https://doi.org/10.33778/kcsa.2021.21.3.049

# 프라이버시 침해에 대응하는 분할 학습 모델 연구

유 지 현*, 원 동 호*, 이 영 숙**

## 요 약

현대의 인공지능은 사회를 구성하는 필수적인 기술로 여겨지고 있다. 특히, 인공지능에서 프라이버시 침해 문제는 현대 사회에서 심각한 문제로 자리 잡고 있다. 개인정보보호를 위해 2019년 MIT에서 제안된 분할 학습은 연합 학습의 기술 중 하나로 개인정보보호 효과를 지닌다. 본 연구에서는 데이터를 안전하게 관리하기 위해 알려진 차분 프라이버시를 이용하여 안전하고 정확한 분할 학습 모델을 연구한다. 또한, SVHN과 GTSRB 데이터 세트를 15가지의 차등적인 차분 프라이버시를 적용한 분할 학습 모델에 학습시키고 학습이 안정적으로 되는지를 확인한다. 최종적으로, 학습 데이터 추출 공격을 진행하여, 공격을 예방하는 차분 프라이버시 예산을 MSE를 통해 정량적으로 도출한다.

# A Study of Split Learning Model to Protect Privacy

Jihyeon Ryu*, Dongho Won*, Youngsook Lee**

## ABSTRACT

Recently, artificial intelligence is regarded as an essential technology in our society. In particular, the invasion of privacy in artificial intelligence has become a serious problem in modern society. Split learning, proposed at MIT in 2019 for privacy protection, is a type of federated learning technique that does not share any raw data. In this study, we studied a safe and accurate segmentation learning model using known differential privacy to safely manage data. In addition, we trained SVHN and GTSRB on a split learning model to which 15 different types of differential privacy are applied, and checked whether the learning is stable. By conducting a learning data extraction attack, a differential privacy budget that prevents attacks is quantitatively derived through MSE.

Key words : Differential Privacy, Inversion Attack, Neural Networks, Learning Model, Extraction Attack

접수일(2021년 09월 01일), 게재확정일(2021년 09월 30일)

* 성균관대학교 소프트웨어학과
** 호원대학교 IT소프트웨어보안학과(교신저자)

# 1. 서 론

현대 사회에서 필수불가결하게 사용되고 있는 인공지능은 학습 과정에서 대량의 데이터가 사용되고 있어 수집한 데이터에 대한 정보보호 관련 문제가 제기되고 있다. 2021년 스캐터랩에서 개발한 인공지능 챗봇 '이루다'는 학습 과정에서 수집한 60만명의 94억여 건의 문장 중, 이름, 휴대전화번호, 주소 등의 데이터가 챗봇을 사용하는 도중 그대로 노출되어 프라이버시 침해 문제가 발생했다[1]. 또한, 아마존의 인공지능 음성비서 Alexa는 녹음된 사용자들의 명령 내용을 청취하는 도중 사용자들의 주소 등의 개인정보가 노출되는 문제가 발생했다[2]. 이런 개인정보 침해 문제는 대량의 데이터를 확실하게 전처리하지 않으면 인공지능 학습에서 그대로 결과가 나타날 위험이 있다. 또한, 사용자가 클라우드에 자신의 데이터를 보낼 때, 원본 데이터를 그대로 사용하는 경우가 많으므로 사용자가 보내는 데이터만을 탈취해도 쉽게 민감한 정보를 지닌 데이터가 갈취당할 수 있다[3]. 위의 사례처럼 인공지능에서의 개인정보 침해 문제는 어렵지 않게 발생할 수 있다.

인공지능을 사용했을 때, 개인정보 침해 문제가 발생하자 이를 막기 위해 여러 방법이 제안되었다[4-19]. 이중 사용자와 클라우드로 모델을 분리하여 학습하는 방법이 많은 관심을 끌었는데[6-16], 특히 최근 Google과 MIT에서 연구된 연합 학습 방법과 분할 학습 방법이 세계적으로 주목된다[15-16]. 또한, 데이터 보호 방식 중 하나인, 원본 데이터 자체에 적당량의 노이즈를 추가하여 개인정보를 보호하는 차분 프라이버시 기법이 같이 주목받고 있다[17-19]. 따라서, 모델에 노이즈를 주입하는 정도에 따라 학습이 안정적이며 안전한 모델임을 확인하는 정량적인 연구가 필요하다. 본 논문은 사용자와 클라우드로 분리된 모델인 분할 학습 모델에 차분 프라이버시를 적용하여 차분 프라이버시에 따른 모델의 적합도를 평가한다.

본 논문에서는 인공지능 모델의 개인정보를 보호하는 방식인 분할 학습을 사용하여 모델을 구현한다. 또한, 분할 학습을 사용한 모델에 차분 프라이버시를 활용하여 학습 데이터 추출 공격에 저항하는 모델을 구현한다. 모델의 정확성과 안전성을 확인하기 위해 모델의 정확도를 확인하고, 학습 데이터 추출 공격을 시행하여 원본 이미지로 얼마나 복구되는지를 확인한다.

본 논문은 다음과 같이 구성된다. 2장은 배경 지식에 관한 내용으로 연합 학습, 분할 학습, 학습 데이터 추출 공격과 차분 프라이버시에 대해 정리한다. 3장에서는 실험 방법에 관해 설명한다. 실험 방법에는 실험 환경과 실험 모델, 공격과 결과 분석, 그리고 실험 과정에 관해 설명한다. 4장에서는 실험 결과에 관한 내용과 이를 분석하는 내용이 포함되며, 5장에서 결론으로 마무리한다.

# 2. 배경 지식

## 2.1 연합 학습

연합 학습(federated learning)은 2016년 Google에서 기계학습에서 사용 가능한 데이터의 탈중앙화에 대해 처음 제안한 방법으로, 모든 사용자가 전체 모델의 복사본을 가지고 가중치를 업데이트한다[15]. 클라우드는 모든 사용자로부터 가중치 업데이트한 값을 받아 평균을 내어 가중치를 업데이트한다. 이 업데이트된 가중치는 사용자에게 다시 받아지고 이 값이 수렴할 때까지 반복한다.

## 2.2 분할 학습

2018년 처음 발표된 분할 학습(split learning)은 MIT에서 SplitNN이라는 모델을 제안하며 기계학습 모델 원본 데이터 추론을 예방하고 사용자의 개인정보 보호하기 위해 연구되고 있는 분야로, 이미 사용자 모델과 클라우드 모델로 나누어진 모델로부터 시작된다[16]. 사용자가 지닌 모델은 이미 학습된 모델로 더는 가중치가 업데이트되지 않는 특성이 있다.

## 2.3 학습 데이터 추출 공격

학습 데이터 추출 공격은 모델 학습을 위해 사용된 데이터를 탈취하는 공격을 의미한다. 분할학습에서 수행되는 공격은 분할학습 모델을 재구성하여 원본 이미지를 복구하도록 하며[23], Query-free 공격, Black-box 공격, White-box 공격으로 다음 표와 같이 나뉜다.

<표 1> 학습 모델, 학습 데이터 세트, 데이터 쿼리 유무에 따른 분할 학습 모델 공격기법 [23]

| 구분 | 학습 모델 | | 학습 데이터 세트 | | 데이터 쿼리 |
|---|---|---|---|---|---|
| | 매개 변수 | 모델의 구조 | 데이터 | 데이터 분포 | |
| Query-free attack | | | | O | |
| | | O | | O | |
| | | | O | O | |
| | | O | O | O | |
| Black-box attack | | | O | O | O |
| | | | | O | O |
| | | | | | O |
| White-box attack | O | O | | | |

## 2.4 $\epsilon$-차분 프라이버시

차분 프라이버시(differential privacy)는 개인정보의 노출 여부를 판별하는 것이 아닌 노출 정도를 정량화하는 것으로, 개인정보를 보호하기 위해 해당 데이터 세트에 임의의 노이즈를 주입함으로써 개인정보가 제 3자에게 노출되지 않도록 보호하는 기법이다 [17-19]. 이때 데이터베이스에서 개인정보를 사용할 수 있는 정도를 계산하는 정도를 프라이버시 예산(privacy budget)이라 하며, 본 논문에서는 $\epsilon$이라고 표현한다.

임의의 알고리즘 $M$이 모든 이웃 데이터베이스 $D_1$과 $D_2$에 대해 (1)을 항상 만족하면 $M$은 $\epsilon$-차분 프라이버시($\epsilon$-differential privacy)를 만족한다.

$$\Pr[A(D_1) \in S] \leq e^{\epsilon}\Pr[A(D_2) \in S] \quad (1)$$

$\epsilon$이 0이 되면 전혀 프라이버시가 보장되지 않는 상황이 되며, 반대로 큰 값을 갖게 되면 정확성이 떨어지는 특성이 있다.

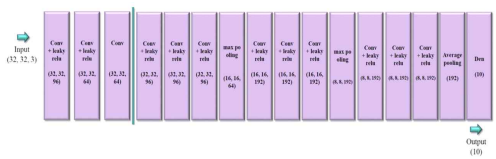
# 3. 실험 방법

## 3.1 실험 환경

본 연구에서는 하드웨어 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz와 NVIDIA Corporation GP102 [TITAN Xp]를, 소프트웨어 Anaconda python 3.8 환경에서 작업하였으며, tensorflow 2.2.0 버전으로 작업하였으며, matplotlib 3.3.1, scikit-learn 0.23.1, keras 2.3.1, pillow 7.2.0을 사용한다.

## 3.2 실험 모델

본 연구에서 사용한 실험 모델은 논문 [13]을 참고하여 만들어졌다. (그림 1)와 같이 학습을 위해 필요한 사용자 모델과 클라우드 모델로 구성되며, 세부 내용은 다음과 같다.



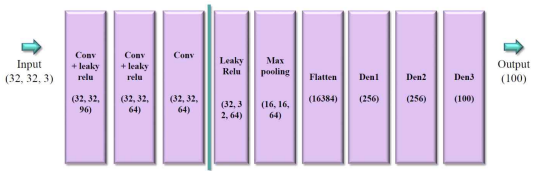
(그림 1) 실험 모델의 구조

### 3.2.1 사용자 모델

사용자 모델은 CIFAR-100 [20] 데이터 세트로 미리 학습하여 구성하였으며, 사용자 모델 전체 학습을 위한 구조는 (그림 2)과 같다.

- **CIFAR-100 데이터 세트** : CIFAR(Canadian Institute for Advanced Research) 데이터 세트는 각각 600개의 이미지를 포함하는 100개의 클래스가 있다. 각 클래스당 500개의 훈련 이미지와 100개의 검증 이미지가 있으며, 100개 클래스는 20개의 수퍼 클래스로 그룹화된다. 각 수퍼 클래스는 수중 포유류, 물고기, 꽃, 식기류, 과일 및 채로, 가정용 전자기기, 가정용 가구, 곤충, 큰 육식동물 등으로 구성된다.

이때 사용자 모델은 입력 값에 가까운 세 개의 레이어가 되며, 사용자의 입력값이 세 개의 레이어를 통과하여 클라우드 모델에 전송하게 된다.



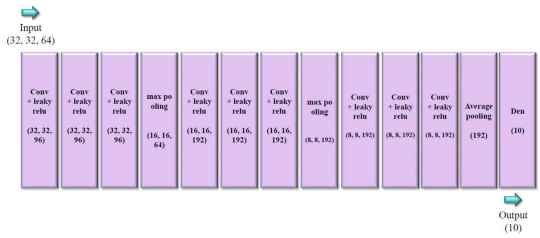

(그림 2) 사용자 모델 학습을 위한 구조

### 3.2.2 클라우드 모델

클라우드 모델은 SVHN [21]와 GTSRB [22] 데이터 세트를 이용하여 학습한다.

- **SVHN 데이터 세트** : SVHN(The Street View House Numbers) 데이터 세트는 구글 거리뷰 이미지에서 따온 0부터 9까지의 숫자 이미지로, (32, 32, 3)의 사이즈를 지니고 있다. 73,200개의 이미지를 학습 데이터로 사용하였으며, 26,000개의 이미지를 시험 데이터로 사용하였다.
- **GTSRB 데이터 세트** : GTSRB(German Traffic Sign Recognition Benchmark) 데이터 세트는 도로 표지판으로 구성된 이미지로, 40개 이상의 값으로 라벨링 되어 있다. 50,000개 이상의 이미지를 지니고 있으나, 이 중 10개 클래스를 뽑아 14,600개의 이미지를 학습 데이터로, 4,800개의 이미지를 시험 데이터로 사용하였다. 또한, 각각의 이미지 크기를 (32, 32, 3)로 조정하였다.

클라우드 모델의 구조는 (그림 3)와 같다.




(그림 3) 클라우드 모델의 구조

여기서 클라우드 모델은 사용자 모델에서 전송된 학습된 데이터를 받아 마저 학습을 시키게 된다. 이후 10가지 값으로 구분하여 결과를 내보낸다.

## 3.3 모델 공격

본 연구에서 사용한 공격 방식은, 학습 데이터 추출 공격 중 White-box 공격으로 pytorch로 작업된 [23]를 참고하여 tensorflow 2.2로 작업되었다. 공격은 사용자 모델과 클라우드 모델 사이의 값을 알고 있고, 사용자 모델의 구조와 매개변수를 모두 알고 있다고 가정한다.

## 3.4 결과 분석

본 연구는 MSE(Mean Square Error) 연산을 사용해 결과를 분석한다. MSE는 수식 (2)와 같이 정의된다. 수식에 나타난 $X$와 $Y$는 두 개의 다른 이미지를 나타내며, 두 이미지는 모두 $m \times n$ 행렬로 이루어져 있다. 이때, 각 행렬의 $i$행 $j$열의 값을 $X(i,j)$, $Y(i,j)$로 나타낸다고 가정한다.

$$MSE(X, Y) = \frac{1}{m \cdot n} \sum_{i,j=1,1}^{m,n} \| X(i,j) - Y(i,j) \|^2 \quad (2)$$

## 3.5 실험 과정

본 연구에 사용된 실험은 다음의 프로세스를 거쳐 실험 및 분석되었다.

- 데이터 전처리를 수행한다. 각 이미지 데이터를 (32, 32, 3)으로 만들어 사용자 모델에 넣을 수 있게 만든다.
- 사용자 모델을 구축하기 전에, 사용자 모델의 일부가 포함되어있는 모델 전체에 CIFAR-100 데이터 세트를 학습시킨다.
- 전 프로세스에서 학습한 모델에서 입력값에 가까운 세 개의 레이어를 추출하고 차분 프라이버시를 적용하여 사용자 모델로 구축한다. 차분 프라이버시는 프라이버시 정도에 따라 모델의 정확성 및 사용 가능성을 판단하기 위해 차등적으로 15단계 ($\epsilon$ = 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000)로 세분화한다
- 사용자 모델에 클라우드 모델을 붙이고 SVHN 데

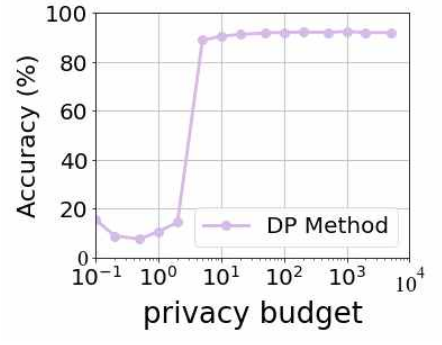
이터 세트를 혹은 GTSRB 데이터 세트를 학습시킨다. 이때 사용자 모델에 사용되는 가중치 값은 고정이며, 클라우드 모델의 값만 학습된다.
- 학습 정도를 확인하여 모델이 사용 가능한지 판별한다.
- 이후 사용자 모델 부분까지 학습된 데이터로부터 학습 데이터 추출 공격을 진행한다. 공격으로 복구된 이미지를 추출한다.
- 복구된 이미지와 원본 이미지를 정량적으로 비교한다. 비교하는 알고리즘에는 MSE를 사용한다.

# 4. 결과 및 분석

## 4.1 SVHN 실험

### 4.1.1 모델의 유용성

SVHN 데이터 세트를 모델에 적용했을 때의 시험 정확도는 다음과 같다. 각 정확도는 프라이버시 예산($\epsilon$)에 따라 확인하였으며 (그림 4)과 같이 나타난다. 프라이버시 예산이 2일 때, 시험 정확도는 14.493%로 높은 강도의 노이즈로 인해 학습이 이루어지지 않았다. 그러니 프라이버시 예산이 5일 때의 시험 정확도는 88.945%로 학습이 이루어지는 것이 드러난다.



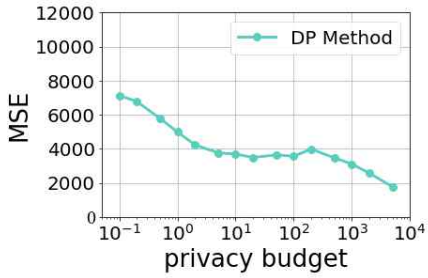

(그림 4) SVHN 데이터 학습 정확도

따라서, 프라이버시 예산이 5 이상일 때부터 모델은 SVHN 데이터를 학습할 수 있으며, 2 이하일 경우 모델에서 데이터를 학습하지 못하여, 사용하기 어려운 것으로 나타났다.

### 4.1.2 공격에 대한 방어성

사용자 모델을 통과하고 난 SVHN 이미지 데이터를 공격 모델인 학습 데이터 추출 공격을 수행할 경우, 다음의 (그림 5)와 같은 결과가 나타났다. 프라이버시 예산이 클수록 원본 이미지를 복구하기 쉬우며, 프라이버시 예산이 작을수록 노이즈 정도가 심해 원본 이미지를 복구하는 것은 불가능에 가깝다. 이에 대한 MSE 분석은 (그림 6)과 같다.


(그림 5) SVHN 학습 데이터 추출 공격 결과

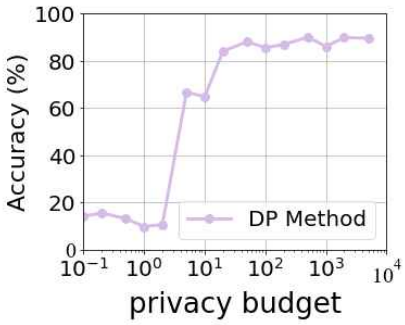



(그림 6) SVHN 데이터 세트에 대한 MSE 결과

## 4.2 GTSRB 실험

### 4.1.1 모델의 유용성

GTSRB 데이터 세트를 모델에 적용했을 때의 시험 정확도는 다음과 같다. 각 정확도는 프라이버시 예산($\epsilon$)에 따라 확인하였으며 (그림 7)과 같이 나타난다. 프라이버시 예산이 2일 때, 시험 정확도는 10.7%로 높은 강도의 노이즈로 인해 학습이 이루어지지 않았다. 그러니 프라이버시 예산이 5일 때의 시험 정확도는 66.768%로 학습이 이루어지는 것이 드러난다. 그리고 프라이버시 예산이 50일 때, 84.16%로 원활하게 학습되는 것으로 나타난다.



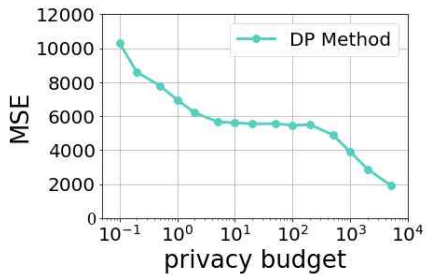

(그림 7) GTSRB 데이터 학습 정확도

따라서, 프라이버시 예산이 50 이상일 때부터 학습이 원활하게 이루어지며, 2 이하일 경우 모델에서 데이터를 학습하지 못하여, 사용하기 어려운 것으로 나타났다.

#### 4.1.2 공격에 대한 방어성

사용자 모델을 통과하고 난 GTSRB 이미지 데이터를 공격 모델인 학습 데이터 추출 공격을 수행할 경우, 다음의 (그림 8)과 같은 결과가 나타났다. 프라이버시 예산이 클수록 원본 이미지를 복구하기 쉬우며, 프라이버시 예산이 작을수록 노이즈 정도가 심해 원본 이미지를 복구하는 것은 불가능에 가깝다. 이에 대한 MSE 분석은 (그림 9)와 같다.


(그림 8) GTSRB 학습 데이터 추출 공격 결과




(그림 9) GTSRB 데이터 세트에 대한 MSE 결과

### 4.3 결과에 대한 논의

SVHN과 GTSRB 데이터 세트를 학습한 분할 학습 모델은 프라이버시 예산이 50 이상일 때부터 학습이 원활하게 이루어지는 것으로 나타났다. 또한, 학습 데이터 추출 공격을 진행했을 때, 프라이버시 예산이 200 이하일 때 공격에 예방되는 것으로 나타났다. 따라서 본 연구는 프라이버시 예산 값이 50에서 200일 경우에 제안하는 모델이 SVHN과 GTSRB 데이터 세트에 대해 효과적으로 개인정보보호와 정확성 둘 다 만족할 수 있는 것으로 도출한다.

## 5. 결 론

본 논문은 SVHN과 GTSRB를 프라이버시를 이용한 분할 학습에 적용하여 학습시킨 모델의 성능을 확인한다. 또한, 사용자 모델에 학습 데이터 추출 공격하고 MSE 연산을 사용해 공격된 이미지를 원본 이미지와 비교하였다. 본 연구에서는 SVHN과 GTSRB 데이터 세트에 대해 프라이버시 예산 값 5 이상에서 학습됨을 확인했으며, 50 이상부터 활용에 용이함을 확인하였다. 그리고 학습 데이터 추출 공격을 피하기 위해 프라이버시 예산 값 200 이하부터 활용에 용이함을 육안으로 감지하고, MSE 결과로 도출한 프라이버시 예산에 대해 분석하였다.

본 논문은 프라이버시 예산을 분할 학습 모델에 적용하여 실생활에 사용할 수 있는 프라이버시 예산 값을 도출하였으며, 본 연구를 바탕으로 인공지능 모델의 개인정보보호에 기여한다. 본 연구는 분할 학습을 이용하여 사용자의 프라이버시를 보호할 뿐만 아니라, 프라이버시 예산을 적용하여 학습 데이터 추출 공격을 예방하였다. 특히, 본 논문에서 제안한 프라이버시 예산을 적용한 분할 학습 모델은, 인공지능 모델의 학습에 영향을 미치지 않으며 사용자의 민감한 정보를 보호할 수 있는 적절한 프라이버시 예산 값을 정량화 한다. 본 연구는 향후 프라이버시 예산을 사용할 인공지능 모델에 유용하게 활용될 수 있을 것으로 기대한다.

# 참고문헌


[1] 송상훈 (Ed.). 대한민국 정책브리핑, “AI 챗봇 ‘이루다’ 관련 조사 결과 발표”, Retrieved June 19, 2021, from https://www.korea.kr/news/policyBriefingView.do?newsId=156449232, 2021.

[2] 삼정KPMG 경제연구원, “음성 AI 시장의 동향과 비즈니스 기회”, ISSUE MONITOR, 제 126호, 2020.

[3] Eunjung Jun, Hakbeom Kim, and Heungyoul Youm, “미국의 개인정보보호 법. 제도 동향”, Review of KIISC 22.1 pp.47-57, 2012.

[4] Herv´e Chabanne, Amaury de Wargny, Jonathan Milgram, Constance Morel, Emmanuel Prouff, “Privacy-Preserving Classification on Deep Neural Network.” IACR Cryptol. ePrint Arch, 2017.

[5] Mauro Barni,, Claudio Orlandi, Alessandro Piva, “A privacy-preserving protocol for neural-network-based computation.” Proceedings of the 8th workshop on Multimedia and security, 2006.

[6] Menghan Liu, Haotian Jiang, Jia Chen, Alaa Badokhon, Xuetao Wei, and Mingchun Huang, “A collaborative privacy-preserving deep learning system in distributed mobile environment”, IEEE International Conference on Computational Science and Computational Intelligence (CSCI), pp.192-197, 2016.

[7] Rigaki Maria, and Sebastian Garcia, “A survey of privacy attacks in machine learning” arXiv preprint arXiv:2007.07646, 2020.

[8] Jonghwan Ko, Taesik Na, Mohammad Faisal Amir, and Saibal Mukhopadhyay, “Edge-host partitioning of deep neural networks with feature space encoding for resource-constrained internet-of-things platforms”, 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, pp.1-6, 2018.

[9] Teerapittayanon Surat, Bradley Mcdanel, and Hsiangtsung Kung, “Distributed deep neural networks over the cloud, the edge and end devices”, 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), IEEE, pp.328-339, 2017.

[10] Shokri, Reza, and Vitaly Shmatikov, “Privacy-preserving deep learning”, Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp.1310-1321, 2016.

[11] Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, Shiho Moriai, “Privacy-preserving deep learning: Revisited and enhanced”, International Conference on Applications and Techniques in Information Security. Springer, Singapore, pp.100-110, 2017.

[12] Seyed Ali Osia, Ali Shahin Shamsabadi, Sina Sajadmanesh, Ali Taheri, and Kleomenis Katevas, “A hybrid deep learning architecture for privacy-preserving mobile analytics”, IEEE Internet of Things Journal 7(5), pp.4505-4518, 2020.

[13] Ji Wang, Jianguo Zhang, Weidong Bao, Xiaomin Zhu, Bokai Cao, and Philip S. Yu, “Not just privacy: Improving performance of private deep learning in mobile cloud” Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.2407-2416, 2018.

[14] Aaron Harlap, Deepak Narayanan, Amar Phanishayee, Vivek Seshadri, Nikhil Devanur, Greg Ganger, and Phil Gibbons, “Pipedream: Fast and efficient pipeline parallel dnn training” arXiv preprint arXiv:1806.03377, 2018.

[15] Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtarik, Ananda Theertha Suresh, Dave Bacon, “Federated Learning: Strategies for Improving Communication Efficiency.” NIPS Workshop on Private Multi-Party Machine Learning, 2016.

[16] Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, Ramesh Raskar. “Split learning for health: Distributed deep learning without sharing raw patient data.” ICLR AI for social good workshop, 2019.

[17] 한국인터넷진흥원, “글로벌 기업의 차등 프라이버시 기술 적용 오픈 소스 지원 현황”, 2020.

[18] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, “Calibrating noise to sensitivity in private data analysis” Theory of cryptography conference. Springer, Berlin, Heidelberg, pp.265-284, 2006.

[19] Jihyeon Ryu, Yifeng Zheng, Yansong Gao, Sharif Abuadbba, Junyaup Kim, Dongho Won, Surya Nepal, Hyoungshick Kim, and Cong Wang, “Can Differential Privacy Practically Protect Collaborative Deep Learning Inference for the Internet of Things?” arXiv preprint arXiv:2104.03813, 2021.

[20] Krizhevsky, Alex, and Geoffrey Hinton, “Learning multiple layers of features from tiny images”, 7, 2009.

[21] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng, “Reading digits in natural images with unsupervised feature learning”, 2011.

[22] Peyman Hosseinzadeh Kassani, Andrew Beng Jin Teoh, “A new sparse model for traffic sign classification using soft histogram of oriented gradients.” Applied Soft Computing, pp.231-246, 2017.

[23] Zecheng He, Tianwei Zhang, and Ruby B. Lee, “Model inversion attacks against collaborative inference” Proceedings of the 35th Annual Computer Security Applications Conference, pp.148-162, 2019.

## [저자소개]

유 지 현 (Jihyeon Ryu)
2018년 3월 ~ 현재 성균관대학교 소프트웨어학과 석박사 통합과정
2018년 2월 성균관대학교 수학과 학사
2018년 2월 성균관대학교 컴퓨터공학과 학사
email : jhryu@security.re.kr

원 동 호 (Dongho Won)
2018년 3월 ~ 현재 성균관대학교 소프트웨어학과 명예교수
2015년 ~ 2018년 성균관대학교 컴퓨터공학과 행단석좌 교수
1982년 ~ 2015년 성균관대학교 컴퓨터공학과 교수
2002년 ~ 2003년 한국정보보호학회장
1985년 ~ 1986년 일본 동경공업대학교 객원연구원
1976년 ~ 1988년 성균관대학교 전자공학 (학사, 석사, 박사)
email : dhwon@security.re.kr

이 영 숙 (Youngsook Lee)
2009년 3월 ~ 현재 호원대학교 IT소프트웨어보안학과 교수
2008년 8월 성균관대학교 컴퓨터공학 박사
2005년 2월 성균관대학교 석사
1987년 2월 성균관대학교 정보공학사
email : ysooklee@howon.ac.kr