(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0028811
(43) 공개일자 2025년03월04일

(51) 국제특허분류(Int. Cl.)
G06N 3/098 (2023.01) *G06N 3/096* (2023.01)
G06N 3/0985 (2023.01)
(52) CPC특허분류
G06N 3/098 (2023.01)
G06N 3/096 (2023.01)
(21) 출원번호 10-2023-0109939
(22) 출원일자 2023년08월22일
심사청구일자 2023년08월22일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이상근
서울특별시 강남구 선릉로 221(도곡동, 도곡렉슬아파트)
김예찬
경기도 남양주시 늘을1로 55 - 18 (호평동 삼우에이치타운빌라2차) 204동 01호
(74) 대리인
윤귀상

전체 청구항 수 : 총 8 항

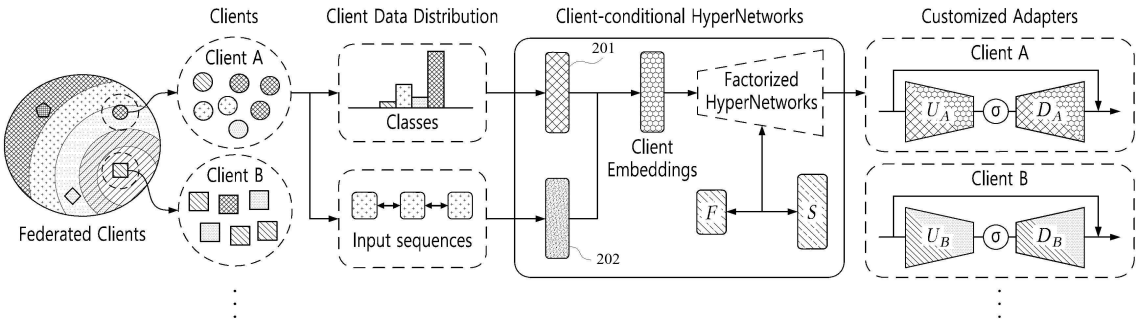
(54) 발명의 명칭 **사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템**

(57) 요 약

본 발명의 일 실시예는 연합학습 환경에서의 클라이언트 디바이스에서 수행되는 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 방법으로서, 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력 시퀀스를 생성하는 단계; 상기 클라이언트 데이터 분포와 상기 입력 시퀀스를 하이퍼네트워크에 입력하는 단계; 상기 하이퍼네

(뒷면에 계속)

대 표 도 - 도3



트워크를 통해 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력시퀀스로부터 특징벡터를 추출하고 결합하여 결합벡터를 생성하는 단계; 상기 결합벡터를 기초로 클라이언트 커스토마이징된 네트워크 파라미터를 생성하는 단계; 및 상기 하이퍼네트워크로부터 출력된 클라이언트 커스토마이징된 네트워크 파라미터를 기초로 상기 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT) 모델을 생성하는 단계를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템을 제공한다.

(52) CPC특허분류
G06N 3/0985 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193973
과제번호	2019-0-00079-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(고려대학교)
기 여 율	1/2
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711188552
과제번호	2021R1A2C3010430
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	미플랫폼: 자기지도학습과 연합학습으로 강화되는 작은 인공지능 플랫폼
기 여 율	1/2
과제수행기관명	고려대학교
연구기간	2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

연합학습 환경에서의 클라이언트 디바이스에서 수행되는 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 방법으로서,

클라이언트 데이터 분포와 레이어 정보를 포함하는 입력 시퀀스를 생성하는 단계;

상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력 시퀀스를 하이퍼네트워크에 입력하는 단계;

상기 하이퍼네트워크를 통해 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력시퀀스로부터 특징벡터를 추출하고 결합하여 결합벡터를 생성하는 단계;

상기 결합벡터를 기초로 클라이언트 커스토마이징된 네트워크 파라미터를 생성하는 단계; 및

상기 하이퍼네트워크로부터 출력된 클라이언트 커스토마이징된 네트워크 파라미터를 기초로 상기 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT) 모델을 생성하는 단계를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법.

청구항 2

청구항 1에 있어서,

상기 하이퍼네트워크는 연합학습에 참여하는 각 클라이언트 디바이스가 가지고 있는 상기 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT)를 생성하도록 학습되는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법.

청구항 3

청구항 2에 있어서,

상기 결합벡터를 인수분해된 하이퍼네트워크에 입력하여 상기 클라이언트 커스토마이징된 네트워크 파라미터가 생성되는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법.

청구항 4

청구항 3에 있어서,

상기 클라이언트 커스토마이징된 네트워크 파라미터는 어댑터 파라미터를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법.

청구항 5

청구항 4에 있어서,

상기 어댑터 파라미터는 병목(Bottleneck) 구조를 갖는 두 개의 선형 레이어로 구성되며, 각 선형 레이어의 파라미터는 2차원의 웨이트(Weight) 파라미터와 1차원의 바이어스(Bias) 파라미터로 구성된 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법.

청구항 6

연합학습 환경에서의 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 적어도 하나 이상의 클라이언트 디바이스로서, 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 생성하고, 상기 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 하이퍼네트워크에 입력하고, 상기 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 기초로 상기 하이퍼네트워크로부터 제공된 어댑터 파라미터를 이용하여, 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT) 모델을 생성하는, 상기 클라이언트 디바이스; 및

상기 클라이언트 디바이스의 클라이언트 데이터에 대한 훈련이 완료되면 각각의 상기 커스토마이징된 파라미터 효율 파인 튜닝 모델을 전송받아 글로벌 모델을 업데이트하는 중앙 집중식 서버를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템.

청구항 7

청구항 6에 있어서,

상기 클라이언트 디바이스는 상기 하이퍼네트워크에 입력된 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력시퀀스로부터 특징벡터를 추출하고 결합하여 결합벡터를 생성하고, 상기 결합벡터를 기초로 클라이언트 커스토마이징된 상기 어댑터 파라미터를 생성하며,

상기 클라이언트 디바이스는 상기 하이퍼네트워크로부터 출력된 상기 어댑터 파라미터를 기초로 상기 PEFT 모델을 생성하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템.

청구항 8

청구항 7에 있어서,

상기 클라이언트 디바이스는 상기 글로벌 모델을 업데이트하기 위해 상기 하이퍼네트워크 및 레이어 인덱스 임베딩과 관련된 매개변수를 상기 중앙 집중식 서버로 전송하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템에 관한 것이며, 보다 상세하게는 모바일 디바이스, IoT 기기 등 자원의 제약이 있는 환경에서 효율적으로 연합학습을 하기 위한 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템에 관한 것이다.

배 경 기 술

[0003] 일반적으로, 연합학습(Federated Learning)은 여러 사용자가 연합해서 하나의 글로벌한 신경망(Neural Network)을 학습하는 기술이다. 특히, 연합학습은 참여하는 사용자의 데이터 프라이버시 보존을 위하여, 하나의 서버에서 학습을 하는 것이 아닌, 사용자 각각의 디바이스에서 신경망을 학습시키고 학습된 모델을 서버에서 취합하는 방식으로 학습이 구성된다. 이 때, 서버에 사용자의 데이터 대신 사용자의 학습된 모델을 보냄으로써 사용자 데이터 프라이버시를 보존한다.

[0004] 이러한 연합학급에서 파라미터-효율적 미세조정(Parameter-efficient fine-tuning, 이하 PEFT)기술로 거대언어모델(Large Language Model)을 주어진 작업에 효율적으로 미세조정 하는 방법이 사용되고 있다. 구체적으로 거대언어모델에 적은 수의 파라미터를 추가하고 학습 시에 추가된 파라미터만을 미세조정하여 주어진 작업을 학습

하게 된다.

[0005] 대표적인 PEFT 기술로는 어댑터(Adapter) 가 있다. 이는 언어모델에 병목구조를 가지고 있는 신경망(Neural Network)을 추가하고, 학습 시에는 언어모델을 구성하는 전체 파라미터가 아닌 어댑터를 구성하는 추가된 파라미터만을 학습시키는 기술이다. 이러한 어댑터는 병목구조를 구성하기 위한 인코더-디코더(Encoder-Decoder)구조의 신경망을 이루고 있으며, 언어모델의 셀프어텐션(Self-attention)계층과 위치별 피드포워드(Position-wise Feed-Forward)계층 사이에 위치하고 있다.

[0006] 그러나, 이러한 거대언어모델이 연합학습에서 사용되는 경우 비효율성이 발생하는 문제가 있다. 즉, 연합학습은 서버와 클라이언트 디바이스 간에 모델 가중치를 전송하여 데이터 프라이버시를 보존하는데, 이 때 거대언어모델의 높은 메모리 요구사항으로 인해 리소스가 제한된 환경에서의 적용이 어려운 현실이다. 또한, 연합학습에서 모델을 클라이언트 디바이스 간에 전송하는데 큰 연산 자원과 대역폭이 필요로 하여, 고사양의 자원을 포함한 기기만 거대언어모델의 효과성을 얻을 수 있는 문제가 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명이 이루고자 하는 기술적 과제는 FL(연합학습)의 맥락에서 PEFT(매개 변수 효율적 미세 조정)를 사용하되, 일반적인 PEFT는 FL 시나리오에서 학습데이터 간의 이질성으로 인해 불안정하고 느린 수렴을 초래하는 경향이 있는 문제점을 극복하기 위해 각 클라이언트 데이터(학습데이터)를 컨디셔닝하여 클라이언트 디바이스별 어댑터를 생성하는 새로운 하이퍼네트워크 기반 FL 프레임워크인 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템을 제공하는 것이다.

[0009] 본 발명이 이루고자 하는 기술적 과제는 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0011] 상기 기술적 과제를 달성하기 위하여, 본 발명의 일 실시예는 연합학습 환경에서의 클라이언트 디바이스에서 수행되는 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 방법으로서, 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력 시퀀스를 생성하는 단계; 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력 시퀀스를 하이퍼네트워크에 입력하는 단계; 상기 하이퍼네트워크를 통해 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력시퀀스로부터 특징벡터를 추출하고 결합하여 결합벡터를 생성하는 단계; 상기 결합벡터를 기초로 클라이언트 커스토마이징된 네트워크 파라미터를 생성하는 단계; 및 상기 하이퍼네트워크로부터 출력된 클라이언트 커스토마이징된 네트워크 파라미터를 기초로 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT) 모델을 생성하는 단계를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법을 제공한다.

[0012] 본 발명의 실시예에 있어서, 상기 하이퍼네트워크는 연합학습에 참여하는 각 클라이언트 디바이스가 가지고 있는 상기 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT)를 생성하도록 학습될 수 있다.

[0013] 본 발명의 실시예에 있어서, 상기 결합벡터를 인수분해된 하이퍼네트워크에 입력하여 상기 클라이언트 커스토마이징된 네트워크 파라미터를 생성할 수 있다.

[0014] 본 발명의 실시예에 있어서, 상기 클라이언트 커스토마이징된 네트워크 파라미터는 어댑터 파라미터를 포함할 수 있다.

[0015] 본 발명의 실시예에 있어서, 상기 어댑터 파라미터는 병목(Bottleneck) 구조를 갖는 두 개의 선형 레이어로 구성되며, 각 선형 레이어의 파라미터는 2차원의 웨이트(Weight) 파라미터와 1차원의 바이어스(Bias) 파라미터로 구성될 수 있다.

[0016] 상기 기술적 과제를 달성하기 위하여, 본 발명의 다른 실시예는 연합학습 환경에서의 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 적어도 하나 이상의 클라이언트 디바이스로서, 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 생성하고, 상기 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 하이퍼네트워크에 입력하고, 상기 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 기초로 상기 하이퍼네트워크로부터 제공된 어댑터 파라미터를 이용하여, 클라이언트 데이터 분포에 적합한 파라미터 효율 파인 튜닝(Parameter-Efficient Fine-tuning, PEFT) 모델을 생성하는 상기 클라이언트 디바이스; 및 상기 클라이언트 디바이스의 클라이언트 데이터에 대한 훈련이 완료되면 각각의 상기 커스토마이징된 파라미터 효율 파인 튜닝 모델을 전송받아 글로벌 모델을 업데이트하는 중앙 집중식 서버를 포함하는 것을 특징으로 하는, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템을 제공한다.

[0017] 본 발명의 실시예에 있어서, 상기 클라이언트 디바이스는 상기 하이퍼네트워크에 입력된 상기 클라이언트 데이터 분포와 상기 레이어 정보를 포함하는 입력시퀀스로부터 특징벡터를 추출하고 결합하여 결합벡터를 생성하고, 상기 결합벡터를 기초로 클라이언트 커스토마이징된 상기 어댑터 파라미터를 생성하며, 상기 클라이언트 디바이스는 상기 하이퍼네트워크로부터 출력된 상기 어댑터 파라미터를 기초로 상기 PEFT 모델을 생성할 수 있다.

[0018] 본 발명의 실시예에 있어서, 상기 클라이언트 디바이스는 상기 글로벌 모델을 업데이트하기 위해 상기 하이퍼네트워크 및 레이어 인덱스 임베딩과 관련된 매개변수를 상기 중앙 집중식 서버로 전송할 수 있다.

발명의 효과

[0020] 본 발명의 실시예에 의하면, 기존의 언어모델의 연합학습에서 적은 수의 파라미터만을 학습시키고, 이를 연합학습에서 서버와의 통신에 사용한다. 이에 따라, 언어모델을 활용하는 연합학습이 가지는 근본적인 계산적, 메모리 측면의 비효율성을 크게 개선시킬 수 있는 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템을 제공할 수 있다.

[0021] 또한, 기존의 연합학습 방법론의 취약점으로 알려져 있는 데이터의 비항등분포 환경(Non-IID, Non-Independent and Identical Distributuion)에서도 높은 강건성을 얻을 수 있는 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법 및 이를 위한 시스템을 제공할 수 있다.

[0022] 본 발명의 효과는 상기한 효과로 한정되는 것은 아니며, 본 발명의 상세한 설명 또는 특허청구범위에 기재된 발명의 구성으로부터 추론 가능한 모든 효과를 포함하는 것으로 이해되어야 한다.

도면의 간단한 설명

[0024] 도 1은 단일 글로벌 어댑터를 모든 이질적 데이터 분포에 단순히 적응시키는 기존 PEFT의 개념도이다.

도 2는 본 발명의 일 실시예에 따른 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템을 나타낸다.

도 3은 도 2에서 제시된 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템의 어댑터 생성 프로세스의 일 예를 나타낸다.

도 4는 본 발명의 일 실시예에 따른 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법을 나타낸다.

도 5는 본 발명의 실시예의 비항등분포 연합학습 환경에서의 정량적 실험결과를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0025] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 본문에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 개시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0026] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의

미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0027] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다.

[0028] 도 1은 단일 글로벌 어댑터를 모든 이질적 데이터 분포에 단순히 적응시키는 기존 PEFT의 개념도이다. 도 2는 본 발명의 일 실시예에 따른 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템을 나타낸다.

[0029] 도 2를 참조하면, 본 실시예의 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템은 중앙 집중식 서버(100)와 하나 이상의 클라이언트 디바이스(200)를 포함한다. 연합학습(FL)의 목표는 클라이언트 디바이스(200) 간에 개인 데이터를 공유하지 않고 단일 글로벌 모델을 공동으로 훈련하는 것이다. 이를 위해 FL은 클라이언트 디바이스(200)와 중앙 집중식 서버(100)(이하, '서버') 간의 학습 모델 통신을 라운드 바이 라운드 방식으로 진행할 수 있다. 각 라운드에서 서버(100)는 먼저 단일 글로벌 모델 Θ를 샘플 클라이언트 디바이스 집합에 배포하고 참여 클라이언트 디바이스(200)는 자체 데이터(학습데이터)에 대해 로컬 최적화를 수행한다. 최적화가 완료되면 서버(100)는 로컬에서 훈련된 모든 로컬 모델을 다시 집계하여 글로벌 모델을 업데이트한다.

[0030] i번째 클라이언트 디바이스(200)의 데이터 세트를 Di라고 하면 위의 글로벌 모델 업데이트 프로세스는 다음과 같이 공식화할 수 있다.

[0031] $$\widetilde{\theta} = \sum_{i=1}^{K} \alpha_i \cdot \mathcal{L}(\mathcal{D}_i; \theta) \quad (1)$$

[0032] 여기서 L(Di ; Θ)는 주어진 데이터 세트와 초기 모델을 기반으로 훈련된 모델을 반환하는 함수이고, K는 참여 클라이언트 수이며 αi는 글로벌 모델을 구축하기 위한 클라이언트 i의 기여 요소이다. αi는 일반적으로 각 클라이언트의 데이터 세트 크기, 즉 $\alpha_i = \frac{|\mathcal{D}_i|}{\sum_i |\mathcal{D}_i|}$ 에 의해 결정된다.

[0033] 그러나 FL의 통신 프로세스에 번거로운 PLM을 사용하는 것은 두 가지 문제를 야기한다. 첫째, 함수 L(·)은 PLM과 관련된 훈련 가능한 매개변수가 많기 때문에 높은 컴퓨팅 리소스가 필요하다. 둘째, 집계 단계(즉, 가중 합산)에서 모델을 송수신하는 데 상당한 네트워크 대역폭이 필요하다. 따라서 이러한 제약을 완화할 수 있는 최적의 솔루션을 찾아, PLM을 사용하는 FL에 대해 보다 효율적이고 자원 집약적인 메커니즘을 제공하는 것이 중요하다.

[0034] 본 발명의 실시예에 따른 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템은 전술된 제약을 극복하기 위해 각 클라이언트의 다양한 데이터 분포 정보를 활용하는 새로운 하이퍼네트워크 기반 FL 프레임워크를 제공한다. 본 발명은 도 1에 도시된 바와 같은 단일 글로벌 어댑터(A)를 모든 클라이언트 데이터 분포에 단순히 맞추는 것이 아니라, 도 2에 도시된 바와 같이 클라이언트 데이터 분포 정보를 가져와 하이퍼네트워크(H)를 통해 각 클라이언트에 맞는 어댑터 매개변수(A_A, A_C)를 생성하는 방법으로 전술된 제약을 극복할 수 있다.

[0035] 하이퍼네트워크(HyperNetworks)는 다른 네트워크의 가중치를 생성하는 네트워크(일반적으로 신경망)이다. 각각에 대해 별도의 모델을 교육할 필요 없이 다양한 작업 또는 데이터 세트에 대한 가중치를 동적으로 생성하는 데 사용할 수 있다. 본 실시예에서 서버는 글로벌 모델을 업데이트하기 위해 다른 클라이언트의 학습 결과(로컬 모델)를 집계하고 통합하는 데 있어서 하이퍼네트워크를 이용할 수 있다. 즉, 서버(100) 측에서 하이퍼네트워크를 사용하여 어댑터 매개변수를 파생한 다음 로컬 모델을 조정하기 위해 클라이언트 디바이스(200)로 다시 전송할 수 있다.

[0036] 이 시스템에서 서버(100)는 클라이언트 디바이스(200)에 전역 모델을 배포하며, 클라이언트 디바이스(200)는 자체 데이터(클라이언트 데이터)에 대해 로컬 최적화를 수행할 수 있다. 최적화가 완료되면 서버(100)는 로컬에서 훈련된 로컬 모델을 집계하여 글로벌 모델을 업데이트할 수 있다. 서버(100)에 의해 배포된 글로벌 모델은 특정 하이퍼파라미터를 사용하여 구성될 수 있으며, 이는 각 클라이언트 디바이스(200)가 각 클라이언트 데이터를 학습하는데 사용될 수 있다. 이러한 하이퍼파라미터에는 학습 속도, 배치 크기, 정규화 용어 등이 포함될 수 있다. 서버(100)는 또한 각 라운드에서 샘플링할 클라이언트 수, 집계 방법 등과 같은 연합학습 자체와 관련된 하이퍼파라미터를 관리할 수 있다. 집계 중에 서버(100)는 하이퍼파라미터를 활용하여 다른 클라이언트 디바이스(200)에서 학습된 모델의 기여도를 평가할 수 있다.

[0037] 도 3은 도 2에서 제시된 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템의 어댑터 생성 프로세스의

일 예를 나타낸다.

[0038] 도 3을 참조하면, 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 시스템에서, 클라이언트 디바이스(200)는 제공되거나 저장되어 있는 클라이언트 데이터(학습데이터)로부터 클라이언트 데이터 분포와, 레이어 정보를 포함하는 입력시퀀스를 추출하고, 이를 하이퍼네트워크에 입력할 수 있다. 본 실시예에서는 클라이언트 데이터 간에 존재하는 이질성으로 인한 부정적인 영향을 완화하기 위해 각 클라이언트 데이터에 맞는 맞춤형 PEFT 모듈을 생성하는 것이 필요하다. 이를 달성하기 위해 각 클라이언트 디바이스(200)는 클라이언트 데이터 분포와 입력시퀀스로부터 각각 특징벡터(201,202)를 추출하고 이들을 결합할 수 있다. 결합된 특징벡터를 미리 마련된 하이퍼네트워크(Hyper Network)에 입력할 수 있다. 하이퍼네트워크에 입력된 상기 특징벡터를 조건으로 하여 특정한 네트워크의 파라미터를 생성하도록 학습될 수 있다.. 어댑터파라미터를 생성하기 위해서, 하이퍼네트워크로부터 출력된 파라미터를 기초로 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성한다. 본 실시예는 연합학습에 참여하는 각 클라이언트 디바이스(200)가 가지고 있는 클라이언트 데이터 분포에 적합한 Parameter-Efficient Fine-tuning (이하, PEFT)를 생성하는 네트워크를 학습시킨다. 이를 달성하기 위해서는 생성할 PEFT 모듈에 대한 선택과 선택된 PEFT 모듈의 파라미터를 생성하는 것이 필요하다. 연합학습 과정에서 본 실시예는 생성할 PEFT 모듈로써 어댑터구조 설정한다. 이는 Bottleneck 구조를 갖는 두 개의 선형 레이어로 구성이 되어 있으며, 각 선형 레이어의 파라미터는 2차원의 웨이트(Weight) 파라미터와 1차원의 바이어스(Bias) 파라미터로 구성되어 있다. 본 발명은 연합학습에 참여하는 클라이언트의 데이터 분포에 따른 파라미터를 만드는 것이 목적이므로, 데이터 분포를 벡터형태로 나타내고 이를 하이퍼네트워크에 입력으로 하여 어댑터 파라미터(상기한 병목구조를 가진 선형변환에 사용되는 웨이트와 바이어스 파라미터를 의미함)를 생성할 수 있다.

[0039] 도 4는 본 발명의 일 실시예에 따른 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법을 나타낸다. 본 실시예의 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법은 연합학습 환경에서의 클라이언트 디바이스(200)에서 수행되는 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성하는 방법에 적용될 수 있다.

[0040] 본 실시예에서 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법은 연합학습에 참여하는 각 클라이언트가 가지고 있는 데이터 분포에 적합한 Parameter-Efficient Fine-tuning (이하, PEFT)를 생성하는 네트워크를 학습시키는 방법이다. 사용자 맞춤형 학습 모델 파라미터 효율적 연합학습 방법에서, 먼저, 클라이언트 디바이스(200)가 클라이언트 데이터(학습데이터) 분포와 레이어 정보를 포함하는 입력시퀀스를 추출한다(S10). 이후, 클라이언트 디바이스(200)가 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 하이퍼네트워크에 입력한다(S20). 이후, 하이퍼네트워크를 통해 상기 학습데이터 분포와 입력시퀀스로부터 각각 특징벡터를 추출하고 결합벡터를 생성한다(S30). 이후, 상기 결합된 특징벡터(결합벡터)를 미리 마련된 인수분해된 하이퍼네트워크(Factorized HyperNetworks)에 입력하며, 인수분해된 하이퍼네트워크는 입력된 결합벡터를 조건으로 하여 어댑터 파라미터를 생성하도록 학습될 수 있다(S40). 인수분해된 하이퍼네트워크는 연합 학습을 포함하여 다양한 학습 시나리오에 적용할 수 있는 특정 유형의 하이퍼네트워크 아키텍처로서, 계산 복잡성과 하이퍼네트워크의 매개변수 수를 줄이기 위해 행렬 분해를 활용하는 특정한 변형을 의미한다. 인수분해된 하이퍼네트워크로부터 출력된 어댑터 파라미터를 기초로 클라이언트 디바이스(200)에서 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성한다(S50). 즉, 연합학습에 참여하는 각 클라이언트 디바이스(200)가 가지고 있는 클라이언트 데이터 분포에 적합한 PEFT를 생성하는 로컬 모델을 학습시킨다.

[0041] 이하, 각 과정을 더 상세히 설명한다.

[0042] 먼저, 도 4에 제시된 바와 같이, 먼저, 클라이언트 디바이스(200)가 클라이언트 데이터(학습데이터) 분포와 레이어 정보를 포함하는 입력시퀀스를 추출한다(S10). 이후, 클라이언트 디바이스(200)가 클라이언트 데이터 분포와 레이어 정보를 포함하는 입력시퀀스를 하이퍼네트워크에 입력한다(S20).

[0043] 생성할 PEFT 모듈의 구조를 정의하는 것으로 시작한다. 다양한 모듈이 많이 개시되었지만 본 실시예에서는 시각 및 이미지, 오디오와 같은 도메인 전반에 걸친 다재다능함과 주어진 작업을 수행하는 데 입증된 효능을 고려하여 어댑터에 중점을 둔다. 어댑터는 PLM의 모든 블록 내에서 셀프어텐션(self-attention) 레이어와 피드포워드 레이어 사이에 인터리브되는 하향 및 상향 투영 기능으로 구성된다.

[0044] 적응 프로세스는 다음과 같이 공식화할 수 있다.

[0045] $$\mathcal{A}^l(x) = \mathbf{U}^l \mathrm{GeLU}(\mathbf{D}^l x) + x \quad (2)$$

[0046] 여기서 $D^l \in R^{rXd}$ 및 $U^l \in R^{dXr}$은 각각 PLM의 l번째 계층에서 하향 및 상향 투영에 대한 가중치이고, d는 PLM의

숨겨진 차원이고 r은 병목 현상 차원이다.

[0047] 다음으로, 이후, 각 클라이언트 디바이스는 하이퍼네트워크를 통해 상기 학습데이터 분포와 입력시퀀스로부터 각각 특징벡터를 추출하고 결합벡터를 생성한다(S30). 클라이언트의 특성을 나타내기 위해 레이블 임베딩과 컨텍스트 임베딩의 두 가지 유형의 정보를 고려한다. 레이블 임베딩은 각 클라이언트의 클래스 분포에 대한 명시적인 정보를 전달하는 역할을 한다. 미니 배치는 일반적으로 균일 분포에 의해 샘플링되기 때문에 미니 배치의 레이블 분포는 클라이언트의 데이터 분포를 충분히 나타낼 수 있다. 따라서 미니 배치의 레이블 분포에서 레이블 임베딩을 구성한다. 클라이언트 i의 미니 배치를 B ⊂ Di 라고 하면 다음과 같이 레이블 임베딩을 도출할 수 있다.

[0048] $$L(\mathcal{B}) = \mathbf{W}_L \text{avg}([y_1; ...; y_{|B|}]) + b_L \quad (3)$$

[0049] 여기서 yi는 인스턴스 xi, [;]는 연결 함수를 나타내고, avg(·)는 미니 배치 내의 평균 풀링을 나타내고, $W_L \in R^{CXt}$ 및 $b_L \in R^t$는 클래스 C 수에 대한 선형 변환 가중치 및 편향이고, t는 입력 임베딩의 차원이다.

[0050] 테스트 데이터의 레이블에 액세스할 수 없기 때문에 우세한 클래스에 편향되지 않는 어댑터를 생성하기 위해 추론 단계에 대해 균일한 분포를 선택한다는 점에 유의해야 한다.

[0051] 컨텍스트 임베딩 데이터의 컨텍스트 정보를 고려하면 보다 포괄적인 관점(예: 언어, 텍스트 스타일)을 취함으로써 각 클라이언트에 대한 이해를 높일 수 있다. 구체적으로, 계층별 어댑터를 생성하기 위해 모든 계층에서 컨텍스트 정보를 추출한다. 문장 임베딩에서 영감을 받아 컨텍스트 임베딩은 ℓ_2 정규화로 길이에 대한 단어 벡터를 평균화하여 추출된다. PLM의 l번째 레이어에서 샘플 xj의 결과 벡터를 fl(xj)라고 하면 l번째 레이어의 컨텍스트 임베딩은 다음과 같이 파생된다.

[0052] $$F^l(\mathcal{B}) = \mathbf{W}_F \max([f^l(x_1); ...; f^l(x_{|\mathcal{B}|})]) + b_F \quad (4)$$

[0053] 여기서 max(·)는 배치 전체의 최대 풀링을 나타내고 $W_F \in R^{dXt}$ 및 $b_F \in R^t$는 각각 선형 변환 가중치 및 편향이다.

[0054] 포괄적인 클라이언트 임베딩 $\mathcal{I}_B^l$ 는 두 가지 유형의 임베딩을 요약하여 구성된다. 또한 각 레이어의 클라이언트 임베딩에 레이어 인덱스 임베딩을 추가하여 생성기가 더 다양한 레이어별 정보를 인코딩하도록 권장한다.

[0055] 이후, 하이퍼네트워크에서 상기 결합된 특징벡터는 미리 마련된 인수분해된 하이퍼네트워크(Factorized HyperNetworks)에 입력된다. 인수분해된 하이퍼네트워크는 입력된 결합벡터를 조건으로 하여 어댑터 파라미터를 생성하도록 학습된다(S40). 인수분해된 하이퍼네트워크를 통해 클라이언트 임베딩을 기반으로 각 이질적 클라이언트에 맞게 어댑터를 조정한다. 주어진 입력 임베딩을 기반으로 매개변수를 생성하는 인수분해된 하이퍼네트워크를 통해 클라이언트 임베딩 I1B를 입력으로 받아 어댑터 매개변수를 생성할 수 있다.

[0056] 공식적으로 어댑터의 매개변수(즉, U^l, D^l)는 하이퍼네트워크의 기능에 따라 생성된다.

[0057] $$(\mathbf{U}_{\mathcal{B}}^l, \mathbf{D}_{\mathcal{B}}^l) := h(\mathcal{I}_{\mathcal{B}}) = (\mathbf{W}_U, \mathbf{W}_D)\mathcal{I}_{\mathcal{B}}^l \quad (5)$$

[0058] 여기서 $\mathcal{I}$ 는 차원이 t인 입력 임베딩이고, $W_D^l \in R^{(rXd)Xt}$, $W_U^l \in R^{(dXr)Xt}$는 하이퍼네트워크의 가중치이다. 하이퍼네트워크는 입력 임베딩에 인코딩된 레이어별 정보를 사용하여 서로 다른 레이어 간에 공유될 수 있다.

[0059] 전술된 하이퍼네트워크에서 맞춤형 어댑터를 생성할 수 있지만 하이퍼네트워크는 일반적으로 상대적으로 많은 수의 매개변수로 구성된다. 따라서 하이퍼네트워크를 두 개의 더 작은 가중치로 분해한다. 또한 분해된 구성요소의 결과 행렬은 ℓ_2 정규화되어 생성된 매개변수가 클라이언트의 데이터 분포에서 로컬 다수 클래스로 편향되지 않는다. 공식적으로 식(5)는 다음 식(6)과 같이 두 개의 인수분해된 구성요소로 재구성된다.

[0060] $$\mathbf{U}_B^l = \mathbf{W}_U \mathcal{I}_B = \sigma(\mathbf{F}_U \mathbf{S}_U)\mathcal{I}_B \quad (6)$$

[0061] 여기서 $F_U \in R^{dXs}$ 및 $S_U \in R^{sX(rXt)}$는 잠재 요인이 있는 WU에서 인수화된 구성 요소를 나타내고, σ(·)는 Frobenius 정규화를 나타낸다.

[0062] 팩터화 분해(factorization)의 경우 잠재 요인 s는 결과 어댑터의 복잡성과 표현성을 결정하는 데 중요한 역할을 한다. 잠재 요인의 더 큰 차원을 허용하기 위해 두 개의 투영 가중치는 연결된 자동 인코더, 즉 $D_B^l = U_B^{l\top}$인 것처럼 유사하게 연결된다. 이 전략을 사용하면 작업 정확도를 손상시키지 않고 메모리 요구 사항을 절반으로 줄일 수 있다.

[0063] 하이퍼네트워크로부터 출력된 파라미터를 기초로 커스토마이징된 파라미터 효율 파인 튜닝 모델을 생성한다(S50). 본 실시예는 연합학습에 참여하는 각 클라이언트 디바이스(200)가 가지고 있는 클라이언트 데이터 분포에 적합한 Parameter-Efficient Fine-tuning (이하, PEFT)를 생성하는 네트워크를 학습시킨다.

[0064] 각 학습데이터에 대한 훈련 단계가 완료되면 각각의 훈련된 모델은 글로벌 모델을 업데이트하기 위해 중앙 집중식 서버(100)로 다시 전송된다(식(1)). 훈련 모델이 하이퍼네트워크라는 점을 고려하여 각 클라이언트는 글로벌 하이퍼네트워크를 업데이트하기 위해 하이퍼네트워크와 레이어 인덱스 임베딩과 관련된 매개변수를 서버(100)로 전송한다.

[0065] 도 5는 본 발명의 실시예의 비항등분포 연합학습 환경에서의 정량적 실험결과를 나타낸다.

[0066] 본 발명을 실제 연합학습에 적용한 실험결과는 도 5에 제시된 표1과 같다. 실험 환경은 100명의 클라이언트를 가정하고, 텍스트 분류 작업을 연합학습한 실험 결과이다. 특히, 클라이언트의 비항등 데이터 분포에서의 강건성을 확인하기 위하여, 각 클라이언트들은 서로 매우 다른 데이터 분포를 가지고 있다. 다름의 정도는 표의 β이며, 낮을수록 높은 비항등분포 수준을 의미한다.

[0067] 본 발명의 실시예에 의하면, 각 클라이언트의 데이터 분포를 통합하여 페더레이션 시나리오(연합학습)에서 PEFT의 클라이언트 드리프트 문제를 완화하며, 각 클라이언트의 데이터 분포를 통합하여 클라이언트 맞춤형 어댑터를 생성하는 하이퍼네트워크 기반 FL 프레임워크를 제공한다.

[0068] 본 발명의 실시예에 의하면, 다양한 분산 시나리오에서 최첨단 결과를 달성한다는 것을 보여준다.

[0069] 본 발명의 실시예에 의하면, FL 시나리오에서 기존 매개변수를 최소한으로 변경하여 PLM을 교육한다.

[0070] 본 발명의 실시예에 의하면, 개인 정보 유출과 관련하여 사전 학습 과정에서 개인 정보 문제를 해결하지는 않았지만 FL 프레임워크는 미세 조정 단계에서 데이터 개인 정보 문제를 완화할 수 있다. 또한 환경에 미치는 영향과 관련하여 전체 미세 조정의 필요성을 제거할 수 있으며 메모리 또는 분포된 서버 측면에서 비용을 크게 줄일 수 있다.

[0071] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0072] 본 발명의 범위는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

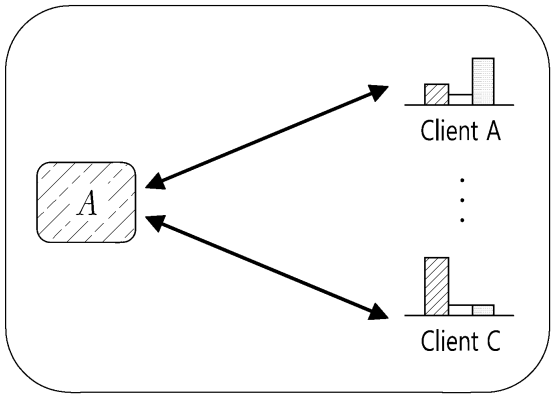
[0074] 100 : 중앙 집중식 서버

200 : 클라이언트 디바이스

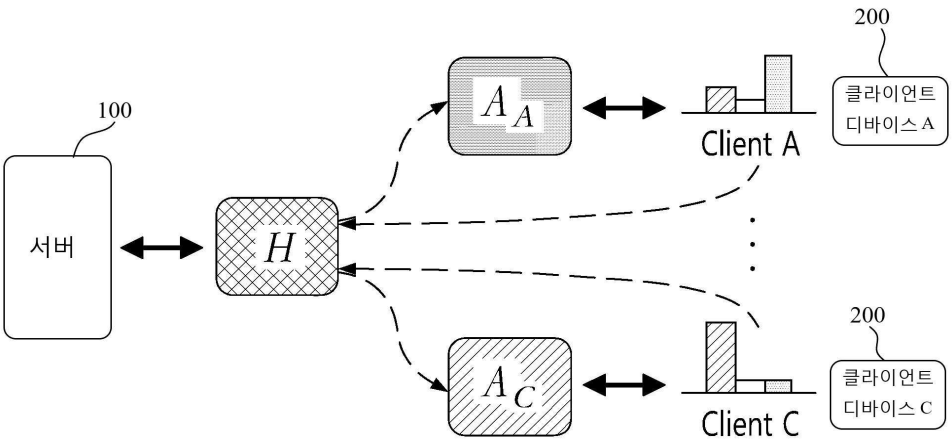
201, 202 : 특징벡터

도면

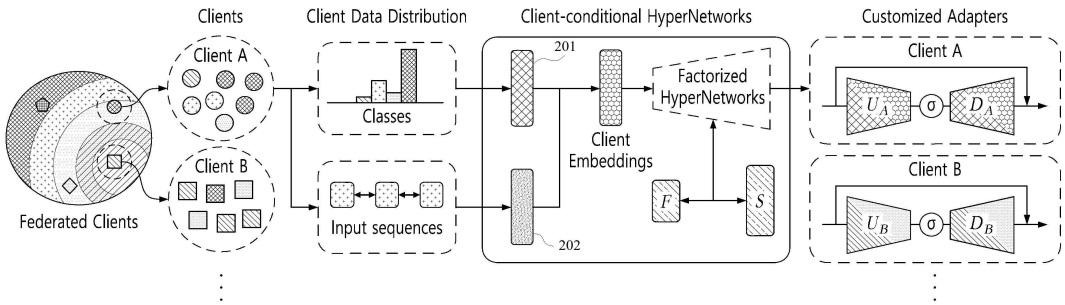
도면1



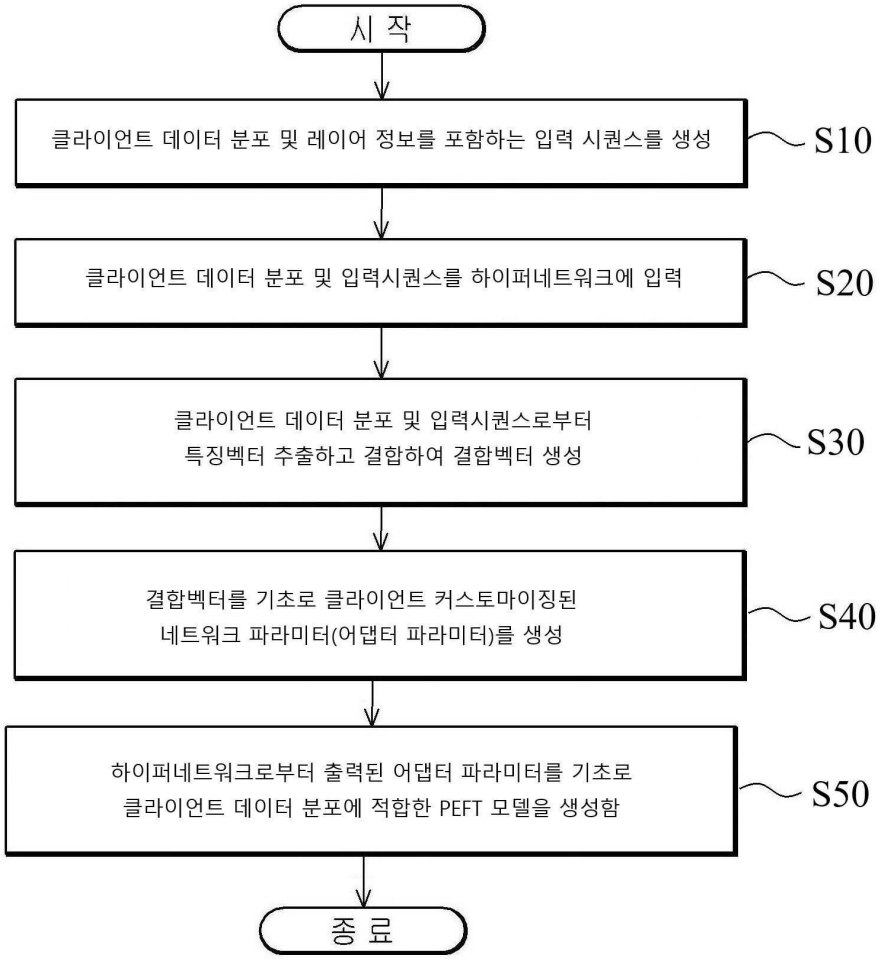
도면2



도면3



도면4



도면5

Methods	Params (%)	Non-Fed	Federated scenario		
			$\beta = 5.0$	$\beta = 2.0$	$\beta = 0.5$
Full Fine-tuning	100%	87.6	84.5	83.7	80.7
Adapter (Houlsby et al., 2019)	0.225%	87.5	78.6	75.0	74.3
LoRA (Hu et al., 2022)	0.055%	**87.8**	80.4	78.4	74.6
Compacter (Karimi Mahabadi et al., 2021)	0.021%	87.3	75.9	73.4	71.0
Prompt-tuning (Li and Liang, 2021)	0.017%	85.6	61.2	60.6	58.0
BitFit (Zaken et al., 2022)	0.038%	87.3	78.4	76.8	72.1
AdaMix (Yaqing Wang and Gao, 2022)	0.277%	87.6	79.6	79.1	76.6
C2A (ours.)	0.049%	87.4	**82.8**	**82.2**	**80.2**