망 모델

도면7


Feature constraint
Feature extractor
Header
Classifier
FedAvg
KL Divergence
CKA Similarity
0.07
0.06
0.05
0.04
0.03
0.02
0.99
0.98
0.97
0.96
0.95
0.94
0 1 2 3 4 5 6 7 8 9
Local optimization steps

도면8


1102
1104
처리 장치
1108
1106
시스템 메모리
1112
RAM
1110
ROM
1130
운영 체제
1132
애플리케이션
1134
모듈
1136
데이터
1124
인터페이스
1114
내장형 HDD
1114
외장형 HDD
1116
FDD
1118
디스크
1126
인터페이스
1120
광 드라이브
1122
디스크
1128
인터페이스
1144
모니터
1138
키보드
1140
마우스
1146
비디오 어댑터
1142
입력 장치 인터페이스
(유선/무선)
1154
WAN
1148
원격 컴퓨터(들)
1158
모뎀
1156
네트워크 어댑터
(유선/무선)
1152
LAN
1150
메모리/저장장치