사용자 개인정보보호를 위한 음성 데이터 분할 학습 모델 연구

장형범 [1], 유지현 [1]
[1] 광운대학교 컴퓨터정보공학부

hbrepark@kw.ac.kr, jhryu@kw.ac.kr

A Study of the Audio Data Split Learning Model to Protect User Privacy

Hyung-beom Jang[1], Jihyeon Ryu[1]
[1]Dept. of Computer Information and Communication Engineering, Kwangwoon University

요 약

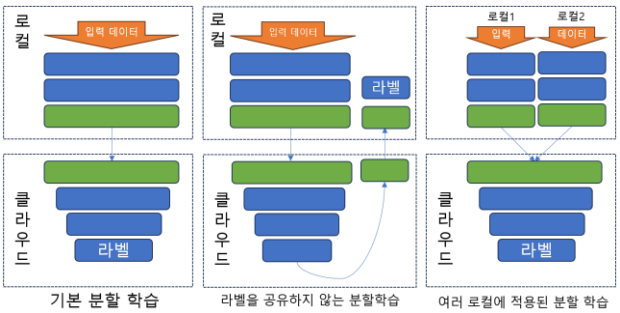
머신 러닝의 학습을 위한 데이터는 개인정보가 포함된 데이터인 경우가 존재한다. 특히 음성인식 모델을 학습시키기 위해서 사용자의 음성 데이터가 필요하며, 이는 개인의 민감한 정보가 포함될 수 있다. 인공지능 학습을 위해 수집한 음성 데이터에 대한 정보보호 침해 공격이 발생할 수 있고, 해당 데이터에 대한 보호 조치가 필요하다. 본 연구는 음성 데이터를 안전하게 관리하기 위해 분할학습을 이용한 음성 데이터 학습 모델을 제안한다.

1. 서론

머신 러닝에서 데이터의 보호는 점점 중요한 이슈로 부각되고 있다. 특히 음성 인식 분야에서는 민감한 정보가 포함된 음성 데이터세트를 사용하는 경우, 데이터 보호 문제가 더욱 심각한 고려 사항이 된다. 음성 데이터는 많은 응용 분야에서 중요한 역할을 한다. 음성 인식 기술은 음성 명령을 텍스트로 변환하거나 음성을 통해 사용자와 상호작용하며 다양한 서비스를 제공한다. 음성 데이터에는 사용자의 음성 패턴, 억양, 언어, 성별 등과 같은 개인정보가 포함될 수 있다. 이러한 정보는 사용자를 식별하거나 사용자의 개인 정보를 노출시킬 수 있다. 이를 방지하기 위해 연합 학습 (Federated Learning)과 분할 학습 (Split learning) 기술이 연구되고 있다[1, 2]. 본 연구는 음성 데이터를 분할 학습 모델에 적용하여 클라우드와 로컬 사이의 음성 데이터 공유를 위한 안전한 머신러닝 기술을 적용하는 방법론을 제안한다.

2. 분할 학습

분할 학습은 원본 데이터의 복호화를 방지하기 위한 모델로 로컬 모델과 클라우드 모델을 나누어 학습을 하는 방식이다. 분할 학습은 (그림 1)과 같이 로컬 모델이 학습한 데이터와 라벨을 공유 받는 방식, 라벨을 공유하지 않는 방식, 데이터를 공유하지 않고 서로 다른 데이터를 가진 여러 디바이스에서 분할 모델을 훈련하는 방식이 있다. 이때 로컬 모델은 학습이 완료된 상태로 더는 가중치가 업데이트 되지 않는다.



(그림 1) 분할학습 간단한 모델

분할 학습은 이미지 데이터를 학습하는 데 사용할 수 있다고 연구되어 왔다. [3, 4]. MNIST, SVHN 데이터 및 학습 모델에 차분 프라이버시를 적용하여 각 9

7%, 88%의 정확도가 확인되었다 [3]. 또한, 분할 학습에 차분 프라이버시를 적용했을 때, White box attack을 예방할 수 있는 차분 프라이버시에 대한 연구가 진행되었다 [4].

3. 모델 및 학습 데이터 세트

본 연구에서는 이미지 데이터를 학습하기 위해 분할 학습에 사용할 모델을 Deep speech 2 [5]로 채택한다. Deep speech 2는 ASR (Automatic Speech Recognition)으로 입력을 Spectrogram으로 받으며, 모델의 구조는 컨볼루션 신경망 (CNN), 순환 신경망 (RNN), 그리고 CTC 손실 함수를 사용한다.

로컬 모델에서의 학습 데이터셋은 Libri speech 데이터세트를 사용하고 클라우드 모델의 데이터셋은 Audio MNIST 데이터 세트를 사용한다.

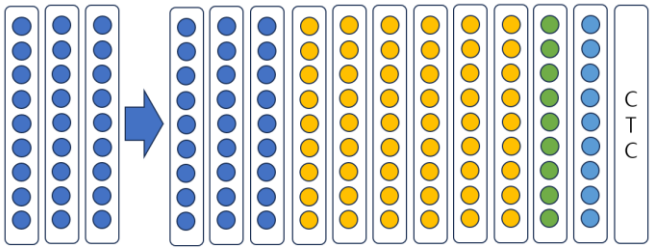
- Libri speech 데이터 세트: 대규모 음성 인식 작업을 위한 공개 음성 데이터세트 미국 영어로 발음된 대화문
- Audio MNIST 데이터 세트: MNIST 데이터 세트를 기반으로 만들어져 숫자를 읽는 음성 데이터 세트

3.1 음성 데이터 전처리

음성 데이터 전처리에는 대표적으로 두 가지 방식이 존재한다. Spectrogram 추출과 MFCC 특징 추출이 있다. Spectrogram은 시간에 따른 주파수의 변화를 시각화 하는 데 사용되는 그래픽 표현 방법이다. 시간을 x 축, 주파수를 y 축으로 나타내 오디오 신호의 주파수 내용이 시간에 따라 어떻게 변하는지 확인이 가능하다. MFCC는 Spectrogram에서 파생된 특징 벡터 중 하나로 음성 데이터를 벡터화 하는 특징이 있다. 이는 사람이 인지하기 편한 주파수만 자르는 Mel-Scale을 거친 전처리 방식이다 [6].

3.2 분할학습 모델 구성 및 학습

분할학습을 진행하기 전 로컬 모델의 학습을 완료시킨다. 로컬 모델의 학습이 완료되면 CNN 레이어에 Libri speech 데이터셋에 대한 가중치가 채워져 있는 상태이다. 이때 입력 값에 가까운 CNN 레이어의 일부분이 로컬 모델이 된다. 로컬 모델을 통과해 클라우드 모델로 데이터가 전달되는 방식을 표현한 (그림 2)와 같이, 클라우드 모델은 로컬 모델에서 전송된 학습 도중의 데이터를 가지고 모델을 업데이트한다. 이때 클라우드 서버는 원본 음성 데이터를 받지 않고도 모델의 가중치를 업데이트한다.



(그림 2) Deep speech 2 모델 분할 학습 제안 모델

4. 결론

본 연구는 분할 학습을 활용해 음성 데이터를 보호하는 모델을 개발하는 방법에 대해 논의했다. 본 연구는 분할 학습은 원본 데이터의 개인 정보 보호를 고려하면서도 효율적인 학습을 가능하게 하는 기술로, 이를 음성 데이터에 적용하는 방법에 대해 제안한다. 본 연구에서 제안한 모델은 로컬 모델이 클라우드 모델에 원본 데이터를 보내지 않고도 음성 데이터 인공지능 학습을 수행할 수 있다.

Acknowledgement

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00239728).

참고문헌

[1] Thapa, Chandra, Mahawaga Arachchige Pathum Chamikara, Seyit A. Camtepe, “Advancements of federated learning towards privacy preservation: from federated learning to split learning”, Federated Learning Systems: Towards Next-Generation AI, pp. 79-109, 2021.

[2] Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, Ramesh Raskar, “Split learning for health: Distributed deep learning without sharing raw patient data”, arXiv preprint arXiv:1812.00564, 2018.

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

[4] Jihyeon Ryu et al. “Can differential privacy practically protect collaborative deep learning inference for IoT?”, Wireless Networks, pp. 1 - 21, 2022.

[5] Dario Amodei et al. “Deep speech 2: End-to--end speech recognition in english and mandarin”, In International conference on machine learning, pp. 173 - 182, PMLR, 2016.

[6] Roy Rudolf Huizen, Florentina Tatrin Kurniati, "Feature extraction with mel scale separation method on noise audio recordings", Indonesian Journal of Electrical Engineering and Computer Science, vol. 24, no. 2, pp. 815 - 824, 2021.