The Journal of the Convergence on Culture Technology (JCCT)
Vol. 9, No. 3, pp.647-654, May 31, 2023. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2023.9.3.647

JCCT 2023-5-74

멀티 파티 시스템에서 딥러닝을 위한 프라이버시 보존 기술

Privacy Preserving Techniques for Deep Learning in Multi-Party System

고혜경*

Hye-Kyeong Ko*

요 약 딥러닝은 이미지, 텍스트와 같이 복잡한 데이터를 분류 및 인식하는데 유용한 방법으로 딥러닝 기법의 정확도는 딥러닝이 인터넷상의 AI 기반의 서비스를 유용하게 하는데 기초가 되었다. 그러나 딥러닝에서 훈련에 사용되는 방대한 양의 사용자 데이터는 사생활 침해 문제를 야기하였고 사진이나 보이스와 같이 사용자이 개인적이고 민감한 데이터를 수집한 기업들이 데이터들을 무기한으로 소유한다. 사용자들은 자신의 데이터를 삭제할 수 없고 사용되는 목적도 제한할 수 없다. 예를 들면, 환자 진료기록에 대한 딥러닝 기술을 적용하기 원하는 의료기관들과 같은 데이터 소유자들은 사생활과 기밀유지 문제로 환자의 데이터를 공유할 수 없고 딥러닝 기술의 혜택을 받기 어렵다. 우리는 멀티 파티 시스템에서 다수의 작업자들이 입력 데이터집합을 공유하지 않고 신경망 모델을 공동으로 사용할 수 있는 프라이버시 보존 기술을 적용한 딥러닝 방법을 설계한다. 변형된 확률적 경사 하강에 기초한 최적화 알고리즘을 이용하여 하위 집합을 선택적으로 공유할 수 있는 방법을 이용하였고 결과적으로 개인정보를 보호하면서 학습 정확도를 증가시킨 학습을 할 수 있도록 하였다.

주요어 : 딥러닝, 프라이버시 보존, 신경망, 멀티-파티 시스템, 확률적 경사 하강

Abstract Deep Learning is a useful method for classifying and recognizing complex data such as images and text, and the accuracy of the deep learning method is the basis for making artificial intelligence-based services on the Internet useful. However, the vast amount of user da vita used for training in deep learning has led to privacy violation problems, and it is worried that companies that have collected personal and sensitive data of users, such as photographs and voices, own the data indefinitely. Users cannot delete their data and cannot limit the purpose of use. For example, data owners such as medical institutions that want to apply deep learning technology to patients' medical records cannot share patient data because of privacy and confidentiality issues, making it difficult to benefit from deep learning technology. In this paper, we have designed a privacy preservation technique-applied deep learning technique that allows multiple workers to use a neural network model jointly, without sharing input datasets, in multi-party system. We proposed a method that can selectively share small subsets using an optimization algorithm based on modified stochastic gradient descent, confirming that it could facilitate training with increased learning accuracy while protecting private information.

Key words : Deep Learning, Privacy Preserving, Neural Network, Multi-Party System, Stochastic Gradient Descent

*정회원, 성결대학교 컴퓨터공학과 조교수 (단독저자)
접수일: 2023년 3월 10일, 수정완료일: 2023년 3월 25일
게재확정일: 2023년 4월 6일

Received: March 10, 2023 / Revised: March 25, 2023
Accepted: April 6, 2023
*Corresponding Author: ellefgt@sungkyul.ac.kr
Dept. of Computer Engineering, Sungkyul Universkty, Korea

Ⅰ. 서 론

인공신경망에 기초한 딥러닝 기법의 최근의 발달은 이미지, 텍스트 인식, 언어 번역과 같이 오래된 AI의 연구에 대한 돌파구를 마련하였다[1]. 구글, 페이스북, 애플과 같은 대기업들은 그들의 사용자로부터 수집된 대용량 훈련 데이터를 이용하였고 대규모의 딥러닝을 효율적으로 사용하기 위해 많은 GPU의 연산을 이용하였다[2, 3]. 딥러닝을 이용한 연구들은 학습에 좋은 결과를 가지고 왔으나 학습을 위해 사용하는 훈련 데이터는 심각한 개인정보 문제를 제기하였다. 중앙 집중적으로 수집된 많은 개인의 글, 사진, 영상 등은 개인정보 위협과 함께 증가되었다. 예를 들면, 기업들이 수집한 데이터를 지울 수 없고 어떻게 이용되는지 통제할 수 없으며 이를 통해 알아낼 수 있는 것이 무엇인지에 대해 관여할 수 없다. 또한, 이미지와 음성녹음은 얼굴, 차량등록번호, 다른 사람들의 음성과 같은 민감한 것들을 포착한다. 뿐만 아니라, 많은 사용자로부터 수집한 빅데이터에 대한 인터넷 거대기업의 독점은 이 데이터로 학습된 AI 모델의 독점으로 이어졌고 사용자들은 개인비서, 외국어 웹페이지의 기계번역의 도움을 받을 수 있지만 공동의 데이터로 만들어진 대부분의 모델들은 그것을 만든 기업들이 소유하게 된다[1].

많은 영역에서 사람들의 데이터를 공유하는 것은 법이나 규정으로 허락되어지지 않고 결과적으로 임상연구자들은 그들의 기관에 속해있는 데이터집합으로만 딥러닝을 수행할 수 있다. 훈련할 수 있는 데이터집합이 증가하고 다양해질수록 신경망 모델을 이용하는 것이 모델을 훈련시킬 때 도움이 된다고 알려져 있다. 그러나 연구자들이 모델을 훈련시킬 때 다른 기관의 데이터를 활용할 수 없기 때문에 결국은 학습 모델이 열악해질 수밖에 없다[4, 5].

본 논문은 데이터집합의 유용성과 개인정보의 프라이버시를 보존할 수 있는 공동 딥러닝을 위한 시스템을 설계한다. 제안된 시스템은 다양한 사람들이 입력으로 신경망 모델을 학습할 수 있고 입력한 데이터를 공유하지 않으면서 다른 참가자들로부터 혜택을 얻을 수 있는 방법을 제안한다. 제안된 방법은 훈련 중에 매개변수를 선택적으로 공유하여 참가자들이 훈련 입력의 공유 없이 다른 참가자들의 모델로부터 이익을 얻도록 한다.

본 논문의 구성은 다음과 같다. 2장에서는 기계학습과 딥러닝에 대한 경사 하강법에 대한 기존 연구들을 살펴보고 3장에서는 본 논문에서 제안하는 선택적 확률적 경사 하강법을 이용한 방법을 통해 제안된 방법이 어떻게 작동하는지 살펴본다. 4장에서는 제안된 기법에 대해 실험을 통해 분석하고 마지막으로 5장에서는 결론을 통해 제안된 방법이 모델의 정확도를 감소시키지 않고 참가자의 훈련데이터의 개인정보를 보호하는 부분을 살펴본다.

Ⅱ. 관련연구

기계학습에 있어서 개인정보 보호에 대한 기존 논문들은 주로 딥러닝과 다른 전통적인 기계학습알고리즘을 겨냥하였고 모델을 학습하거나 기존모델의 입력으로 사용되는 데이터의 개인정보, 모델의 개인정보, 모델의 출력에 대한 개인정보에 대한 부분을 다루었다[12]. 안전한 다자간 연산(multi-party computation, MPC)에 기초한 기법들은 다자가 그들만의 입력으로 공동의 기계학습을 수행할 때 연산의 중간단계를 보호하는데 도움을 줄 수 있다[6]. 일반적으로 MPC 기법은 적지 않은 성능 오버헤드를 일으키고 개인 정보 보호 딥러닝에 이를 적용하는 것은 해결되지 않은 문제로 남아있다. 이 모델의 개인정보를 보호하는 기법은 개인정보 보호 확률적 추론, 개인정보 보호 화자 식별, 암호화된 데이터 연산을 포함한다 [7, 8]. 본 논문의 목적은 각 참가자가 개인적으로 사용할 수 있는 신경망을 공동으로 훈련시키는 것이다.

딥러닝은 복잡한 데이터로 비선형특징과 함수를 학습하는 과정이다. 딥러닝은 이미지 인식, 담화 인식, 안면 탐지에 있어 과거의 기법들을 능가하는 것으로 나타났다[1]. 딥러닝은 암과 유전학과 관련된 생체의학 데이터를 분석하는데 유망하다[1, 9]. 이러한 모델들을 만드는데 사용되는 훈련 데이터는 개인정보관점에서 보면 민감한 부분이 있어 개인정보 보호 딥러닝 기법의 필요성을 부각시킨다. 병렬 딥러닝 연구들에 있어

GPU/CPU 클러스터의 확률적 기울기 하강의 병렬화 신경망 훈련 중의 연산 분산화를 위한 기술 등과 같은 최근의 연구들이 있다[9, 10]. 하지만 이러한 기술들은 훈련데이터의 개인정보에 관심을 두지 않고 이 훈련을 통제하는 방법을 사용하였다.

2. 경사 하강을 이용한 망 매개변수 학습

딥러닝은 다차원의 데이터에서 복잡한 특징을 추출하고 이를 이용하여 입력과 출력을 연결시키는 모델을 만드는 것을 목표로 한다. 딥러닝 아키텍쳐는 동시적으로 다층의 망으로 만들어져서 추상적인 특징들은 비선형 함수로 계산한다. 이 논문에서는 훈련입력은 지도학습에 초점을 맞춘다.

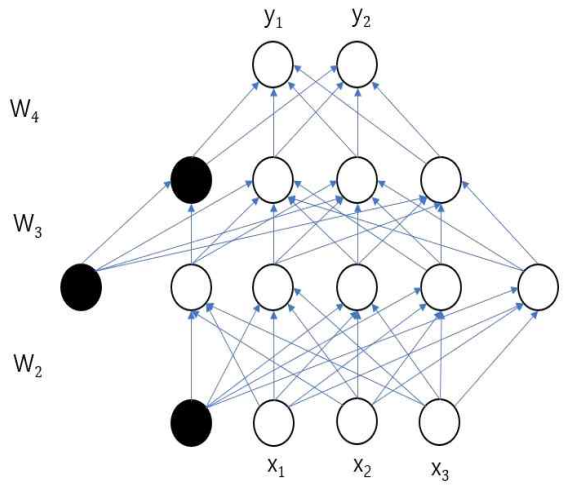

그림 1. 두 개의 숨겨진 층이 있는 신경망
Figure 1. Neural network with two hidden layers

다층 신경망은 딥러닝에서 가장 보편적인 형태로 그림 1은 2개의 은닉층을 가진 신경망을 나타낸다. 그림에서 각각의 노드는 하나의 신경망을 모형화하였고 검은색 노드는 편향 노드를 나타낸다. 행렬 W_k는 각 계층에서 활성화 함수를 계산하는 데 사용되는 가중치를 포함한다. 전형적인 다층망에서 각 신경은 이전 층에 있는 신경의 출력뿐 아니라 내보내는 신경에서 나오는 신호를 받는다. 그리고 전체 입력에 비선형 활성화 함수를 적용해서 계산된다[9, 11]. k층 신경의 출력벡터는 $a_k = f(W_k a_k - 1)$이다.

여기에서 f는 활성화함수이고 W_k는 각각의 입력신호의 기여를 결정하는 가중치행렬이다. 일반적으로, 더 높은 층에서 곧바로 추출된 가공되지 않은 특징으로 이루어지고 이전 층의 출력은 그 모델에 의해 만들어진 추상적 응답과 상응한다. 비선형 함수 f와 가중치 행렬은 각각의 층에서 추출된 특징을 결정하고 신경망의 분류 정확성을 최대화하는 매개변수 (가중치의 행렬)의 값에 대한 데이터를 자동으로 학습하는 것이다[9]. 신경망의 매개변수를 학습하는 것은 비선형 최적화의 문제로 지도학습에서 목적 함수는 신경망의 출력이다.

이러한 문제를 해결하기 위해 사용된 알고리즘들은 일반적으로 변형된 경사 하강 방법이다. 경사 하강은 신경망에 대한 매개변수의 집합에서 시작하고 각 단계에서 최적화된 비선형함수의 기울기를 연산하고 기울기를 감소시키기 위해 매개변수를 업데이트한다. 이런 과정을 알고리즘이 최적에 수렴될 때까지 계속한다.

3. 확률적 경사 하강법

확률적 경사 하강 (stochastic gradient descent, SGD)은 추출된 데이터 한 개에 대해서 그래디언트를 계산하고, 경사 하강 알고리즘을 적용하는 방법이다. 전체 데이터를 사용하는 것이 아니라, 랜덤하게 추출한 일부 데이터를 사용하고 학습 중간 과정에서 결과의 진폭이 크고 불안정하며, 속도가 매우 빠르다. 또한, 데이터 하나씩 처리하기 때문에 오차율이 크고 GPU의 성능을 모두 활용하지 못하는 단점이 있다[11]. 이러한 단점들을 보완하기 위해 나온 바업들이 Mimi batch를 이용한 방법이며, 확률적 경사 하강법의 노이즈를 줄이면서 전체 배치보다 더 효율적인 방법이다[9]. 매개변수의 기울기는 모든 사용가능한 데이터에 걸쳐 평균화될 수 있고 큰 데이터집합을 학습하는데 비율적이다. 확률적 경사 하강은 전체 데이터 집합에 있는 극도로 작은 하위집합의 기울기를 계산할 때 가장 단순한 경우, 최대 확률에 대응하는 데이터샘플은 각 최적화 단계에서 무작위로 선택된다.

확률적 경사하강법에서는 가중치 업데이트 수식에서 아래 수식 (1)과 같이 i번째 트레이닝 데이터에 대해서만 계산한 값을 이용한다.

$$W_j = W_j + \eta(y^{(i)} - y^{\wedge(i)})x_j^{(i)} \quad (1)[11]$$

예를 들어, 100만개의 트레이닝 데이터가 있다고 가정하면, 배치 경사 하강법으로 계산하려면 100만 X 100만=1조 번의 연산이 필요하다. 이는 가중치를 업데이트 하는 식에서 식 (1)은 $y^{\wedge}(i)$를 계산하기 위해 100만 번의 덧셈을 해야 하고, W_j는 최종적으로 업데이터하기 위해 전체적으로 또 100만 번의 덧셈을 수행해야 한다.

그러나 확률적 경사 하강법을 적용하면 100만 개의 모든 데이터를 활용하여 머신러닝을 수행하더라고 100만 번의 연산만 필요하게 된다. 확률적 경사 하강법을 이용하면 배치 경사하강법의 근사치로 계산되지만 가중치를 업데이터하는 시간이 빠르기 때문에 실제로는 비용함수의 수렴 값에 더 빨리 도달하게 된다. 본 논문에서 다루는 개인정보보호와 확률적 기울기 하강을 위한 시스템 설계 시 이 식을 이용하여 설계한다.

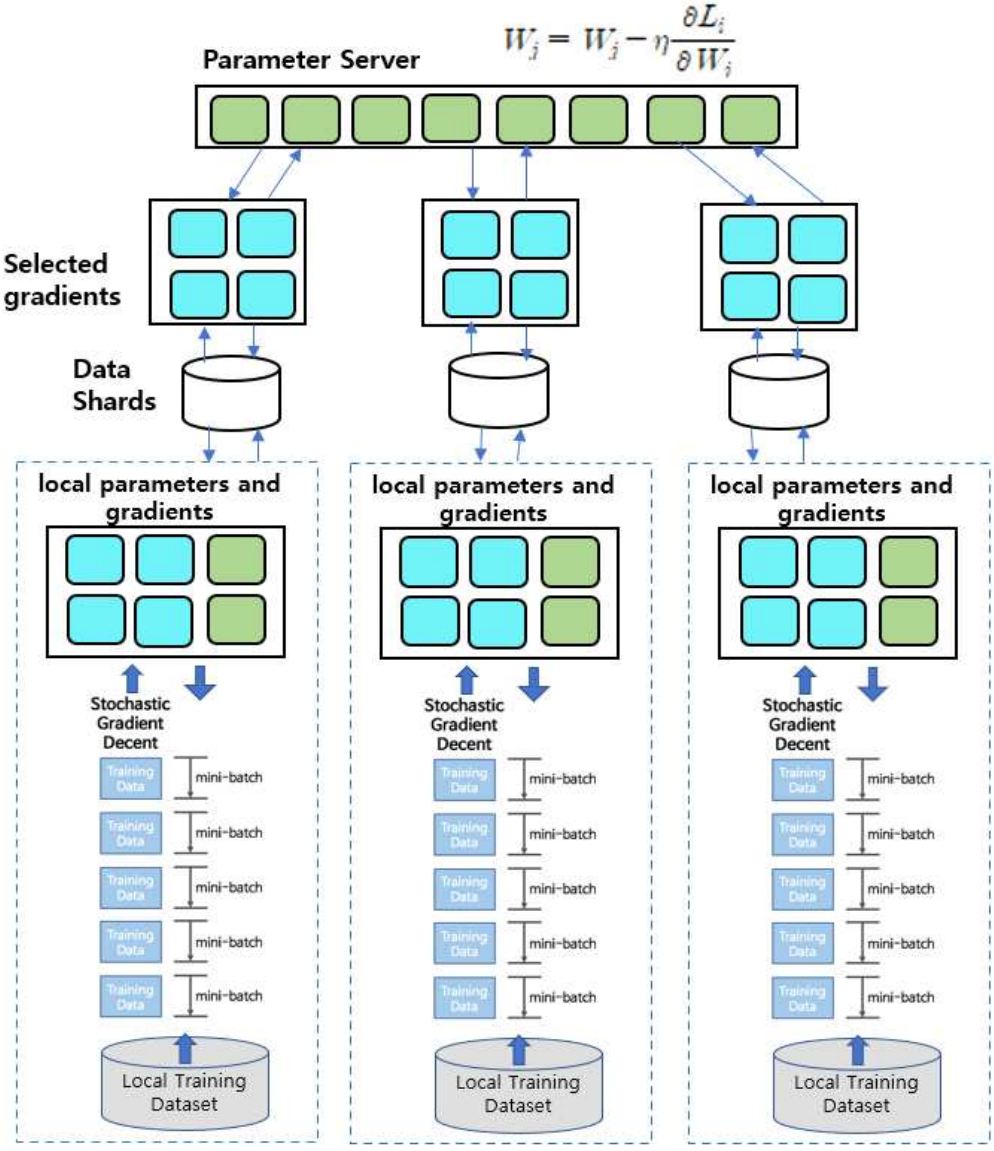

그림 2. 선택적 확률적 경사 하강법의 동작 원리
Figure 2. Principle of operation of selective stochastic gradi

III. 선택적 확률적 경사 하강법

신경망의 매개변수를 학습하는 것은 비선형 최적화의 문제이다. 지도학습에서 목적함수는 신경망의 출력인데 이 문제를 해결하기 위해 사용된 알고리즘들은 일반적으로 경사 하강 (gradient descent, GD)법의 변형이다. 경사 하강법은 신경망에 대한 매개변수의 집합에서 시작하고 각 단계에서 최적화된 비선형함수의 경사를 연산하고 경사를 감소시키기 위해 매개변수를 업데이트하고 이러한 과정들은 알고리즘이 최적에 수렴할 때까지 반복한다. 매개변수의 경사는 모든 사용가능한 데이터에 평균화될 수 있고 경사 하강으로 알려진 알고리즘은 큰 데이터집합을 학습하는 데는 비효율적이다. SGD는 전체 데이터 집합에 있는 작은 하위집합의 경사를 계산하는 방법으로 본 논문에서 제안한 방법은 선택적인 공동 딥러닝 방법이다. 본 논문에서 가정하는 방법은 경사 하강 시 다른 매개변수를 업데이트 시 독립적으로 수행하고, 선택적인 확률적 경사 하강법은 각각의 학습 반복에서 적은 매개변수를 업데이트 하게 된다.

1. 선택적인 매개변수 업데이트

선택적인 SGD에서 학습자는 첫째로, 각각의 반복에서 업데이트 될 매개변수의 부분을 선택한다. 선택 시 현재 값에서 더 큰 기울기를 가진 매개변수를 선택한다.

$$W_j = W_j - \eta \frac{\partial L_i}{\partial W_j} \quad (2)\ [11]$$

식 (2)는 확률적 경사 하강법을 나타낸다. 식 (2)에서 W는 갱신할 가중치 매개변수, $\frac{\partial L_i}{\partial W_j}$는 매개변수에 대한 결과 값의 기울기를 나타낸다. η는 학습률을 나타내는데, 실제로는 0.01과 같이 미리 정해서 사용한다. 각 미니배치 i에 대해 SGD 같이 모든 매개변수 W_j에 대한 손실함수 $\frac{\partial L_i}{\partial W_j}$를 연산한다. 선택적인 SGD는 둘 이상의 참가자들이 독립적이면서 동시에 학습하는 것을 목표로 한다.

그림 2는 본 논문에서 제안한 선택적 SGD 방법의 동작원리를 나타낸다. 국지적 훈련에서 참가자들은 서로 몇몇 매개변수들에 대해 연산된 기울기를 공유하고 각 참가자는 어떤 기울기를 공유할지, 얼마나 자유할지를 통제한다. 주어진 매개변수에 대해 연산된 모든 기울기의 합이 매개변수의 국지적 최적화에 대한 하강의 규모를 결정한다.

2. 훈련 방법

참가자들은 실제로 데이터를 보지 않지만 각각의 다른 참가자들의 훈련 데이터의 혜택을 받고 그들만의 제한된 훈련 데이터를 통해 더 정확한 모델을 만들 수 있다. 참가자들은 기울기를 직접 공유할 수 있고 신뢰할 수 있는 중앙 서버를 통해서나 업데이트 출처를 숨기거나 안전한 연산을 할 수 있다. 서버는 매개변수의 값에 기울기를 더하고 각 참가자들은 서버로부터 그 매개변수의 하위 집합을 다운로드하고 이를 이용하여 자신의 모델을 업데이트하게 된다. 주어진 매개변수의 다운로드 기준은 업데이트 빈도와 매개변수에 더해진 기울기의 이동평균을 이용한다.

예를 들어, 훈련에 이용할 수 있는 국지적 개인 데이터 집합을 가진 참여자들 N이 있다고 가정하자. 모두가 사용가능한 매개변수들의 최신 값을 유지하는데 책임을 지는 매개변수 서버가 있다고 가정하고 이 매개변수는 실제 서버에서 구현되거나 분산시스템에 의해 설계된다. 표 1은 실험에서 사용되는 파라미터의 종류를 나타낸다.

표 1. 실험 파라미터
Table 1. Simulation parameters

η	- 확률적 경사 하강법의 학습 속도
θ_d, θ_u	- 다운로드 및 업로드를 위해 선택된 매개변수 비율
v	- 다른 참가자들과 공유한 경사도 값에 대한 경계
τ	- 경사도 선택을 위한 임계값

각 참가자는 매개변수를 초기화하고 데이터 집합을 만들어 훈련을 시작한다. 알고리즘은 참가자들이 매개변수 서버로 선택된 신경망 매개변수의 기울기를 업로드하고 각 SGD의 최신 매개변수 값을 다운로드할 수 있게 하는

교환 프로토콜을 포함한다.

알고리즘 1은 참가자의 선택적 SGD 알고리즘의 의사코드를 보여준다.
참가자들은 독립적으로 매개변수 집합에 수렴할 수 있고 이러한 매개변수가 단일 참가자의 훈련 데이터 집합에 과적되는 것을 피할 수 있고 각 참가자는 다른 참가자들과 상호작용하지 않고도 새로운 데이터로 독립적이고 개인적으로 평가할 수 있다. 매개변수들은 학습되고 있는 실제 신경망 매개변수와 반대로 공동 학습과정을 통제할 수 있고 각각의 참가자들은 신경망 매개변수 W를 유지할 수 있고 매개변수 서버는 별도의 매개변수 벡터 W를 유지한다.

알고리즘 1. 참자가의 선택적 SGD 알고리즘
Algorithm 1. Selective SGD algorithm for participants

초기 파라미터 W와 학습 속도 η를 선택한다.
비슷한 최소값을 얻을 때까지 반복한다.

1. 서버로부터 $|\theta_d * W$ 파라미터들을 다운로드하고 대응되는 지역 파라미터들을 재배치한다.
2. 로컬 데이터셋에서 SGD를 실행하고 (1)에 대응되는 로컬 파라미터 W를 업데이트한다.
3. SGD에 대하여 모든 로컬 파라미터들내에서 바뀐 벡터 △W의 경사도를 계산한다.
4. 아래의 조건에 따라 선택된 $|\theta_u * W$의 경사도의 인덱스들의 집합인 S를 파라미터 서버에 업로드한다.

분산된 SGD는 일반적으로 어떤 매개변수들이 다른 참가자들에 의해 업데이트될 필요가 있는지 혹은 업데이트 속도에 대해 어떠한 가정을 하지 않는다. 몇몇 참가자들은 더 나은 연산능력과 처리량 때문에 더 많은 업데이트를 할 수 있다.

- 기울기를 선택하고 공유하는 방법은 두 개의 기준을 고려한다.
  1. 경사 하강 알고리즘에서 큰 값을 가려내고 정확하게 θ_u 값을 선택한다.
  2. 임계치 τ보다 큰 값의 임의의 하위집합을 선택하는 것으로 τ보다 큰 기울기의 수는 매개변수의 θ_u 부분보다 작을 수 있기 때문에 더 적은 기울기가 공유된다.

선택된 기울기 △ W를 업로드하기 전에 그 값들은 $[-v, v]$ 범위 내로 잘리고 이러한 값들이 훈련 데이터에 대한 너무 많은 정보를 흘리지 않게 하기 위해 무작위 노이즈를 더한다. 참가자는 bound(△ W, v)로 △ W를 업데이트하고 업로드하기 전에 무작위 노이즈를 추가한다. 미니배치는 M 크기의 무작위로 선택된 훈련 데이터 집합이다.

Ⅳ. 실험 및 결과

본 논문에서는 관련연구에서 딥러닝 데이터 셋으로 많이 사용되는 MNIST 데이터 셋을 이용하여 제안된 방법을 평가한다[11]. MNIST 데이터 셋은 손 글씨 숫자가 쓰인 28X28 크기의 이미지의 집합으로 이미지의 중앙에 숫자가 위치하도록 정규화 되었다. 데이터 셋은 60,000개의 훈련용 셋과 10,000개의 실험용 셋으로 구성된다. 실험에서 사용되는 훈련세트는 평균을 빼고 데이터 샘플의 표준편차로 나우어서 데이터 집합을 정규화하였다. MNIST에 대한 신경망 입력 층의 크기는 1024이고 학습 목적은 입력을 10개의 가능한 숫자 중 하나로 분류하는 것이며 출력 층의 크기는 10이다.

실험평가에서는 합성 곱 신경망 (convolutional neural network, CNN) 구조를 사용한다. CNN은 이미지와 영상 인식으로 널리 쓰이며 본 논문에서는 Torch7 nn 패키지를 이용하여 출력된 CNN 값을 이용하여 평가하였다[11],[12]. 이 그림들은 각층에서 사용된 함수와 계층 간의 연결을 보여준다. 표 2는 매개변수의 수를 나타내고 표 3은 인공신경망 CNN의 파라미터의 수를 나타낸다.

표 2. 데이터셋의 크기
Table 2. Size of datasets

	MNIST
train	60,000
test	10,000

표 3. 인공신경망 파라미터의 수
Table 3. Number of neural network parameters

	MNIST
CNN	100,000

실험에서는 두 개의 시나리오를 가지고 분석하였다. 첫 번째는 전체 데이터집합에 대해 중앙 집중화된 SGD 방법으로 모든 훈련 데이터가 하나의 데이터 집합으로 모이고 표준 확률적 경사 하강을 활용하여 해당 데이터 집합으로 훈련되는 방법으로 개인정보를 침해하는 시나리오이다. 두 번째는 어떤 기울기를 업로드할지 선택하는 기준을 이용하여 선택적 SGD 방법을 이용하는 방법으로 개인정보 침해를 줄일 수 있다.

SGD 미니 배치 크기는 1와 32로 학습속도 (η=0.01과 0.001)에 대한 설정을 평가하였다. MNIST 600개 데이터 샘플이 사용되었고 선택적 SGD에서 공유를 위해 선택된 매개변수의 부분 θ_u={1, 0,1, 0.01, 0.001}에서 값을 갖는다. 다운로드 시 매개변수 θ_d는 1로 설정되었다. 표 4는 CNN 아키텍처에서 선택적 SGD에 의해 달성된 최대 정확도를 나타낸다.

표 4. CNN 아키텍처에서 선택적 SGD에 의해 달성된 최대 정확도
Table 4. Maximum accuracy achieved by selective SGD for CNN

	SGD	0.1	0.01	0.001	독립SGD
MNIST.CNN	0.9919	0.9912	0.9852	0.9652	0.9214

미니 배치는 사이즈는 1이고 독립 SGD와 비교하여 정확도를 계산하였다. 전통적인 경사 하강과 비교하여 본 논문에서 제안한 방법의 효율성을 보여주기 위해서 MNIST 데이터 집합으로 합성 곱 신경망을 훈련했을 때 얻은 SSGD와 SGD의 정확성을 평가하였다. 경사 하강 단계에서 기울기의 작은 부분으로 공유함으로써 SGD와 거의 같은 정확도를 달성하였고 선택적인 매개변수 공유도 전체적인 SGD에는 거의 영향을 주지 않았다. 훈련과정에서 미니 배치 1로 설정하는 것은 높은 확률성을 달성하고 매우 빨리 수렴한다.

V. 결 론

전통적인 딥러닝에서 모든 훈련 데이터는 일반적으로 학습을 수행하는 회사에 노출되고 데이터를 기여한 개인들은 어떠한 통제 권한을 갖지 못한다. 그들의 민감한 정보는 해당 회사에 데이터 저장소를 손상시킨 공격자들에게 데이터에 접속할 수 있는 기관에 유출될 수 있는 위험이 있다. 또한, 전통적인 딥러닝 학습법에서 학습된 모델은 데이터 소유자가 직접적으로 사용할 수 없고 사용 시에는 모델을 소유하고 있는 기업에 입력을 통해 개인정보 위험에 노출되고 있다. 본 논문에서 제안한 시스템은 딥러닝과 관련된 개인정보 위협들을 해결하는 것을 목표로 한다. 제안된 프라이버시 보존 딥러닝 시스템은 이러한 문제들을 해결하고 훈련 데이터의 개인정보를 보호하고 학습 목적에 대한 권리를 보장할 수 있다. 본 논문에서는 선택적 확률 경사 하강법에 기초한 선택적 훈련 기법을 제안하였다. 제안된 방법은 CNN 신경망에서 효과적 있었고 결과 모델의 정확도를 감소시키지 않고 참가자의 훈련데이터의 개인정보를 보호한다. 따라서 데이터소유자가 기밀유지 문제로 인해 데이터를 공유할 수 없는 부분에서 딥러닝 학습을 시킬 경우 효율적인 개인 정보 보호를 제공한다.

본 논문에서 제안된 방법은 이미지 분류 알고리즘의 기준으로 사용되는 MNIST 데이터 집합으로 평가하였고 분산된 참가자들에 의해 만들어진 모델들의 정확도는 단일한 단체가 데이터집합 전체를 쥐고 이를 이용하여 중앙 집중화된 모델을 훈련시켜 사생활을 침해하는 경우와 비슷하였다. MNIST 데이터 집합의 경우 참가자들이 매개변수 10%를 공유했을 경우, 99.12%의 정확도를1%일 때 98.52%의 정확도를 얻었다. 그에 비해 중앙 집중화된 사생활 침해 모델의 최대 정확도는 99.19%였다. 제안된 방법은 모든 훈련 데이터를 직접적으로 드러내지 않기 때문에 시스템상의 유출은 신경망 매개변수의 부분에 걸친 간접적인 부분이다. 제안된 방법은 CNN 신경망에서 효과적 있었고 결과 모델의 정확도를 감소시키지 않고 참가자의 훈련데이터의 개인정보를 보호한다. 따라서 데이터소유자가 기밀유지 문제로 인해 데이터를 공유할 수 없는 부분에서 딥러닝 학습을 시킬 경우 효율적인 개인 정보 보호를 제공한다. 제안된 방법은 CNN 신경망에서 효과적 있었고 결과 모델의 정확도를 감소시키지 않고 참가자의 훈련데이터의 개인정보를 보호한다. 따라서 데이터소유자가

기밀유지 문제로 인해 데이터를 공유할 수 없는 부분에서 딥러닝 학습을 시킬 경우 효율적인 개인 정보 보호를 제공한다.

향후 연구로 우리는 간접적인 유출도 최소화하기 위해 차등화 된 개인정보를 매개변수 업데이터에 어떻게 적용하는지에 대한 연구를 수행할 예정이다.

References

[1] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R, Prenger, S. Satheesh, S.sengupta, and A. Coates, "Deepspeech: Scaling up end-to-end speech recognition," 2014. DOI: 10.48550/arXiv.1412.5567

[2] S. W. Lee, "Development of a Method for ACF Bonding Based on Machine Vision," The Journal of the Convergence on Culture Technology (JCCT), Vol. 4, No. 3, pp. 209-212, 2018. DOI: 10.17703/JCCT.2018.4.3.209

[3] H. Lee and J. Choi, "Implementation of Smart Ventilation Control System using IoT and Machine Learning," The Journal of the institute of Internet, Broadcasting and Communication (JIIBC), Vol. 20, No. 2, 2020. DOI:10.7236/JIIBC.2020.20.2.283

[4] D. Shultz, "When your voice betrays you," Science, Vol. 347, No. 6221, 2015. DOI: 10.1126/science.347.6221.494

[5] Y. S. Lee, "Analysis on machine learning-as-a-service," International Journal of Advanced Culture Technology (IJACT), Vol. 6, No. 4, pp. 303-308, 2018. DOI: 10.17703//IJACT2018.6.4.303

[6] J. Bos, K. Lauter, and M. Naehrig, "Private predictive analysis on encrypted medical data," Informatics, Vol. 50, pp. 234-243, 2014. DOI: 10.1016/j.jbi.2014.04.003

[7] M. Pathak, S. Rane, W. Sun, and B. Raj, "Privacy preserving probabilistic inference with hidden markov models," In proc. of 2011 IEEE International Conference on acoustics, Speech and Signal Processing, pp. 1-3, 2011. DOI: 10.1109/ICASSP.2011.5947696

[8] P. Xie, M. Bilenko, T. Finley, R. Gilad-Bachrach, K. Lauter, and M. Naehrig, "Crypto-nets: Neural networks over encrypted data,"arXiv:1412.6181,2014. DOI: 10.48550/arXiv.1412.6181

[9] M. Liang, Z. Li, T. Chen, and J. Zeng, "Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach," In Proc. of IEEE/ACM transactions on computational iology and bioinformatics, Vol. 12, No. 4, pp. 928-937, 2015. DOI:10.1109/TCBB.2014.2377729

[10]M. Pathak, S. Rane, W. Sun, and B. Raj, "Multiparty different privacy via aggregation of locally trained classifiers," In Proc. of International Coference on Neural Information Processing Systems, pp. 1876-1884, 2010.

[11]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition,"IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998. DOI: 10.1109/5.726791

[12]https://github.com/torch/nn

※ 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.NRF-2021R1A2C1012827)