



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0144834
(43) 공개일자 2023년10월17일

(51) 국제특허분류(Int. Cl.)
G06N 20/20 (2019.01) G06N 3/08 (2023.01)
(52) CPC특허분류
G06N 20/20 (2021.08)
G06N 3/08 (2023.01)
(21) 출원번호 10-2022-0044004
(22) 출원일자 2022년04월08일
심사청구일자 2022년04월08일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대 학교)
(72) 발명자
최동완
인천광역시 미추홀구 용정공원로 33, 108동 1303호(용현동, 인천 SK Sky VIEW)
김성웅
서울특별시 강남구 삼성로 212, 22동 1209호(대치동, 은마아파트)
(74) 대리인
양성보

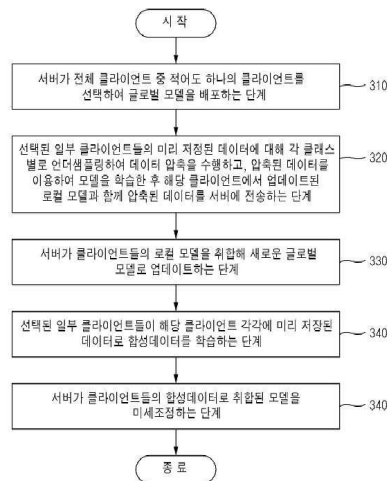
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법 및 시스템

(57) 요약

데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법 및 시스템이 제시된다. 본 발명에서 제안하는 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법은 서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 단계, 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 압축된 데이터를 서버에 전송하는 단계, 상기 서버가 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계, 상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계 및 상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 단계를 포함한다.

대표도 - 도3



이 발명을 지원한 국가연구개발사업

과제고유번호 1711160586
 과제번호 2022-0-00448-001
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 방송통신산업기술개발사업
 연구과제명 [Ezbaro] 인간처럼 회상이 가능한 인공 신경망 지속학습 플랫폼 개발(1차년도)
 기 여 율 50/100
 과제수행기관명 인하대학교 산학협력단
 연구기간 2022.04.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711135511
 과제번호 2021R1F1A1060160
 부처명 과학기술정보통신부
 과제관리(전문)기관명 한국연구재단
 연구사업명 기본연구
 연구과제명 [Ezbaro] 질의 가능한 인공지능: 상호작용형 XAI 서비스를 실현하기 위한 질의 기반

딥러닝 모델 해석 기술

기 여 율 40/100
 과제수행기관명 인하대학교
 연구기간 2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711152759
 과제번호 2020-0-01389-003
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 인공지능융합연구센터지원사업(국고)
 연구과제명 [Ezbaro][정부] 인공지능융합연구센터지원(3차년도)
 기 여 율 10/100
 과제수행기관명 인하대학교 산학협력단
 연구기간 2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 단계;

상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 단계;

상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계;

상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계; 및

상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 단계

를 포함하는 연합학습 방법.

청구항 2

제1항에 있어서,

상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 단계는,

상기 클라이언트들이 각 클래스 별로 언더샘플링된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송하고,

상기 서버는 클라이언트들의 클래스 개수에 비례한 일정한 양의 데이터를 수신하여 클라이언트들이 불균형한 클래스 데이터를 가지고 있는 경우라도 균형된 서버 데이터 집합을 획득하도록 하여 개인 정보를 보호하는 동시에 글로벌 모델의 안정된 성능을 유지하기 위해 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하고, 상기 클라이언트들이 각 클래스 별로 언더샘플링된 상기 압축된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송하는

연합학습 방법.

청구항 3

제1항에 있어서,

상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계는,

상기 서버는 라운드마다 변화하는 클래스 분포를 제한하여 클라이언트의 클래스 개수만큼 압축된 데이터를 수신 받게 되고, 해당 라운드의 전체 클래스 분포를 리밸런싱한 상기 압축된 데이터의 세트를 이용하여 새로운 글로벌 모델로 업데이트하는

연합학습 방법.

청구항 4

제1항에 있어서,

상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계는,

상기 합성데이터의 학습 속도를 향상하기 위해 추가 모델 학습 과정을 제거하고, 학습률을 조정하여 학습률에

따른 학습의 불안정성을 제거하기 위해 경사도 클리핑(gradient clipping) 기법을 최적화에 추가하는
연합학습 방법.

청구항 5

서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 통신부;
상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수
행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께
상기 압축된 데이터를 서버에 전송하는 연합 학습부;
상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하면, 상기 선택된 일부
클라이언트들 각각에 미리 저장된 데이터로 합성데이터를 학습하는 합성데이터 학습부; 및
상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 모델 업데이트부
를 포함하는 연합학습 시스템.

청구항 6

제5항에 있어서,
상기 연합 학습부는,
상기 클라이언트들이 각 클래스 별로 언더샘플링된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서
버에 전송하고,
상기 서버는 클라이언트들의 클래스 개수에 비례한 일정한 양의 데이터를 수신하여 클라이언트들이 불균형한 클
래스 데이터를 가지고 있는 경우라도 균형된 서브 데이터 집합을 획득하도록 하여 개인 정보를 보호하는 동시에
글로벌 모델의 안정된 성능을 유지하기 위해 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하고,
상기 클라이언트들이 각 클래스 별로 언더샘플링된 상기 압축된 데이터를 해당 클라이언트에서 업데이트 된 모
델과 함께 서버에 전송하는
연합학습 시스템.

청구항 7

제5항에 있어서,
상기 모델 업데이트부는,
상기 서버는 라운드마다 변화하는 클래스 분포를 제한하여 클라이언트의 클래스 개수만큼 압축된 데이터를 수신
받게 되고, 해당 라운드의 전체 클래스 분포를 리밸런싱한 상기 압축된 데이터의 세트를 이용하여 새로운 글로벌
모델로 업데이트하는
연합학습 시스템.

청구항 8

제5항에 있어서,
상기 합성데이터 학습부는,
상기 합성데이터의 학습 속도를 향상하기 위해 추가 모델 학습 과정을 제거하고, 학습률을 조정하여 학습률에
따른 학습의 불안정성을 제거하기 위해 경사도 클리핑(gradient clipping) 기법을 최적화에 추가하는
연합학습 시스템.

발명의 설명

기술 분야

본 발명은 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법 및 시스템에 관한 것이다.

[0001]

배경 기술

- [0002] 최근, 딥러닝은 대규모 데이터 집합을 활용해 여러 분야에서 괄목할만한 성장을 보이고 있지만, 수백만장이 넘어가는 데이터를 학습하는 일은 막대한 보관 비용과 계산 비용을 필요로 한다. 또한, 기존의 중앙 학습 방식은 모든 데이터를 하나의 데이터 서버에 저장해 학습한 반면, 실제 환경에선 데이터들은 여러 IoT 장치, 모바일 장치, 자율주행차 등에서 생성되거나 수집되어, 분산된 형태로 존재한다. 따라서, 분산 데이터를 서버로 모으기에는 많은 수집 비용과 개인 정보 보호 문제를 야기하며, 데이터 서버를 사용하지 않고 딥러닝 어플리케이션 사용자가 개인 장치에 존재하는 데이터로만 모델을 학습하기에는 데이터의 양이 충분하지 않다. 이러한 분산 데이터 환경과 중앙 데이터 서버 학습 방식의 간극을 해결하기 위해, 분산 데이터를 활용하면서도 클라이언트의 개인 데이터를 전송하지 않고 학습하는 연합학습이 등장하였다.
- [0003] 연합학습이란 개인 정보 보호를 보장하는 딥러닝 모델 학습 방식으로, 자신의 데이터로 모델을 학습하는 클라이언트들과 학습된 모델을 취합하는 하나의 파라미터 서버로 이루어져 있다. 클라이언트들은 개인 정보를 보호받으면서도 좋은 성능의 모델을 제공받고, 파라미터 서버를 운영하는 회사는 데이터의 수집 비용, 모델 학습 컴퓨팅 비용을 절감 가능한 장점이 존재한다. 연합학습은 서비스 사용자와 제공자 모두 만족하는 Win-Win 전략이라 볼 수 있다. 그러나, 데이터 서버에서 전체 데이터를 학습, 검증, 테스트 데이터로 나누어 모델을 완전 학습 후, 평가하던 전통적인 방식과 달리, 연합학습 시나리오에선 모델의 성능을 검증할 데이터가 파라미터 서버에 존재하지 않는다. 따라서, 연합학습에선 학습 중인 모델의 성능이 어느 정도인지 가늠할 수 없어, 하이퍼파라미터를 조정하기 어렵고, 모델의 성능이 불안정할 경우 잘못된 모델을 배포하게 되는 등 여러 문제를 일으킬 수 있다.
- [0004] 종래기술에서는 테스트 데이터 집합을 서버에 두고 학습 과정 중에 모델 성능을 확인하였다. 이는 연합학습에서는 실현 불가능한 방법으로, 오직 실제 데이터가 아닌 공공 데이터를 활용한 실험 환경에서만 가능하다. 다른 방법으로는 일정 클라이언트가 학습에 참여하지 않고 검증을 위한 클라이언트가 되거나 전체 클라이언트들의 일부 데이터를 검증 데이터로 사용하는 방법도 있다. 하지만, 이러한 방법은 참여하는 클라이언트의 수가 적거나 클라이언트의 데이터가 검증 데이터까지 나누기에 충분하지 않을 경우 불가능한 실용적이지 않은 방법이다.
- [0005] 모델 성능을 확인할 수 없는 상황에서, 모델의 갑작스러운 성능 저하는 심각한 문제를 초래할 수 있다. 예를 들어, 다음 단어 예측이나 이모티콘 추천처럼 키보드 어플리케이션 등에서도 성능 저하로 클라이언트에게 불편을 야기할 수도 있지만, 특별히 뇌 종양 식별이나, 코로나 환자들의 임상 결과 예측과 같은 의료 영역에서는 모델의 성능이 갑자기 하락하는 경우 더욱 심각한 악영향을 끼칠 수가 있다. 따라서, 연합학습 모델의 안정적인 성능은 필수적으로 보장되어야 한다.
- [0006] 연합학습의 주요 문제점은 데이터 이질성 문제로, 각 클라이언트의 데이터는 클라이언트(예를 들어, 모바일 사용자, 병원 등)의 패턴에 따라 각기 다른 데이터 분포를 형성한다. 이질적인 분포에서 학습된 모델들을 취합한 모델은 성능이 저하되고 느리게 수렴하는 문제점이 발생한다.
- [0008] 도 1은 종래기술에 따른 연합학습에서의 데이터 이질성에 따른 성능 그래프를 나타내는 도면이다.
- [0009] 도 1을 참조하면, CIFAR-10에서 50명이 클라이언트가 참가하는 연합학습에서의 데이터 이질성에 따른 세 개(α 0.05, 0.1, 1)의 성능 그래프를 나타내었다.
- [0010] 도 1에서, 파라미터 α 는 클라이언트들의 데이터 분포를 조정하는 파라미터로, α 가 작으면 클라이언트들은 더 편향된 데이터 분포를 가진다. 도 1의 세 그래프를 통해 각 클라이언트에서 수집되는 로컬 데이터 분포 간의 이질성(non-i.i.d)이 강할수록(다시 말해, α 가 작을수록), 모델 성능의 변동 폭이 심해짐을 확인할 수 있다. 따라서, non-i.i.d가 모델의 최종 성능이나 느린 수렴 속도뿐만 아니라, 안정성도 해침을 알 수 있다. 이는 앞서 언급한 바와 같이, 모델의 성능을 검증하지 못하고 학습되는 중간에도 모델을 배포할 필요가 있는 연합학습에서는, 심각한 단점이라고 볼 수 있다. 불안정성 문제가 발생하는 보다 구체적인 원인은 서버에서 모델을 취합할 때 각 라운드마다 전체가 아닌 일부 클라이언트만 참여하기 때문이며, 이렇게 참여하는 클라이언트들의 이질적인 데이터 분포로 인해 해당 라운드에서 클래스 불균형 문제가 발생하기 때문이다.
- [0011] 종래기술의 경우 데이터 이질성으로 인한 성능 하락 문제의 원인을 편향된 클라이언트 모델 업데이트로 파악한다. 이를 'Client-drift', 전체 업데이트와 엇나가는 클라이언트의 업데이트로 인해 발생하는 문제로 보고 클라

이언트의 업데이트를 제한하는 방법들이 주로 연구되었다. 그러나, 해당 방법들은 이전 라운드의 모델을 최대한 업데이트 하는 것을 목적으로 하기 때문에 학습의 효율이 떨어지거나, 이전 시점의 업데이트조차 편향되어 있는 경우 모델이 오히려 계속 잘못된 방향으로 학습될 수 있다. 이는 클라이언트의 업데이트 제한이 이론적으로는 이질성 문제를 완화하지만, 딥러닝 문제에선 실제로 잘 동작하지 않기 때문이다.

- [0012] 또 다른 데이터 이질성 해결 방안으로 극단적으로 이질적인 데이터 상황에서 학습이 되지 않는 것을 방지하기 위해, 모든 클라이언트가 공유하는 공공 데이터 집합을 사용하는 해결 방안이 있다. 하지만, 실제 상황과는 맞지 않은 비실용적인 방법이다.

선행기술문헌

비특허문헌

- [0013] (비특허문헌 0001) [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273-1282

발명의 내용

해결하려는 과제

- [0014] 본 발명이 이루고자 하는 기술적 과제는 각 라운드마다 학습에 참여하는 클라이언트의 데이터 분포가 달라지게 되고 이로 인해 그 전 라운드의 성능을 망각하게 되는 현상 때문에 연합학습의 불안정성이 발생하는 문제를 해결하기 위해 각 클라이언트에서 수집된 데이터를 개인정보를 노출하지 않을 정도의 형태로 요약 및 압축하는 데이터 집합 압축(dataset condensation) 방법 및 시스템을 제공하는데 있다.

과제의 해결 수단

- [0015] 일 측면에 있어서, 본 발명에서 제안하는 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법은 서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 단계, 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 단계, 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계, 상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계 및 상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 단계를 포함한다.
- [0016] 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 단계는 상기 클라이언트들이 각 클래스 별로 언더샘플링된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송하고, 상기 서버는 클라이언트들의 클래스 개수에 비례한 일정한 양의 데이터를 수신하여 클라이언트들이 불균형한 클래스 데이터를 가지고 있는 경우라도 균형된 서버 데이터 집합을 획득하도록 하여 개인 정보를 보호하는 동시에 글로벌 모델의 안정된 성능을 유지하기 위해 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하고, 상기 클라이언트들이 각 클래스 별로 언더샘플링된 상기 압축된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송한다.
- [0017] 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계는 상기 서버는 라운드마다 변화하는 클래스 분포를 제한하여 클라이언트의 클래스 개수만큼 압축된 데이터를 수신 받게 되고, 해당 라운드의 전체 클래스 분포를 리밸런싱한 상기 압축된 데이터의 세트를 이용하여 새로운 글로벌 모델로 업데이트한다.
- [0018] 상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계는 상기 합성데이터의 학습 속도를 향상하기 위해 추가 모델 학습 과정을 제거하고, 학습률을 조정하여 학습률에 따른 학습의 불안정성을 제거하기 위해 경사도 클리핑(gradient clipping) 기법을 최적화에 추가한다.
- [0019] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 시스템

은 서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 통신부, 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 연합 학습부, 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하면, 상기 선택된 일부 클라이언트들 각각에 미리 저장된 데이터로 합성데이터를 학습하는 합성데이터 학습부 및 상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 모델 업데이트부를 포함한다.

발명의 효과

[0020] 본 발명의 실시예들에 따르면 각 클라이언트에서 수집된 데이터를 개인정보를 노출하지 않을 정도의 형태로 요약 및 압축하는 데이터 집합 압축(dataset condensation) 방법 및 시스템을 통해 연합학습의 불안정성을 개선할 수 있다. 이를 통해 데이터를 개인정보를 구분할 수 없을 정도의 극히 적은 양의 데이터로 압축함으로써, 정보의 손실을 최소화하면서도 프라이버시를 보장할 수 있다. 또한, 본 발명의 실시예들에 따르면 다양한 실험 환경에서 종래기술에 따른 알고리즘에 비해 더 빠른 수렴속도와 안정적인 성능을 보장함을 확인하였다. 특히, 데이터 이질성이 강할수록 다른 알고리즘들에 비해 더 많은 격차를 보임을 확인하였다.

도면의 간단한 설명

[0021] 도 1은 종래기술에 따른 연합학습에서의 데이터 이질성에 따른 성능 그래프를 나타내는 도면이다.
 도 2는 종래기술에 따른 클래스 데이터 분포 및 FedAvg, FedDC의 클래스별 성능 비교를 위한 도면이다.
 도 3은 본 발명의 일 실시예에 따른 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법을 설명하기 위한 흐름도이다.
 도 4는 본 발명의 일 실시예에 따른 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 장치의 구성을 나타내는 도면이다.
 도 5는 본 발명의 일 실시예에 따른 FedDC 학습 알고리즘을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 연합학습은 하나의 글로벌 모델을 만들기 위해 여러 클라이언트들이 모델을 데이터 공유 없이 협력 학습하는 기법으로, 최근 개인 정보 보호가 중요해지고 데이터 서버 유지 비용이 증가함에 따라, 딥러닝 커뮤니티에서 대두되고 있는 학습 방법이다. 그러나, 기존의 중앙 학습 방식은 학습이 완료된 모델을 검증 데이터 집합으로 평가 후 클라이언트에게 배포하지만, 연합학습에서는 클라이언트들의 모델을 취합하는 서버에 모델을 검증할 데이터 집합이 없다고 가정하므로 학습 중인 모델의 성능을 평가할 수가 없다. 또한, 서버에서 모델이 취합되어 학습되는 중간에도 각 클라이언트에게 배포되어 사용되는 경우가 많아 연합학습 된 모델의 일시적 성능 하락이 발생할 경우 치명적인 문제로 이어질 수 있다. 연합학습의 학습을 저해하는 주요 원인은 데이터 이질성으로, 각 클라이언트에서 수집되는 로컬 데이터 분포 간의 이질성(non-i.i.d)으로 인해 느린 수렴속도와 성능이 하락하는 문제가 발생한다. 특히, 본 발명에서는 non-i.i.d의 정도에 비례하여 모델의 성능이 급격히 변화하는 문제 또한 발생함을 확인하였다. 이는 클라이언트의 데이터 분포를 미리 확인할 수 없고 모델의 중간 성능을 검증할 수 없는 연합학습에서는 치명적인 단점이라고 할 수 있다.

[0023] 따라서, 본 발명에서는 연합학습의 불안정성 문제의 원인이 각 라운드마다 학습에 참여하는 클라이언트의 데이터 분포가 달라지게 되고 이로 인해 그 전 라운드의 성능을 망각하게 되는 현상 때문임을 확인하고, 이를 해결하기 위해 각 클라이언트에서 수집된 데이터를 개인정보를 노출하지 않을 정도의 형태로 요약 및 압축하는 데이터 집합 압축(dataset condensation) 방법을 제안한다. 이후, 압축된 데이터를 로컬 모델과 함께 서버에 전송하고 서버에서 이를 활용하여 글로벌 모델이 안정되고 일관된 성능을 유지할 수 있도록 하는 방안을 제안한다. 제안하는 방법을 적용한 결과, 연합학습의 글로벌 모델은 항상 안정적인 성능을 유지할 뿐만 아니라 빠른 수렴 속도 또한 보임을 확인하였다.

[0024] 본 발명에서는 매번 달라지는 불균형한 데이터 분포를 맞춰 주기 위해, 각 클라이언트의 클래스 데이터를 언더샘플링하여 서버에 전송한다. 서버는 균형 잡힌 데이터 집합을 활용해 편향된 서버 모델을 미세조정 함으로써, 일반화된 모델을 얻어 안정적인 성능을 보장해준다. 그러나, 실제 데이터는 프라이버시를 보장하지 못해 서버에 전송할 수 없고, 실제 데이터를 보내더라도 적은 데이터를 선택하는 과정에서 랜덤성으로 인해 정보의 손실이

발생한다. 이에 본 발명에서는 학습을 통해 원본 데이터를 개인정보를 구분할 수 없을 정도의 극히 적은 양의 데이터로 압축함으로써, 정보의 손실을 최소화하면서도 프라이버시를 보장하는 방법을 제안한다.

[0025] 실험 결과, 본 발명에서 제안하는 방법은 다양한 실험 환경에서 다른 알고리즘에 비해 더 빠른 수렴속도와 안정적인 성능을 보장함을 확인하였다. 특히, 데이터 이질성이 강할수록 다른 알고리즘들에 비해 더 많은 격차를 보임을 확인하였다. 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0027] 도 2는 종래기술에 따른 클래스 데이터 분포 및 FedAvg, FedDC의 클래스별 성능 비교를 위한 도면이다.

[0028] 도 2를 참조하면, 120라운드(도 2(a))와 121라운드(도 2(b))의 라운드의 클래스 데이터 분포 및 FedAvg, FedDC의 클래스별 성능 비교를 위한 그래프를 도시한다.

[0029] 연합학습[1]은 분산된 모바일 장치 환경에서 학습할 수 있도록 제안한 방식이며, 기존 SGD를 응용한 FedAvg 알고리즘을 고안하였다. 연합학습 알고리즘의 성능을 개선하기 위한 연구는 크게 두 가지 분류로 나눌 수 있다. 로컬 제한 연구와 모델 취합 알고리즘 개선 연구이다.

[0030] 로컬 제한 연구는 클라이언트의 로컬 업데이트와 전체 업데이트의 차이를 제한하는 항을 추가한다. FedProx는 이전 서버 모델의 파라미터와 클라이언트 모델 파라미터의 차이인 Proximal term을 제한항으로 사용한다. SCAFFOLD는 전체 경사도(gradient) 방향과 로컬 경사도 방향의 차이를 제한항으로 사용한다. 또 다른 종래기술에서는 이전 클라이언트 모델을 음성 항으로, 서버 모델을 양성항으로 뺄으로써 대조학습을 통해 모델을 제한한다.

[0031] 모델 취합 개선 알고리즘에서, FedNova는 일반적인 모델 평균 방식은 객관적 불일치를 유발한다고 분석해, 정규화된 평균화 방식을 제안한다. FedAdam은 기존 모델 취합 알고리즘을 의사 경사도(pseudo gradient)로 해석하여 서버 업데이트에도 최적화 알고리즘을 적용한다. FedMA는 단순히 모델 파라미터의 가중 합산이 아닌 모델의 레이어 별로 뉴런 매칭을 통해 최적화된 합산 방식을 제안한다.

[0032] 수백만 개의 샘플로 구성된 대규모 데이터 집합은 딥러닝 모델을 얻기 위해 표준이 되고 있지만, 이러한 규모의 데이터를 저장하고 학습에 사용하는 일은 막대한 비용과 인프라를 필요로 한다. 따라서, 데이터 집합 크기를 줄이면서도 원본 데이터 집합으로 학습한 모델의 성능을 최대한 유지하는 데이터 압축 기법이 등장하였다. 종래기술에서는 인공 노이즈 데이터로 모델을 학습시킨 다음, 실제 데이터의 학습 손실을 최소화하도록 인공 데이터를 최적화하는 방식은 제안하였다. 또 다른 종래기술에서는 동일한 문제상황에서 실제 데이터와 인공 데이터를 동일한 모델에 통과시켰을 때의 경사도를 일치하도록 학습 방법을 제안했다. 또 다른 종래기술에서는 데이터 증강 기법을 활용해 DC의 성능을 더 끌어올렸다.

[0033] 본 발명에서는 종래기술의 동일한 문제상황에서 실제 데이터와 인공 데이터를 동일한 모델에 통과시켰을 때의 경사도를 일치하도록 학습 방법을 연합학습에 활용하여 클라이언트의 로우 데이터를 압축해 정보의 손실을 최소화하고 사람이 알아보지 못하는 합성 데이터를 압축 후 통신하여 개인 정보 보호를 보장하는 방법을 제안한다.

[0034] 모든 클래스의 데이터 수가 균등한 실험 데이터 환경과 달리, 실제 데이터 분포는 클래스 데이터 수가 굉장히 불균형한 롱테일 분포인 경우가 대다수이다. 이러한 롱테일 분포 상황에서 클래스 불균형 문제는 데이터가 적은 마이너리티 클래스의 성능이 굉장히 떨어지고 데이터 수가 많은 메이저리티 클래스의 극도로 높아져, 전체 성능이 특정 클래스에게 지배되는 문제가 발생한다. 이런 클래스 불균형 문제를 해결하기 위해, 데이터 분포를 리밸런싱해 학습하는 오버샘플링이나 언더샘플링 방법이 존재한다. 또한, 기존 지도학습의 크로스 엔트로피 손실함수 대신 각 클래스의 샘플마다 다른 손실 값을 주는 클래스 균형 손실(Class-balanced loss) 방법, 분류기의 불균형을 해결해주는 방법이 존재한다.

[0035] 연합학습에서도 또한 실제 데이터 환경을 고려해 클래스 불균형한 문제를 고려한 연구들이 존재한다. 그러나 본 발명에서는, 전체 데이터 클래스 분포가 불균형한 상황이 아니라, 전체 데이터가 균등한 상황이라도 참여하는 클라이언트들의 랜덤성에 의해 일시적인 데이터 불균형이 발생했을 때의 클래스 불균형이 글로벌 모델의 성능을 저하시키는 문제를 해결한다.

[0036] 앞서 설명된 바와 같이, 연합학습은 데이터 이질적인 상황이 심할수록, 모델의 성능이 불안정해진다는 사실에 대해 분석한다. 다시 도 1을 참조하면, 그래프에서 $\alpha=0.05$ 일 때 성능이 가장 요동치는 원인을 근본적으로 파악하고자 도 2에 도시된 바와 같이 특정 라운드의 실험 결과를 자세히 분석한다. 특별히 FedAvg 알고리즘의 성능

하락이 두드러지게 나타나는 120번째 라운드(정확도 43.9%)와 121번째 라운드(정확도 21.3%)를 참조한다.

[0037] 먼저, 각 라운드에서의 클래스 분포, called 라운드 클래스 분포를 살펴본다. 라운드마다 참여하는 클라이언트들은 서로 달라지기 때문에, 라운드 클래스 분포는 한 라운드 차이로 하더라도 매우 다를 수 있다. 이는 전체 데이터 집합에서는 클래스들이 균등하게 분포하고 있더라도, 클라이언트들마다 다른 분포를 가지고 있는 상황(다시 말해, non i.i.d)으로 인해 해당 라운드에서 일시적으로 발생할 수 있는 일종의 클래스 불균형 문제이다. 즉, 각 클라이언트에서 클래스 별로 데이터 양이 차이가 나기 때문에 발생하게 되는 것이다. 예를 들어, 120번째 라운드는 세 클래스만이 1000개의 데이터가 넘고 나머지 7개 클래스는 400개보다도 더 적으며, 121번째 라운드에서는 특히 첫 번째 클래스 데이터 수가 4000개가 넘지만, 다른 3개의 클래스는 데이터가 50개보다 적다. 이때, FedAvg의 경우 120번째 라운드와 121번째 라운드에서, 해당 라운드의 메이저리티 클래스한테 성능이 잠식됨을 확인할 수 있다. 특히 121 라운드에서 첫 번째 클래스가 다른 클래스에 비해 유독 많은 데이터를 가지고 있어, 120번째 라운드에 그나마 균등했던 성능 분포가 121번째 라운드 첫 번째 클래스에 오퍼버피팅 되면서 전체 성능이 편향됨을 볼 수 있다. 따라서, 기존 FL알고리즘의 경우 이전 라운드에서 학습된 지식을 잊어버리는 치명적 망각현상이 매 라운드마다 발생하게 되어 성능이 요동치는 문제가 발생한다.

[0038] 본 발명에서는 이와 같은 라운드별(round-wise) 치명적 망각 현상을 효과적으로 완화시키기 위한 방법을 제안한다. 기본적인 아이디어는 클라이언트들이 각 클래스 별로 언더샘플링된 적은 양의 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송하는 것이다. 즉, 이 때 서버는 클라이언트의 클래스 개수에 비례한 일정한 양의 데이터를 받게 된다. 이에 클라이언트가 불균형한 클래스 데이터를 가지고 있더라도 상관없이 균형된 서버 데이터 집합을 얻을 수 있다. 따라서, 라운드 클래스 분포에 편향된 서버 모델을 균형 잡힌 서버 데이터 집합을 통해 미세조정하여 모델의 일반화 성능을 회복한다.

[0039] 하지만, 클라이언트들의 데이터는 개인 정보 보호문제로 인하여 그 자체로 서버에 보낼 수가 없으며, 설사 극히 일부를 선택하여 소량의 데이터를 보낸다 하더라도 어떤 데이터가 선택되느냐에 따라서 성능은 매번 달라질 수 있고 선택된 데이터에 모델이 과적합하게 되는 문제가 여전히 발생할 수 있다. 따라서, 본 발명에서는 개인 정보를 보호하면서도 글로벌 모델의 안정된 성능을 유지하기 위해, 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하여 보내는 방법을 제안한다. 이를 통해 매 라운드마다 달라지는 클래스 분포를 효과적으로 통제하여, 안정적인 모델을 얻을 수 있다. 즉, 서버는 클라이언트의 클래스 개수만큼 압축된 데이터를 받게 되고, 이 압축 데이터 세트는 해당 라운드의 전체 클래스 분포를 리밸런싱한 데이터 세트로, 이를 통해 편향되어 있는 기존 모델(FedAvg)을 일반화된 성능을 갖는 모델(FedDC)로 만들 수 있음을 보인다. 120 번째 라운드에선 FedAvg와 FedDC가 비슷한 성능 분포를 보이지만, FedDC는 좀더 소외된 클래스에 대해서도 잘 맞춰 전체적인 성능이 올라감을 확인할 수 있고, 특히 121번째 라운드에서 첫 번째 클래스에 편향되어 있던 성능이 전체 클래스를 모두 공평하게 잘 맞추므로써, FedDC는 이전 라운드 성능51.7%와 비슷한 성능 51.4%를 유지한다.

[0041] 도 3은 본 발명의 일 실시예에 따른 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법을 설명하기 위한 흐름도이다.

[0042] 제안하는 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 방법은 서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포하는 단계(310), 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송하는 단계(320), 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하는 단계(330), 상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습하는 단계(340) 및 상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 단계를 포함한다.

[0043] 단계(310)에서, 서버가 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포한다.

[0044] 단계(320)에서, 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송한다.

[0045] 상기 클라이언트들이 각 클래스 별로 언더샘플링된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송하고, 상기 서버는 클라이언트들의 클래스 개수에 비례한 일정한 양의 데이터를 수신하여 클라이언트들이 불균형한 클래스 데이터를 가지고 있는 경우라도 균형된 서버 데이터 집합을 획득하도록 하여 개인 정보를

보호하는 동시에 글로벌 모델의 안정된 성능을 유지하기 위해 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하고, 상기 클라이언트들이 각 클래스 별로 언더샘플링된 상기 압축된 데이터를 해당 클라이언트에서 업데이트 된 모델과 함께 서버에 전송한다.

[0046] 단계(330)에서, 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트한다.

[0047] 상기 서버는 라운드마다 변화하는 클래스 분포를 제한하여 클라이언트의 클래스 개수만큼 압축된 데이터를 수신 받게 되고, 해당 라운드의 전체 클래스 분포를 리밸런싱한 상기 압축된 데이터의 세트를 이용하여 새로운 글로벌 모델로 업데이트한다.

[0048] 단계(340)에서, 상기 선택된 일부 클라이언트들이 해당 클라이언트 각각에 미리 저장된 데이터로 합성데이터를 학습한다.

[0049] 상기 합성데이터의 학습 속도를 향상하기 위해 추가 모델 학습 과정을 제거하고, 학습률을 조정하여 학습률에 따른 학습의 불안정성을 제거하기 위해 경사도 클리핑(gradient clipping) 기법을 최적화에 추가한다.

[0050] 단계(350)에서, 상기 서버가 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정한다.

[0051] 본 발명에서는 클래스 불균형한 라운드 데이터 분포를 해결하기 위해 데이터 압축 기법을 통한 균형된 미세조정 방법(FedDC)을 제안한다.

[0052] 본 발명의 실시예에 따른 연합학습의 시나리오에서 아래와 같은 수식을 최소화하는 것을 목표로 한다:

$$\min_{\theta} F(\theta) = \sum_{i=1}^M p_i (F_i(\theta) := \mathbb{E}_{x \sim \mathcal{P}_i} [f_i(\theta; x)])$$

$$\text{where } \sum_{i=1}^M p_i = 1 \text{ and } p_i \geq 0 \quad (1)$$

[0053] 딥러닝 모델 파라미터를 $\theta \in \mathbb{R}^d$, 연합학습에 참여하는 전체 클라이언트의 수를 M , i^{th} 클라이언트의 가중치를 p_i 라 정의한다, F_i 는 i^{th} 클라이언트의 로컬 목적 함수를, f_i 는 클라이언트의 손실 함수를 의미한다. 이 때, 클라이언트의 데이터 분포 $\mathcal{P}_i$ 는 클라이언트의 패턴에 따라 매우 다양한 데이터 분포를 가질 수 있다.

[0055] 클라이언트들의 학습 데이터 집합 $\mathcal{D}_i$, 연합학습에서의 전체 데이터집합 $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M\}$ 이라 정의할 때, 연합학습의 가장 기본적인 알고리즘인 FedAvg 은 실제 클라이언트의 데이터 분포 $\mathcal{P}_i$ 에서 위 식(1)의 최적화 문제를 해결하기 위해 아래와 같은 수식을 사용한다:

$$F(\theta) = \sum_{i=1}^M \frac{|\mathcal{D}_i|}{|\mathcal{D}|} F_i(\theta),$$

$$\text{where } F_i(\theta) = \frac{1}{|\mathcal{D}_i|} \sum_{x \in \mathcal{D}_i} \ell_{ce}(\theta; x) \quad (2)$$

[0056] 전체 데이터 집합의 크기는 $|\mathcal{D}| = \sum_{i=1}^M |\mathcal{D}_i|$ 이며, 위 식(1)식에서 각 클라이언트의 가중치는 동일한 값이 아닌 자신의 학습 데이터 수 (i.e., $|\mathcal{D}_i|$)에 비례한 값으로 설정된다. 즉, 클라이언트들의 로컬 파라미터를 가중 합산하여 최종적으로 글로벌 파라미터를 구한다. 클라이언트의 로컬 목적함수 $F_i(\theta)$ 는 경험적 위험 최소화 목적함수(empirical risk minimization objective function)이고, ℓ 는 다중 클래스 분류(multi-class classification)에서 일반적으로 사용하는 크로스 엔트로피 손실(cross entropy loss)이다. 앞서 언급한 클라이언트 로컬 업데이트 제한 알고리즘은 ℓ_{ce} 에 제한항(regularization term) ℓ_{ce} 을 추가한다.

[0058] 연합학습의 글로벌 모델은 총 T 번의 라운드를 반복해 학습되며, 한 라운드는 다음과 같이 세 단계로 진행된다. 첫 번째로, 서버가 전체 클라이언트 중 일부에게 글로벌 모델을 배포한다. 두 번째로, 선택된 일부 클라이언트는 자신의 데이터로 모델을 학습 후 업데이트된 모델을 서버에 전송한다. 마지막으로, 서버는 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트한다. 이 때, FedAvg 알고리즘은 클라이언트의 로컬 모델 업데이트 시, 여러 번의 미니배치 SGD를 수행한다:

[0059] $\theta_{i,k+1}^t = \theta_{i,k}^t - \eta \nabla \ell(\theta_{i,k}^t; x_{i,k}), k = 0, 1, \dots, E - 1 \quad (3)$

[0060] $x_{i,k}$ 는 $\mathcal{D}_i$ 에서 랜덤하게 선택된 데이터, η 는 학습률, $\theta_{i,k+1}^t$ 는 t 라운드에서 i^{th} 클라이언트가 총 E 번의 이터레이션을 반복해 업데이트한 로컬 모델을 의미한다.

[0061] 그러나, 연합학습의 글로벌 모델이 식(1)을 최소화하는 지점으로 수렴하더라도, 중앙 학습 방식으로 학습한 모델과 같은 최적화 지점을 갖지 않는다. 이는, 데이터 이질성 문제로, 중앙학습 방식의 딥러닝 모델은 모든 클라이언트의 학습 데이터를 모은 전체 데이터 셋 $\mathcal{D} = \sum_{i=1}^M \mathcal{D}_i$ 에서 i.i.d로 데이터를 선택(또는, 생성)해 학습한다. 따라서, $\mathcal{D}$ 에서 선택된 미니배치 데이터는 전체 데이터 분포 $\mathcal{P}$ 와 같은 분포를 갖는다, 최근, 이러한 i.i.d 데이터로 학습한(다시 말해, 중앙 학습 방식) 딥러닝 모델은 동일한 i.i.d 테스트셋에서도 사람보다 뛰어난 성능을 보였다. 반면, 연합학습에선 각 클라이언트의 학습 데이터 집합 $\mathcal{D}_i$ 는 전체 데이터 분포 $\mathcal{P}$ 와 다르고, $\mathcal{D}_i$ 에서 선택된 데이터 또한 $\mathcal{P}$ 를 표현할 수 없다. 따라서, non-i.i.d 데이터 집합 $\mathcal{D}_i$ 으로 학습한 클라이언트 모델 $F_i(\theta)$ 과 클라이언트 모델들을 취합한 서버모델 $\sum_{i=1}^M \frac{|\mathcal{D}_i|}{|\mathcal{D}|} F_i(\theta)$ 은 전체 데이터 집합 $\mathcal{D}$ 에서 i.i.d로 학습한 모델과 같은 수렴 지점을 갖지 않는다.

[0062] 앞서 설명된 바와 같이, 연합학습 문제를 해결하기 위해 본 발명에서는 클라이언트의 원본 데이터를 요약 및 압축하는 합성데이터 학습 방법을 제안한다.

[0063] m^{th} 클라이언트의 학습 데이터 집합을 $\mathcal{D}_m = \{(x_n, y_n)\}_{n=1}^{|\mathcal{D}_m|}$ 이라 할 때, 데이터 압축기법의 목표는 기존 학습 데이터 집합 $\mathcal{D}_m$ 보다 작은 합성 데이터 집합 $\mathcal{S}_m = \{(\tilde{x}_n, y_n)\}_{n=1}^{|\mathcal{S}_m|}$ 을 학습하는 것으로, 아래의 식을 만족해야 한다:

[0064] $\mathbb{E}_{x \sim \mathcal{P}_m} [\ell(\theta^{\mathcal{D}_m}; x)] = \mathbb{E}_{\tilde{x} \sim \mathcal{P}_m} [\ell(\theta^{\mathcal{S}_m}; \tilde{x})] \quad (4)$

[0065] 여기서, $\theta^{\mathcal{D}_m} = \min_{\theta} \ell(\theta, x^*, y^*)$, $\theta^{\mathcal{S}_m} = \min_{\theta} \ell(\theta, x^*, y^*)$ 이고, $|\mathcal{S}_m| \ll |\mathcal{D}_m|$ 이다.

[0066] 합성 데이터 집합 $\mathcal{S}_m$ 과 기존 학습 데이터 집합 $\mathcal{D}_m$ 으로 학습한 딥러닝 파라미터를 각각 $\theta^{\mathcal{D}_m}, \theta^{\mathcal{S}_m}$ 라 한다. 이때, 극히 적은 양의 데이터로 학습한 모델 $\theta^{\mathcal{S}_m}$ 이 실제 클라이언트의 데이터 분포 $x^* \sim \mathcal{P}_m$ 에서 $\theta^{\mathcal{S}_m}$ 만큼 좋은 일반화 성능을 내는 것을 기대한다. 위의 식(4)를 만족하는 합성 데이터를 만들기 위해, 두 손실 값의 경사도 차이를 최소화하는 합성 데이터 셋 $\mathcal{S}_m$ 을 학습한다.

[0067] $\min_{\mathcal{S}_m} \mathbb{E}_{\theta \sim \mathcal{P}_{\theta}} [D(\nabla_{\theta} F(\theta; \mathcal{S}_m), \nabla_{\theta} F(\theta; \mathcal{D}_m))] \quad (5)$

[0068] $D(\nabla F_{\mathcal{S}}(\theta), \nabla F_{\mathcal{D}}(\theta)) = \sum_{l=1}^L d(\nabla F_{\mathcal{S}}(\theta^{(l)}), \nabla F_{\mathcal{D}}(\theta^{(l)})) \quad (6)$

[0069] $d(\nabla F_{\mathcal{S}}(\theta^{(l)}), \nabla F_{\mathcal{D}}(\theta^{(l)})) = 1 - \frac{\nabla F_{\mathcal{S}}(\theta^{(l)}) \cdot \nabla F_{\mathcal{D}}(\theta^{(l)})}{\|\nabla F_{\mathcal{D}}(\theta^{(l)})\| \|\nabla F_{\mathcal{S}}(\theta^{(l)})\|}$

[0071] $\nabla_{\theta} F(\theta; \mathcal{S}_m), \nabla_{\theta} F(\theta; \mathcal{D}_m)$ 은 각각 $\mathcal{S}_m, \mathcal{D}_m$ 으로 입력으로 한 경사도이다. 여기서 딥러닝 모델 θ 는 $\mathcal{P}_{\theta}$ 분포에서 랜덤 초기화한 모델이다. 로스 비교 함수 $D(\cdot, \cdot)$ 은 두 경사도의 차이를 1-(코사인 유사도)로 사용하며, 모델의 레이어를 $l = 1, 2, \dots, L$ 이라 할 때, 레이어 단위로 손실 값을 연산한다. $D(\nabla F_{\mathcal{S}}, \nabla F_{\mathcal{D}})$ 은 식(5)의 식을 간단히 표기한 식이다.

- [0073] 도 4는 본 발명의 일 실시예에 따른 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 장치의 구성을 나타내는 도면이다.
- [0074] 제안하는 데이터 집합 압축을 통한 연합학습 학습 안정도 향상 장치는 서버(410) 및 클라이언트(420)를 포함하고, 상기 서버(410)는 통신부(411) 및 모델 업데이트부(412)를 포함하고, 상기 클라이언트(420)는 연합 학습부(421) 및 합성데이터 학습부(422)를 포함한다.
- [0075] 본 발명의 실시예에 따른 서버(410)의 통신부(411)는 전체 클라이언트 중 적어도 하나의 클라이언트를 선택하여 글로벌 모델을 배포한다.
- [0076] 본 발명의 실시예에 따른 클라이언트(420)의 연합 학습부(421)는 상기 선택된 일부 클라이언트들의 미리 저장된 데이터에 대해 각 클래스 별로 언더샘플링하여 데이터 압축을 수행하고, 상기 압축된 데이터를 이용하여 모델을 학습한 후 해당 클라이언트에서 업데이트된 로컬 모델과 함께 상기 압축된 데이터를 서버에 전송한다.
- [0077] 상기 연합 학습부(421)는 클라이언트들이 각 클래스 별로 언더샘플링된 데이터를 해당 클라이언트에서 업데이트된 모델과 함께 서버에 전송하고, 상기 서버는 클라이언트들의 클래스 개수에 비례한 일정한 양의 데이터를 수신하여 클라이언트들이 불균형한 클래스 데이터를 가지고 있는 경우라도 균형된 서버 데이터 집합을 획득하도록 하여 개인 정보를 보호하는 동시에 글로벌 모델의 안정된 성능을 유지하기 위해 개인정보가 식별되지 않는 형태로 각 클래스의 데이터를 압축하고, 상기 클라이언트들이 각 클래스 별로 언더샘플링된 상기 압축된 데이터를 해당 클라이언트에서 업데이트된 모델과 함께 서버에 전송한다.
- [0079] 본 발명의 실시예에 따른 서버(410)의 모델 업데이트부(412)는 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트한다.
- [0080] 상기 서버는 라운드마다 변화하는 클래스 분포를 제한하여 클라이언트의 클래스 개수만큼 압축된 데이터를 수신 받게 되고, 해당 라운드의 전체 클래스 분포를 리밸런싱한 상기 압축된 데이터의 세트를 이용하여 새로운 글로벌 모델로 업데이트한다.
- [0081] 본 발명의 실시예에 따른 클라이언트(420)의 합성데이터 학습부(422)는 상기 서버가 상기 클라이언트들의 로컬 모델을 취합해 새로운 글로벌 모델로 업데이트하면, 상기 선택된 일부 클라이언트들 각각에 미리 저장된 데이터로 합성데이터를 학습한다.
- [0082] 상기 합성데이터 학습부(422)는 합성데이터의 학습 속도를 향상하기 위해 추가 모델 학습 과정을 제거하고, 학습률을 조정하여 학습률에 따른 학습의 불안정성을 제거하기 위해 경사도 클리핑(gradient clipping) 기법을 최적화에 추가한다.
- [0083] 본 발명의 실시예에 따른 서버(410)의 모델 업데이트부(412)는 상기 클라이언트들의 합성데이터로 취합된 모델을 미세조정한다.
- [0085] 도 5는 본 발명의 일 실시예에 따른 FedDC 학습 알고리즘을 나타내는 도면이다.
- [0086] FedDC의 자세한 학습 과정을 도 5의 알고리즘에 나타내었다. 본 발명의 실시예에 따르면, 연합학습의 라운드 기 본 3가지 단계에 추가적으로 두 가지 과정이 더해진다. 클라이언트가 개인 데이터로 합성데이터를 학습하는 과정과 서버가 클라이언트들의 합성데이터로 취합된 모델을 미세조정하는 과정이다. 합성데이터 학습을 위해 총 K 번의 이터레이션을 사용하며, 각 이터레이션 동안 θ 모델은 분포 P_{θ} 에서 새로 랜덤 초기화하여 사용된다. 이때, θ 는 합성 데이터 학습에만 사용되기 때문에 연합학습 알고리즘에는 영향을 끼치지 않는다. 한 번의 이터레이션 동안, 클라이언트의 실제 로우 데이터는 같은 클래스 c 에서만 미니배치 B_D^c 를 샘플링하며, 같은 클래스의 합성데이터 S_i^c 와 비교해 식(6) 손실값을 계산한다. 그 후, 클라이언트들이 개인 데이터로 학습된 모델과 합성데이터를 서버에 전송하면 서버는 클라이언트들의 모델을 취합 후 합성데이터 집합으로 미세조정 한다:

$$\tilde{\theta}^t = \theta^t - \eta \nabla \ell_s(\theta^t) \quad (7)$$

[0088] 여기서, 합성데이터 집합은 $S = \sum_{i=1}^M S_i$ 이며, $\ell_s = \frac{1}{|S|} \sum_{s \in S} \ell_{ce}(\theta^t; s)$ 이다. 연합학습에선 클라이언트마다 제한된 컴퓨터 자원으로 빠르게 학습하는 것이 중요하므로, 좀 더 빠른 학습 방법을 필요로 한다. 기존 데이터 집합 합성 기법은 경사도를 매칭하는 과정에서 파라미터 θ 를 일정량 학습하였다. 그러나, 본 발명에서는 학습 속도를 향상하기 위해, 추가 모델 학습 과정을 제거한다. 대신, 학습률을 조정하여, 학습의 효율을 높인다(예를 들어, $\eta = 0.1$ 내지 $\eta = 3.0$). 높은 학습률에 따른 학습의 불안정성을 제거하기 위해, 본 발명은 다음과 같은 경사도 클리핑(gradient clipping)을 최적화에 추가한다:

[0089]
$$\text{if } \|g\| > v, g \leftarrow \frac{gv}{\|g\|} \quad (8)$$

[0090] 여기서, 경사도 클리핑은 매개변수의 기울기의 노름 $\|g\|$ 이 일정량의 문턱값(threshold)을 넘으면 정규화하는 방식으로, 경사도 바니싱(gradient vanishing)을 방지한다. 본 발명은 합성데이터 학습 중 모델 업데이트 추가 과정은 합성데이터 퀄리티에 큰 차이가 없음을 확인하였다. 이는, 모델 업데이트는 모델 랜덤 초기화처럼, 그래디언트 매칭에 랜덤성을 부과해주는 방법에 불과하다고 추측된다.

[0091] 또한, 기존 데이터 집합 합성기법의 장점이었던 압축된 데이터가 전체 데이터를 잘 표현해준다는 장점이 연합학습에서는 단점으로 작용한다. 합성데이터는 실제 데이터를 변형한 데이터가 아닌 새로 생성된 인공 데이터이지만, 합성데이터가 잘 학습될수록 그 클래스의 특징이 잘 보일 수 있다. 따라서, 개인 프라이버시를 중요시하는 연합학습에서는 합성데이터로 인해 클라이언트의 원본 데이터를 유추할 수는 문제가 발생한다. 즉, 합성데이터의 퀄리티와 개인 프라이버시의 트레이드 오프(Trade-off)가 생기고, 잘 학습된 합성데이터일 수록 프라이버시를 해치는 모순이 발생한다. 따라서, 본 발명에서는 일반적인 모델학습에서 사용하던 모멘텀과 가중치 감쇠를 활용하여 합성데이터를 학습한다. 이를 통해, 합성데이터의 이미지는 노이즈 데이터처럼 구분이 안 가도록 망가진 이미지로 변형되지만, 기존 합성데이터를 활용한 알고리즘과 비교해도 망가진 이미지를 활용한 알고리즘이 유사한 성능을 유지함을 확인하였다.

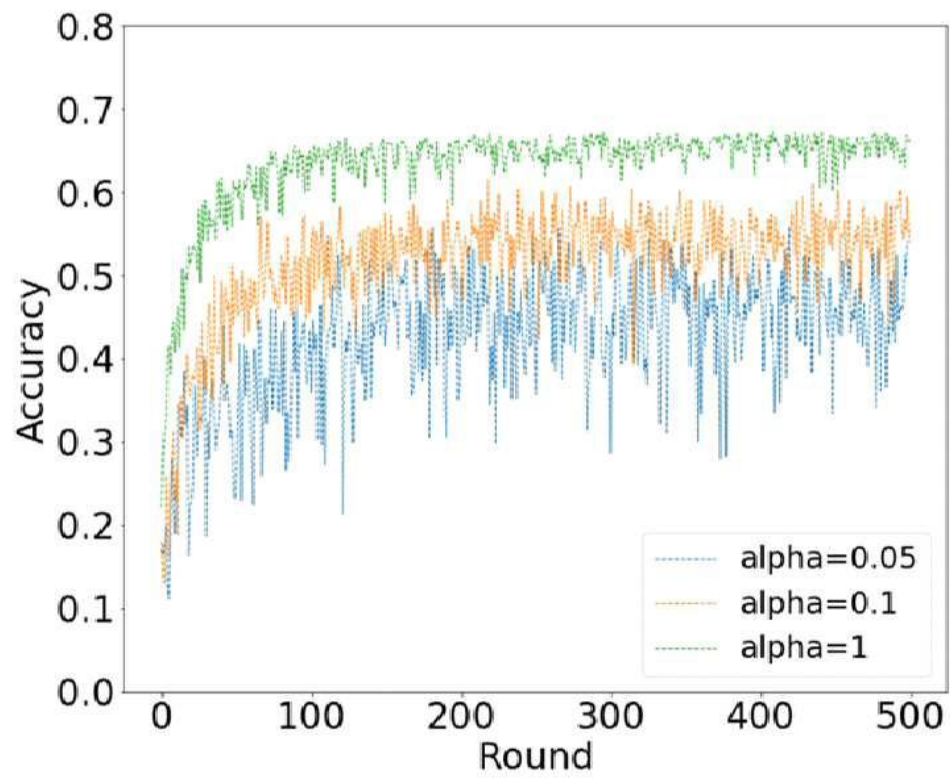
[0092] 본 발명의 실시예에 따르면, 이를 FedDC와 구분한 FedDC+라 제안하며, 이를 통해 개인 프라이버시를 유지하면서도 안정적인 모델을 보장받을 수 있다.

[0093] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

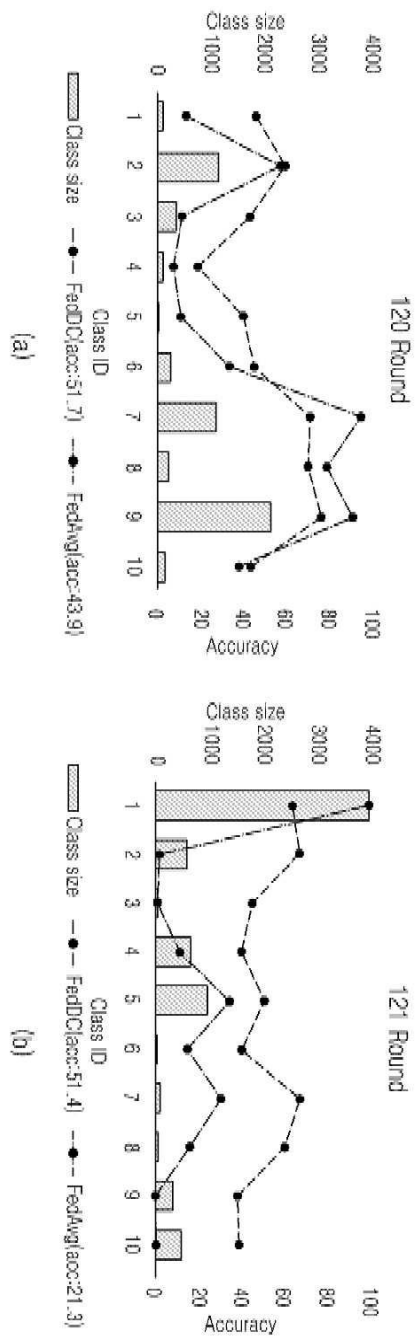
[0094] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

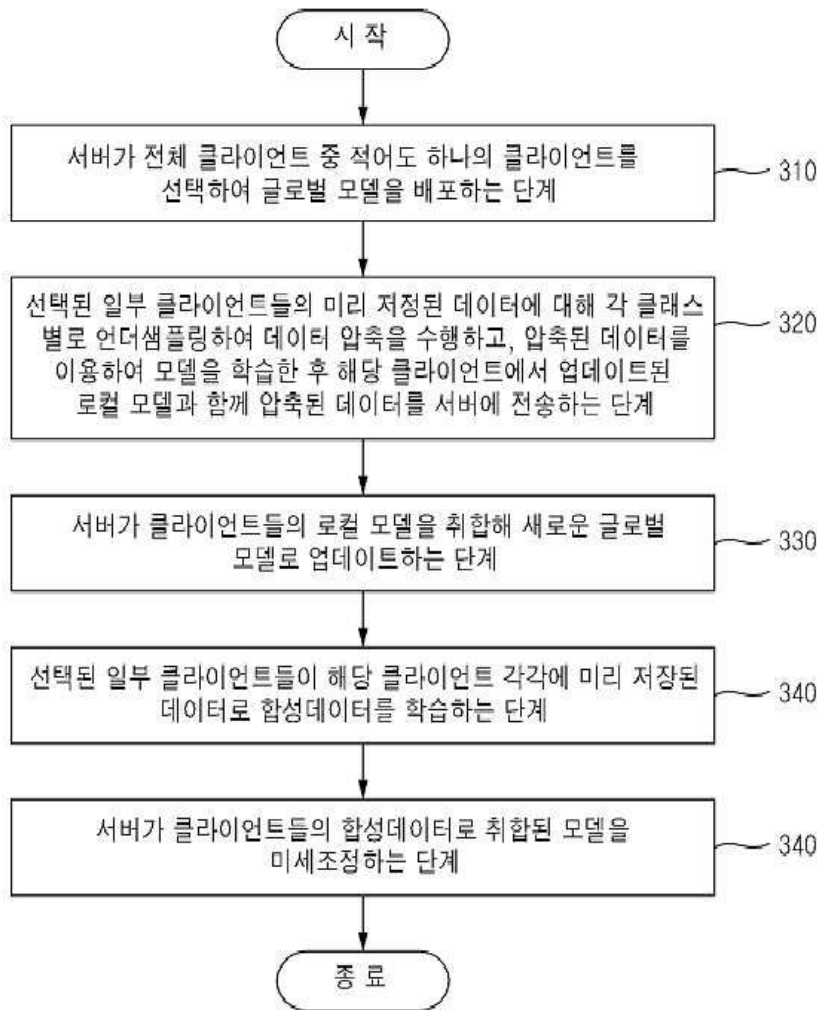
도면1



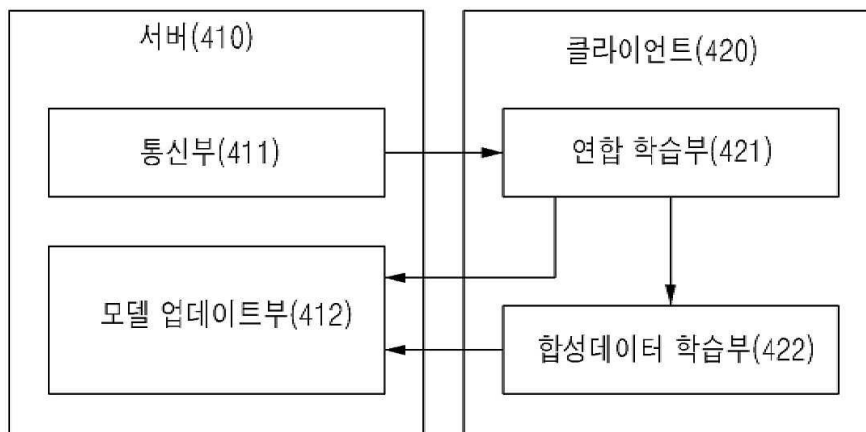
도면2



도면3



도면4



도면5

Algorithm 1 FedDC

Input: random initialize global model θ^t
for each round $t = 0, 1, \dots, T - 1$ **do**
 Select a random subset $\mathbb{S}_t$ for the clients M
 Send the global model θ^t to $\mathbb{S}_t$
 for each client $i \in \mathbb{S}_t$ **in parallel do**
 $\theta_i^t \leftarrow \text{ClientUpdate}(\theta^t, \mathcal{D}_i)$ using Eq. (3)
 $S_i \leftarrow \text{TrainCondensationData}(i, \mathcal{D}_i)$
 send the θ_i^t, S_i to server
 end for
 Aggregate the received client models using Eq. (2)
 Fine-tuning with condensation data using Eq. (7)
end for

TrainCondensationData ($i, \mathcal{D}_i$):
for $k = 0, 1, \dots, K - 1$ **do**
 Initialize $\theta \sim P_\theta$
 for $c = 0, 1, \dots, C - 1$ **do**
 Sample a minibatch $\mathcal{D}_i \sim B_{\mathcal{D}}^c$
 Compute $f_S(\theta) = \frac{1}{|\mathbb{S}_i^c|} \sum_{x \in \mathbb{S}_i^c} \ell_{ce}(\theta; x)$
 and $f_{\mathcal{D}}(\theta) = \frac{1}{|B_{\mathcal{D}}^c|} \sum_{x \in B_{\mathcal{D}}^c} \ell_{ce}(\theta; x)$
 Update $S_i^c = S_i^c - \eta \cdot \nabla D(\nabla F_S(\theta), \nabla F_{\mathcal{D}}(\theta))$
 end for
end for
return condensation dataset S_i
