

강화학습을 통한 분할학습에서의 효율적인 학습방안 연구

이두현, 이정우
서울대학교

{doolee13, junglee}@snu.ac.kr

A Study on an Efficient Learning Approach for Split Learning through Reinforcement Learning

Doo Hyun Lee, Jung Woo Lee*
Seoul National Univ.

요 약

기존 분산 환경에서의 신경망 학습 방식들은 개별 기기가 보유한 데이터가 독립적이지 않으며 같은 분포를 따르지 않기 때문에 경사 하강법이 편향된다는 한계를 갖고 있다. 이를 해결하기 위해 본 논문에서는 개별 기기가 보유한 데이터를 하나의 상태 공간으로 대응시킬 수 있는 강화학습 기반 구조를 제안하고 실험을 통해 극단적인 데이터 분포에서도 좋은 성능의 이미지 분류 결과를 확인할 수 있었다.

I. 서 론

연합학습(Federated Learning)의 고질적인 문제는 각각의 기기(client)에서 모아진 학습 데이터들이 독립적이지 않고 같은 분포를 따르지 않기(Non-IID) 때문에 인공 신경망을 학습시킬 때 최적해가 아닌 편향된 방향으로 경사 하강법이 진행된다는 것이다. 이에 대한 대안으로 개별 기기(client)에서 학습된 내용이 가우시안 분포의 혼합이라는 가정하에 서버단에서 이를 추론해 추가적인 학습을 진행하는 방법[1], 서버단에서 생성모델(generative model)을 추가적으로 학습시켜 전통적인 방식의 연합학습을 진행 후 생성모델이 만든 의사 데이터(pseudo data)를 통해 편향된 경사 하강법을 해결하는 방식[2]이 존재한다. 하지만 이러한 방식들의 한계는 분류하려는 데이터의 가지수가 많아질수록 정확도가 낮아진다는 점, 기존의 연합학습 진행 후 추가적인 훈련을 진행해야 한다는 한계를 갖는다.

연합학습과 다르게 개별 기기(client)와 서버의 인공 신경망 구조가 같지 않아도 되는 분할학습(split learning)에서도 편향된 기울기 문제는 지속되지만 서버와 기기의 신경망이 다른 구조를 가질 수 있다는 자유도 덕분에 강화학습 개념을 도입하여 데이터들이 독립적이지 않고 같은 분포를 따르지 않기(Non-IID) 때문에 발생하는 문제를 해결할 수 있다.

본 논문에서는 분할학습(split learning) 환경에서 각각의 기기가 가진 서로 다른 성질의 학습 데이터를 강화학습을 통해 효율적으로 훈련시키기 위해 본래의 목적에 맞게 상태(state), 행동(action), 보상(reward)에 대한 정의와 함께 제시된 환경에서 효율적으로 동작하기 위해 강화학습 알고리즘이 만족해야 하는 조건을 제시한다. 또한 위의 조건을 만족시키는 선행 연구된 알고리즘을 사용해 실제 분산환경에서 이미지 분류 작업을 효과적으로 수행할 수 있는지 검증한다.

II. 본론

모델과 강화학습 환경 설정 본 논문에서는 개별 기기(client)와 서버에 각각 행동 결정(actor), 행동 평가(critic) 신경망이 분할된 구조를 제안한다. 이러한 구조는 학습이 완료된 후 개별 기기가 서버와의 통신 없이 자신이 가진 결정 신경망만을 통한 추론 과정을 거쳐 결과를 알 수 있다는 장점이 있다. 그림 1 은 제안된 구조를 도식화 한 것이며 행동 결정 신경망은 기기, 행동 평가 신경망은 서버에 위치해 있다.

분할학습의 주요 과제로는 이미지 분류 작업이 있으며 이를 해결하기 위해 본 논문에서는 강화학습 환경에서의 상태집합 S 를 기기가 가진 이미지들의 전체집합, A 를 분류결과로 도출가능한 이미지의 종류(class) 집합에 대응시킬 것이며 $r \in [0,1]$ 로 주어지는 보상은 행동이 현재 상태에 맞는 결과를 택한 경우 1, 그렇지 못한 경우 0 으로 주어진다.

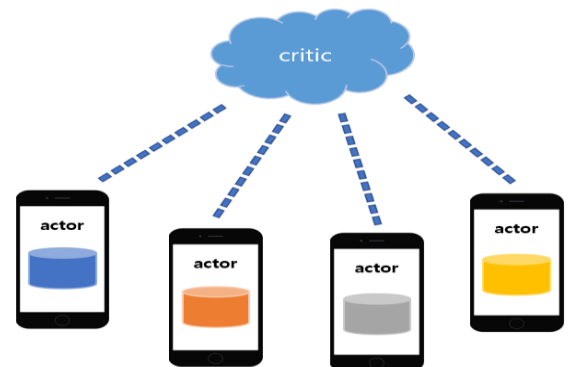


그림 1. 제안된 구조 도식화

분할학습을 위한 조건 제시된 환경에서 효율적인 학습을 위해서는 두가지 조건이 만족되어야 한다. 연합, 분할학습 설정에서 개별 기기는 제한된 연산성능과 저장공간을 가정하므로 강화학습에서 흔히 사용되는 재현 메모리 [3](replay buffer)를 사용하지 않는 방법이어야 하며 이

는 곧 탐색정책이 그대로 업데이트에 사용되는 방식(on-policy)을 의미한다. 또한 여러 기기들이 병렬적으로 자신의 훈련 데이터를 통해 업데이트하고 서버에서 이를 반영하기 위해서는 다중 프로세스를 지원하는 알고리즘이어야 한다. 위의 조건들을 만족시키는 알고리즘에는 A3C[4], PPO[5]가 있으며 본 논문에서는 더 높은 성능을 보이는 PPO 알고리즘을 사용했다.

$L^{clip}(\theta) = E_t[\min(r_t(\theta)A_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)]$ 선정된 알고리즘은 위와 같은 목적 함수를 통해 기존 정책에서 급격히 변하는 정책에 손해를 줘 급격한 정책 변화를 방지하고 A의 부호에 따라 정해지는 양방향의 하방(lower bound)을 두번째 항으로 설정해 안정적인 정책 변화를 유도하는 방식이다. 또한 다수의 정책 결정(actor) 신경망이 모은 데이터를 통해 정책 평가(critic) 신경망을 업데이트하는 환경이 분할학습에 적합하다.

실험조건 및 결과 독립적이지 않고 같은 분포를 따르지 않는(Non-IID) 데이터를 만들기 위해 Dirichlet 분포를 사용한다.

$$f(x_1, \dots, x_k; \alpha_1, \dots, \alpha_k) = \frac{1}{B(\alpha)} \prod x_i^{\alpha_i-1}$$

위와 같은 분포를 통해 나눠진 데이터는 아래의 그림 2와 같으며 사용된 변수는 0.6이다.

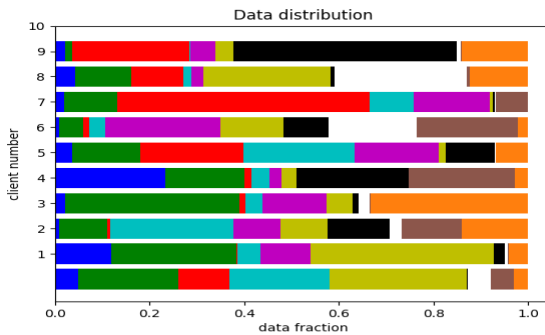


그림 2. MNIST 데이터 분포 그래프

행동 결정(actor), 행동 평가(critic) 신경망은 각각 2층의(2-layer) 64 차원 전결합층(fully-connected layer)으로 구성했으며 구체적인 학습 변수들은 표 1과 같다. 실험환경은 10 개의 기기(client)가 MNIST 이미지 데이터를 소유했음을 가정했고 학습 후 1000 번의 평가를 통해 성공률을 측정했으며 학습 스텝에 따른 분류 정확성은 표 2와 같이 나타난다.

항목	값
Learning_rate	3e-4
n_steps	2048
batch_size	64
n_epochs	10
gamma	0.99
gae_lambda	0.95
clip_range	0.2

표 1. 구체적인 학습 변수들 값

실험조건	FedAvg Non-IID	PPO IID	PPO Non-IID
정확도(%)	94.31	96.74	96.51

표 2. 기존 제시된 방식과의 분류 정확성 비교결과

공평한 비교를 위해 이미 존재하는 방법인 FedAvg 알고

리즘에서도 강화학습의 행동 결정, 평가 신경망과 같이 2 층의 전결합층을 사용하였으며 실험결과 극단적인 Non-IID 환경에서 2.2%p 성능증가를 확인할 수 있었다.

III. 결론

본논문에서는 강화학습 알고리즘을 통해 분할 학습 환경에서 병렬적인 탐색과 서버 중앙화된 훈련 구조를 제안했고 실험을 통해 극단적인 Non-IID 한 데이터 분포 환경에서도 신경망 훈련을 통한 이미지 분류가 효율적으로 이뤄졌음을 확인할 수 있었다. 또한 이는 개별 기기가 보유한 이미지 데이터의 집합이 강화학습 관점으로 봤을 때 하나의 큰 상태공간으로 대응될 수 있음을 의미하기도 한다.

ACKNOWLEDGMENT

This work is in part supported by Center for Applied

Research in Artificial Intelligence(CARAI,

UD190031RD(25%)) grant Funded by Defense Acquisition

Program Administration(DAPA), Agency for Defense

Development(ADD), INMAC, and BK21-plus.

참 고 문 헌

- [1] Luo, Mi, et al. "No fear of heterogeneity: Classifier calibration for federated learning with non-iid data." Advances in Neural Information Processing Systems 34. 2021
- [2] Zhang, Lin, et al. "Fine-tuning global model via data-free knowledge distillation for non-iid federated learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [3] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." nature 518.7540. 2015
- [4] Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." International conference on machine learning. PMLR, 2016.
- [5] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347. 2017