

# 연합학습에서의 Multi-Exit 구조 적용에 관한 연구

신유진, 고정길

연세대학교 IT 융합공학과, 지능형반도체 IT 융합전공

yujin\_shin@yonsei.ac.kr, jeonggil.ko@yonsei.ac.kr

## A Study on the application of multi-exit learning for federated learning

Yu Jin Shin, JeongGil Ko

BK21 Graduate Program in Intelligent Semiconductor Technology, Yonsei University

### 요약

본 논문은 최근 다양한 성능을 가진 모바일 디바이스에서 개인정보를 유출하지 않고 모델의 파라미터만을 공유하여 학습을 진행하는 연합학습(federated learning)에 multi-exit 구조를 적용하였다. 이를 통해 모델의 개인화 성능 및 수렴 속도를 증대시키고, 클라이언트의 컴퓨팅 리소스 소모 및 통신 비용을 줄이고자 한다.

### I. 서론

현대의 디지털 환경에서 개인정보 보호는 중요한 문제로 대두되고 있으며, 특히 다양한 성능을 가진 모바일 디바이스가 보편화됨에 따라 분산된 데이터 처리 및 학습 방법론에 대한 필요성이 증가하고 있다. 이러한 배경에서 연합학습(Federated Learning)[1]은 개인 데이터를 로컬에서 유지하면서 모델의 파라미터만을 공유하여 중앙 서버에서 학습을 진행하는 방식으로, 개인정보 보호와 데이터 보안을 동시에 충족시키는 혁신적인 접근법으로 주목받고 있다.

연합학습은 다양한 클라이언트 장치에서 수집된 데이터의 분포가 균일하지 않고, 각 장치의 계산 능력과 네트워크 자원이 상이한 상황에서도 효과적인 학습을 가능하게 한다. 그러나 기존 연합학습 모델은 모든 클라이언트에 동일한 모델을 적용하여 클라이언트 마다 상이한 학습속도를 보이며, 이는 빠르게 연합학습 네트워크를 수렴시키는데에 어려움을 초래한다. 또한, 클라이언트의 컴퓨팅 리소스와 통신 비용은 연합학습의 실용성을 제한하는 주요 요인으로 작용하고 있다.

본 논문에서는 이를 해결하기 위한 방법으로 multi-exit 구조를 연합학습에 적용하는 방법을 제안하였다. Multi-exit 구조는 모델 내부에 여러 개의 출력 단계를 두어, 중간 단계에서도 예측 결과를 도출할 수 있도록 한다. 이러한 구조는 연합학습에 적용될 경우, 각 클라이언트가 자신의 리소스 상황에 맞게 최적의 출력 단계를 선택할 수 있어, 계산 부담을 줄이고 통신 비용을 최소화하는 동시에 모델의 개인화 성능을 향상시킬 수 있는 가능성을 제공한다.

본 논문에서는 연합학습에 multi-exit 구조를 도입하여 클라이언트의 다양한 요구를 반영하고, 모델의 개인화 성능 및 수렴 속도를 증대시키기 위한 접근을 제안한다. 이를 통해 클라이언트의 컴퓨팅 리소스 소모와 통신 비용을 줄이는 동시에, 개인화된 모델 성능을 극대화할 수 있는 가능성을 탐구하고자 한다.

### II. Multi-Exit 구조의 적용

본 논문에서는 이미지와 시계열 데이터 처리에 가장 보편적으로 사용되는 ResNet[2]과 LSTM[3]을 대상으로 multi-exit 구조를 재구성하였다. 이는 두 네트워크가 각각 컨볼루션 신경망과 순환 신경망이라는 상이한 구조를 가지면서도 multi-exit 구조의 적용 가능성을 확인하고자 함이다.

multi-exit 구조를 적용하기 위해서는 exit 위치 설정과 loss 구성이라는 두 가지 주요 요소가 필요하다. 첫 번째로, Multi-exit 구조를 두 모델에 적용한 방식은 다음과 같다. 이때, 두 모델 모두 총 3 층의 exit level을 같도록 설계하였다. ResNet은 총 5 개의 블록으로 구성되며, 이 중 첫 번째, 세 번째, 다섯 번째 블록의 출력을 사용하여 exit layer를 구성하였다. 각 exit layer는 average pooling 후 0.5의 dropout이 적용된 fully connected (fc) 레이어로 구성되었다. LSTM의 경우, 총 3 개의 LSTM 레이어로 구성되며, 각 레이어의 출력을 사용하여 exit layer를 구성하였다. 각 exit layer는 0.5의 dropout이 적용된 fully connected (fc) 레이어로 구성되었다.

다음으로 학습시에는 각 exit layer에서의 출력을 기반으로 cross entropy loss를 계산하고, 이를 모두 더한 값을 최종 loss로 삼아 학습을 진행한다. 이를 통해 각 exit layer의 학습이 동시에 이루어지도록 한다.

또한, [4]을 참고하여 상위 exit layer 의 출력을 하위 exit layer 에 대한 distillation loss 로 사용하였으며, distillation loss 와 cross entropy loss 의 비율은 하이퍼 파라미터 C 를 통해 결정된다. 이러한 loss 구성 방식은 각 exit layer 의 결과가 너무 달라지지 않도록 조정하여, 일관된 모델 성능을 유지하기 위함이다.

### III. 실험

본 논문에서는 multi-exit 구조를 적용하여 이미지 분류와 시계열 데이터 분류라는 두 가지 태스크에 대해 실험을 진행하였다. 이미지 분류에서는 ResNet 을 사용하였으며, 데이터셋으로는 Fashion MNIST 를 사용하였다. 시계열 데이터 분류에서는 Human Activity Recognition (HAR) 태스크를 수행하도록 하였으며, LSTM 을 사용하여 UNIMIB-HAR[5] 데이터셋에 대해 실험하였다. 실험은 총 24 개의 클라이언트가 참여하도록 구성되었으며, 두 데이터셋은 모두 dirichlet 분포의 alpha 값을 0.2 로 설정하여 non-iid 하게 분포되도록 구성하였다. 연합학습은 총 50 라운드 동안 진행하였으며, 각 라운드 별로 로컬 학습은 5 회씩 진행하였다.

Multi-exit(ME) 구조를 적용한 것과 그렇지 않은 결과를 실험을 통해 비교한다. 우선 정확도(Accuracy)를 비교하였다. 이때, 정확도는 각 클라이언트의 personalized 된 결과를 기준으로 하였다. 이미지 분류에서는 ME 적용시 81.38%, 미적용시 80.18%의 정확도를 달성하였다. HAR 태스크에서는 ME 적용시 60.48%, 미적용시 52.3%의 정확도를 달성하였다. ME 를 적용하였을 때 더 정확도가 높아지는 것을 확인할 수 있다.

	ME 적용	ME 미적용
이미지 태스크	81.38	80.18
HAR 태스크	60.48	52.3

Table 1. 각 태스크 별 Multi-exit(ME) 적용 여부에 따른 정확도(%)

다음으로 수렴 속도와 통신 비용을 비교하였다. 세 모델 모두 모델의 전체 파라미터를 서버로 전송하므로, 같은 정확도를 목표로 할 때 더 빠르게 수렴하는 모델이 더 적은 통신 비용을 필요로 한다. 이를 통해 통신 비용을 비교하면 Table 2 와 같다. 이미지 분류에서 타겟 정확도를 75%로 설정했을 때, ME 적용시 8 라운드, 미적용시 9 라운드가 소요되었다. HAR 태스크에서는 50%의 정확도를 달성하기 위해, ME 적용시 33 라운드, 미적용시 39 라운드가 소요되었다. ME 구조 적용을 통한 빠른 수렴으로 통신 비용의 이득을 가져올 수 있다는 것을 확인할 수 있었다.

	ME 적용	ME 미적용	통신 비용 절감
이미지 태스크	8	9	11.2%
HAR 태스크	33	39	15.4%

Table 2. 각 태스크 별 타겟 정확도 달성을 위해 필요한 라운드 수 및 통신비용 절감량(%)

마지막으로 Table 3 에서는 ME 를 적용하였을 때 각 exit 에서의 정확도를 비교하였다. Exit 별 정확도 차이가 크지 않으므로, 클라이언트 컴퓨팅 성능에 맞는 exit 레벨 선정으로 컴퓨팅 리소스를 절감할 수 있다.

Exit Level	이미지 태스크	HAR 태스크
Level 1	80.81	60.00
Level 2	81.78	59.76
Level 3	82.11	60.48

Table 3. 각 태스크의 Exit 레벨 별 정확도(%)

### IV. 결론

본 논문에서는 ResNet 과 LSTM 에 multi-exit 구조를 적용하여 연합학습의 성능을 향상시키는 방법을 제안하였다. 이를 통해 클라이언트의 컴퓨팅 리소스 소모 및 통신 비용을 줄이면서 모델의 개인화 성능과 수렴 속도를 증대시킬 수 있음을 보였다. 실험 결과, ME 를 적용하였을 때 더 높은 정확도와 빠른 수렴 속도를 달성하였다. 특히, 다양한 exit 포인트에서 일관된 성능을 유지하면서도 효율적인 자원 활용을 가능하게 하였다. 이러한 결과는 multi-exit 구조가 연합학습의 효율성과 효과성을 크게 향상시킬 수 있음을 시사한다.

### ACKNOWLEDGMENT

이 연구는 2023 년도 두뇌한국 21 사업(4 단계 BK21 사업)과 2024 년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0020535, 2024 년 산업혁신인재성장지원사업).

### 참고 문헌

- [1] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.
- [2] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [3] Sak, Hasim, Andrew W. Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." (2014).
- [4] Phuong, Mary, and Christoph H. Lampert. "Distillation-based training for multi-exit architectures." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
- [5] Micucci, Daniela, Marco Mobilio, and Paolo Napolitano. "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones." Applied Sciences 7.10 (2017): 1101.