Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.6, Dec. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.6.1055

산업용 사물 인터넷을 위한 프라이버시 보존 연합학습 기반 심층 강화학습 모델*

한 채 림,[1†] 이 선 진,[2] 이 일 구[3‡]
[1,2,3]성신여자대학교 (학생, 대학원생, 교수)

Federated Deep Reinforcement Learning Based on Privacy Preserving for Industrial Internet of Things*

Chae-Rim Han,[1†] Sun-Jin Lee,[2] Il-Gu Lee[3‡]
[1,2,3]Sungshin Women's University (Student, Graduate student, Professor)

요 약

최근 사물 인터넷을 활용한 산업 현장에서 수집되는 빅데이터를 활용해 복잡한 문제들을 해결하기 위하여 심층 강화학습 기술을 적용한 다양한 연구들이 이루어지고 있다. 심층 강화학습은 강화 학습의 시행 착오 알고리즘과 보상의 누적값을 이용해 자체 데이터를 생성하여 학습하고 신경망 구조와 파라미터 결정을 빠르게 탐색한다. 그러나 종래 방법은 학습 데이터의 크기가 커질수록 메모리 사용량과 탐색 시간이 기하급수적으로 높아지며 정확도가 떨어진다. 본 연구에서는 메타 학습을 적용한 연합학습 기반의 심층 강화학습 모델을 활용하여 55.9%만큼 보안성을 개선함으로써 프라이버시 침해 문제를 해결하고, 종래 최적화 기반 메타 학습 모델 대비 5.5% 향상된 97.8%의 분류 정확도를 달성하면서 평균 28.9%의 지연시간을 단축하였다.

ABSTRACT

Recently, various studies using deep reinforcement learning (deep RL) technology have been conducted to solve complex problems using big data collected at industrial internet of things. Deep RL uses reinforcement learning's trial-and-error algorithms and cumulative compensation functions to generate and learn its own data and quickly explore neural network structures and parameter decisions. However, studies so far have shown that the larger the size of the learning data is, the higher are the memory usage and search time, and the lower is the accuracy. In this study, model-agnostic learning for efficient federated deep RL was utilized to solve privacy invasion by increasing robustness as 55.9% and achieve 97.8% accuracy, an improvement of 5.5% compared with the comparative optimization-based meta learning models, and to reduce the delay time by 28.9% on average.

Keywords: Data optimization, Deep neural networks, Deep reinforcement learning, Federated learning

Received(10. 17. 2023), Modified(11. 27. 2023), Accepted(11. 27. 2023)

* 본 논문은 2023년도 산업통상자원부의 재원으로 한국산업기술진흥원의 지원(P0008703, 2023년 산업혁신인재성장지원사업), 2023년도 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업(IITP-2022-RS-2022-00156310), 2023년도 여대학원생공학연구팀제 지원사업으로 과학기술정보통신부와 한국여성과학기술인육성재단의 지원을 받아 연구되었음.

† 주저자, 20200969@sungshin.ac.kr
‡ 교신저자, iglee@sungshin.ac.kr(Corresponding author)

I. 서 론

산업용 사물 인터넷(Industrial Internet of Things, IIoT)의 시장 규모가 기하급수적으로 증가함에 따라 기계학습(machine learning)은 4차 산업 성장을 위한 필수 기술이 되었다[1]. 강화학습(Deep Reinforcement Learning, Deep RL)은 시행착오를 통해 보상을 받으며 데이터에 내재된 규칙을 찾는 학습 방법이다[2]. 강화학습 에이전트(agent)는 자신 및 주변 환경의 상태 값(value)과 정책(policy)을 바탕으로 행동을 결정하고, 이에 따른 보상(reward)을 받는다. 정책은 최적의 행동을 결정하기 위하여 정하는 일련의 규칙을 의미하고, 보상은 에이전트 행동의 품질과 정책 준수 여부를 판단하는 중요한 지표 역할을 한다. 강화학습 에이전트는 보상 성형(reward shaping)을 통해 자신의 정책을 학습하여 보상을 최대화한다. 그러나, 산업용 사물 인터넷을 활용한 산업 현장에서 보상 성형 알고리즘은 시뮬레이션(simulation, sim) 환경에 최적화되어 있으므로 실제(real) 산업 환경에 적용했을 때 정상적인 학습이 불가능한 sim2real 문제가 발생한다[3]. 일부 에이전트만이 상태-행동(state-action) 값(q-value)에 대해 보상을 제공받고, 나머지 에이전트는 보상이 없는 상태(observed state)로 학습하기 때문에 후자의 경우 정책 결정(decision policy) 학습이 어렵다.

실제와 유사한 환경을 만들기 위하여 심층 신경망 모델(Deep Neural Network, DNN)의 보상에 노이즈(noise)를 추가하는 에이전트 강화학습에 관한 다양한 연구가 이루어지고 있다. 큐러닝(Q-learning) 기반 강화학습[4]은 버퍼에서 무작위로 선택된 상태-행동 값을 가지는 큐 행렬(q-table)로 미니배치(mini-batch)를 만들어 모델을 학습한다. 큐러닝 기반 강화학습은 목표로 하는 정책(target policy)을 학습하는 데에 최적화되어 있지만, 추정값이 최대값보다 큰 경우 심각한 편향이 발생한다(maximization bias).

학습 효율(Training efficiency)을 높이기 위하여 다중 에이전트 기반 강화학습(Cooperative reinforcement learning)[5] 또한 활발하게 연구되고 있지만, 프라이버시 침해라는 트레이드-오프(trade-off)가 발생한다. 효율적인 정책 학습을 위해서는 에이전트 간 정책 파라미터를 공유해야 하지만, 현존하는 연구에서는 다른 에이전트의 작업 정보에 접근할 수 없으므로 보상 성형 알고리즘에서 정책을 갱신할 수 없다[6].

종래 연구에서는 공통적으로 학습 데이터의 크기가 증가함에 따라 큐 행렬의 크기가 기하급수적으로 증가하는 메모리 과부하와 최종 행동을 결정하는 데 선택될 상태-행동 값(true q-value)의 정확한 수치를 알 수 없다는 본질적인 문제가 해결되지 못하였다. 따라서 본 연구에서는 심층 강화학습 모델에 마르코프 결정 과정(Markov Reward Process, MDP)[7] 기반의 연합-메타 학습을 적용하여 정확도와 보안성을 개선할 수 있는 프레임워크를 제안한다. 제안하는 방법의 주요 기여점은 다음과 같다.

- 메타 학습 기반의 다중 에이전트의 환경을 적용하여 sim2real 문제를 해결한다.
- 보상 신호(Reward signal)의 변동성이 큰 손실 값을 학습에서 배제하며 에이전트의 수에 따라 상태-행동 값을 분산 학습하여 산업용 사물 인터넷의 프라이버시 침해 문제를 해결하였다.
- 제안 모델은 종래 최적화 기반 메타 학습 모델 대비 28.9% 탐색 시간을 단축하고, 97.8%의 분류 정확도와 55.9%의 보안성을 개선했다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 메타 학습과 관련된 선행기술을 비교·분석하고, Ⅲ장에서는 메타 러닝을 적용한 연합학습 기반의 심층 강화학습 모델을 제안한다. Ⅳ장에서는 제안모델과 종래 최적화 기반 메타 학습들의 분류 정확도와 메모리 사용량, 지연시간, 보안성을 비교하여 평가하고, 마지막으로 Ⅴ장에서 결론을 맺는다.

II. 관련 연구

본 장에서는 메타 학습의 접근법 중 모델 파라미터를 최적화하는 접근법(optimization-based approach)과 연합학습(federated learning; FL)에 관한 대표 연구들을 분석한다.

2.1 메타 학습

메타 학습(Meta-learning)이란 현재의 문제를 해결하기 위하여 과거의 누적 결과의 근거들을 유추하여 딥 네트워크(deep network)의 파라미터(parameter)를 최적화하는 데 사용하는 방법론으로서 전이학습의 효율성을 증대시켜 예측 성능을 큰

폭으로 향상시킨다[8].

MAML은 경사하강법(gradient descent)을 이용하여 학습하는 모델의 파라미터를 최적화하는 방법이다[9]. 경사하강법은 오차함수의 기울기 보폭 크기에 따라 편향이 크게 좌우된다. 기울기의 간극이 너무 클 경우, 학습의 최저점에 도달하지 못하는 과적합(overfitting)이 발생하고, 간극이 너무 작으면 계산량이 기하급수적으로 증가하여 최저점을 찾지 못하는 과소적합(underfitting)이 발생한다.

MAML[10]에서는 학습 데이터에 대하여 미니배치를 생성하여 샘플링(sampling)하고, 각 샘플(sample)에 대한 오차함수 기울기를 계산[mini-batch gradient descent]하여 최적 파라미터를 찾을 때까지 반복적으로 값을 갱신한다. 이러한 방식으로 계산량을 줄이고, 오차함수를 최소화하는 지점으로 빠르게 수렴할 수 있다.

Daun 외 5명이 제안한 FOMAML[11]은 MAML의 성능은 유지하면서 계산 복잡도를 줄인 모델이다. 이 모델에서는 파라미터를 최적화하는 데 필요한 연산량과 비용을 줄이기 위하여 이계도함수(second derivative)를 생략하였다.

Nichol, Achiam과 Schulman이 제안한 Reptile[12]은 FOMAML과 유사하게 MAML의 성능을 유지하면서 계산 복잡도를 줄인 모델이다. 이 모델에서는 다중 경사하강법(multiple gradient descent)을 이용하여 각 미니배치마다 모델의 가중치를 다르게 설정하여 데이터를 학습한다. 손실값의 기울기의 합으로 최적의 파라미터를 선택한다.

Table.1은 MAML[10], FOMAML[11]과 Reptile[12]의 종래 심층 강화학습 모델에서 발생하는 문제를 해결하기 위한 기능을 비교·분석한 표이다. 종래 메타 학습 파라미터 최적화에 사용되는 방식들은 sim2real 문제를 해결하고, 학습에 필요한 연산량을 줄였다. 그러나, 상기 모델들은 공통적으로 파라미터를 추정하는 방식이기 때문에 정책 결정에 사용되는 상태-행동 값의 수치는 알 수 없다. 학습할 때마다 파라미터를 갱신해야 하며, 이는 모델의 메모리 사용량과 연산량의 과부하를 초래한다.

Table 1. Previous Research on the Optimization-based Approach of Meta Learning

Solve Problem	MAML[10]	FOMAML [11]	Reptile [12]
Sim2real	O	O	O
Train bias	X	△	△
Amount of calculation	X	O	O
True q-value	X	X	X

2.2 연합학습

연합학습은 탈중앙화된 데이터(decentralized data) 환경에서 다수의 로컬 기기 데이터를 중앙 서버로 전송하여 글로벌 모델(global model)을 학습한다. 로컬 모델의 업데이트 정보만 송·수신하기 때문에 네트워크 트래픽(network traffic)과 저장 비용을 줄일 수 있다[13]. 연합학습은 비독립·동일하게 분포되어 있지 않은(Non-Independent and Identically Distributed:Non-IID) 환경에서 데이터가 학습되기 때문에, 해당 환경에서 각 데이터를 공유하지 않으면서 모든 데이터셋을 글로벌하게 최적화하는 방법에 관한 연구가 활발하게 이루어지고 있다.

FedSGD[14]는 기존 로컬에서 사용되는 확률적 경사하강법(Stochastic Gradient Descent, SGD)[15]을 연합학습에 적용한 기초 모델이다. 해당 모델은 각 로컬 기기에서 계산된 가중치를 평균값으로 갱신하기 때문에 통신 비용이 증가하여 수렴 속도가 늦다.

FedAvg[16]는 데이터셋 증류(dataset distillation)을 이용하여 FedSGD의 효율성(efficiency)과 속도(bandwidth)을 높인 방법이다. 학습 기울기만을 연산했던 FedSGD와 달리 FedAvg에서는 로컬 데이터의 크기(batch size), 학습 횟수(epoch)와 모델 갱신에 참여하는 사용자(fraction)를 하이퍼파라미터(hyper-parameter)로 추가하여 더 독립적이고 효율적인 연합학습 환경을 조성하였다. 그러나, 해당 모델은 적대적인 공격과 신뢰할 수 없는 데이터에 대하여 견고성(robustness)을 유지하기 힘들다.

Cross-device FL[17]는 익명 사용자 개인의 이익을 최대화하는 방향으로 학습하는 모델이다. 연산량을 줄이기 위하여 사용자가 거짓 정보를 만들거나, 악의적인 정보를 생성할 수 있다. 더하여, Non-IID 데이터 환경에서의 불안정한 네트워크 환경에 원활한 통신을 보장하지 못한다.

III. 메타 학습을 적용한 연합학습 기반의 심층 강화학습 프레임워크

본 장에서는 메타 학습을 적용한 연합학습 기반의 심층 강화학습 프레임워크를 설명한다. Fig. 1와 Fig. 2는 각각 제안하는 모델의 프레임워크와 메타 학습의 순서도이다.

먼저, 제안 프레임워크는 데이터를 미니배치로 분류한다. 각 에이전트는 할당된 미니배치에 대해 개별 마르코프 결정 환경에서 상태와 행동을 학습하여 보상을 받는다. 본 연구에서 제안하는 마르코프 결정 환경 기반의 연합학습 환경에서는 신경망과 보상의 총합 G만 중앙 서버와 공유하기 때문에 데이터가 유출되더라도 데이터의 기밀 정보를 파악할 수 없다. 중앙 서버에서 G값을 갱신하기 때문에, 서버는 n개의 에이전트에 각각 신경망 학습 파라미터 τ을 공유한다. 이때, 보상(R)의 총합을 수식 (1)과 같이 정의하였다[18].

$$G = \sum_{k=1}^{n} \tau^{k-1} R_k \tag{1}$$

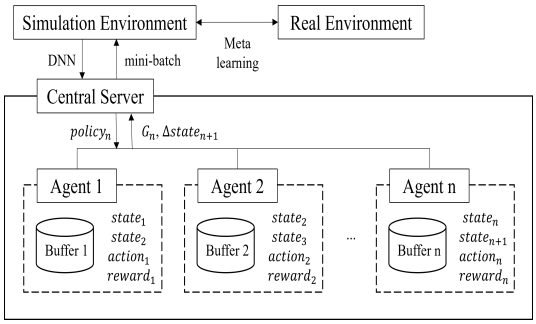

Fig. 1. Architecture of Meta Learning based Federated Reward Shaping Framework

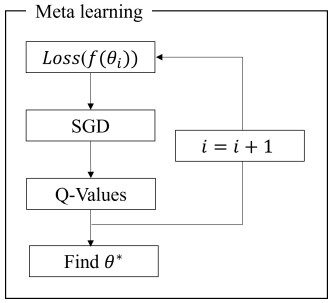

Fig. 2. Flow chart of Meta Learning

종래 심층 강화학습 알고리즘은 상태-행동 값의 수치가 아닌 추정값을 측정하기 때문에 해당 값의 정확도에 따라 확실성이 좌우되는 외삽 오류(extrapolation error)가 발생할 수 있다[19]. 제안 모델에서는 중앙서버가 각 에이전트의 상태 변동값($\triangle state_{n+1}$)과 G를 기반으로 보상을 계산하고, 이전 보상과 비교하여 변동성이 큰 보상 값은 제거함으로써 상태-행동 값이 오측(overestimation)되는 것을 방지하고 메모리 효율을 개선하였다.

학습 데이터가 복사 신경망을 학습하는 파라미터 τ를 가지고 학습 가중치 W를 계산한다고 가정할 때, 모델의 성능을 높이기 위해서는 모든 학습 데이터에 대해 파라미터 θ가 최적화되어야 한다. 각 복사 신경망의 W는 신경망 각각의 확률모델 p로부터 추출된다. 목적함수(Objective function)를 최소화하면서 확률적 경사하강 파라미터에 수렴하기 위한 θ와 W의 최적점은 Fig. 3과 같다. 그래프에서 x축은 τ, y축은 θ, z축은 W으로 설정하였다. 목적함수를 최소화하기 위한 θ와 W의 최적점은 2차 함수 그래프의 꼭짓점에 해당한다.

신경망 파라미터 θ는 미니 배치의 수 i만큼 복제되고, 각 미니배치에 있는 α개의 데이터는 각각의 복사 신경망 θ_i에서 학습된다. θ_i에서 계산된 목적함수 $f(\theta)$에 경사 하강을 사용하여 손실값의 합을 최소화하는 파라미터 θ^*를 찾는다. 해당 값을 바탕으로 상태-행동 값을 찾아 정책으로 선택한다.

랜덤으로 선택된 노이즈를 갖는 데이터셋 중 복사된 신경망들을 학습하기 위하여 모체 신경망에 사용되는 목적함수 $f(\theta^*)$의 θ^*는 수식 (2)와 같다. 각 미니배치에 최적화된 θ에 확률적 경사하강법이 적용된 손실값 L의 기울기 합이 최소일 때 전체 신경망의 성능은 최대화된다.

유클리드 거리(euclidean distance)[20] ∇_θ에

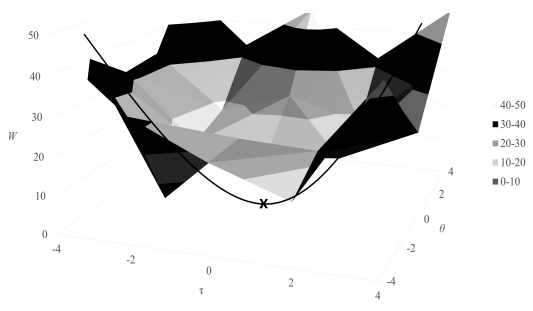

Fig. 3. Optimizer to minimize objective function

대하여 $f(\theta^*)$은 $f(\theta)$에서 각 $f(\theta)$의 손실값에 ∇_θ를 적용한 결과를 뺀 값으로 계산한다. 여기서 각 $f(\theta)$에 대한 L은 i번째 미니배치 데이터에 대하여 초기화된 θ를 기준으로 계산할 수 있다.

$$\theta^* = \min_\theta \sum_{i \sim p(i)} L_{p_i}(f(\theta_i^*)) \\ = \min_\theta \sum_{i \sim p(i)} L_{p_i}[f(\theta) - \nabla_\theta L_{p_i}(f(\theta))] \quad (2)$$

$f(\theta)$가 k단계 동안 확률적 경사하강법을 수행할 때, $f(\theta)$는 k단계의 정책 값 x과 실제 결과값 y간의 평균제곱오차(mean square error, MSE)를 의미한다[21]. 제안 모델의 $L_{p_i}(f(\theta))$는 수식 (3)과 같이 $f(\theta)$의 손실값으로 계산할 수 있다.

$$L_{p_i}(f(\theta)) = \sum_{x_k, y_k \sim p_i} f(\|x_k - y_k\|)^2 \quad (3)$$

제안 모델에서는 k가 0부터 9일 경우의 평균제곱오차 값을 비교하여 확률적 경사하강법을 적용할 최적의 k를 설정하였다. k에 따른 제안 모델의 평균제곱오차 값을 나타낸 실험 결과는 Fig. 4와 같다.

실험을 통하여 최적의 k값을 2로 설정하였고, 이후 단계에서도 과적합이 되지 않는 것을 확인하였다.

본 모델은 메타 학습 기반의 손실값을 최소화하는 최적화 파라미터의 상태-행동 값을 도출하여 사용하는 데이터셋마다 실제 환경에 적절히 조정(fine-tuning)할 수 있으며, 신경망 학습 파라미터만을 공유하여 프라이버시 문제를 개선하였다.

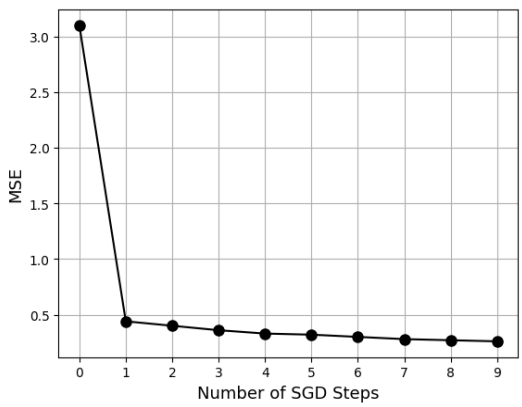

Fig. 4. Mean Square Error Value of the Proposed Model According to the Stochastic Gradient Descent's Steps

IV. 평가 및 분석

본 장에서는 제안한 메타 학습을 적용한 연합학습 기반의 심층 강화학습 모델과 종래 최적화 기반 메타 학습에 사용되는 MAML[10], FOMAML[11], Reptile[12]의 분류 정확도와 메모리 사용량, 지연시간, 보안성을 비교하여 성능을 평가하였다.

4.1 실험 환경

본 실험 환경은 Table 2와 같다. 평가는 Intel(R) CoreTM i7-1065G7 CPU @ 1.30GHz, RAM 16GB, SSD 477GB 환경에서 수행하였으며 시뮬레이터는 Python 3.12.0 버전으로 제작하였다.

CIFAR-10[22] 데이터셋을 활용하여 모델의 분류 정확도 성능에 대하여 샘플링과 배치 정규화(batch normalization)를 계산하였다. 메타 데이터셋 M을 클래스(C)로 나누고, K-shot, N-way 분류[23]를 통해 선택된 N개의 클래스에 대해 $K+1$개의 데이터를 선택하고 샘플링하였다. 전체 데이터셋에 대한 통계량은 배치 정규화 값으로 계산하여 평가를 진행하였다.

분산 학습에서 γ를 학습 비율(rate of learning), δ을 감산 요소(discount factor)라고 정의하였을 때, 실험 파라미터는 γ=0.05, δ=1.0으로 설정하였다. 메타 학습에 확률적 경사하강법(SGD)[15]을 사용하였고, 연합학습 기반의 분산 학습 환경에는 Adam Optimizer[24]을 사용하였다.

Table 2. Experiment Environment

	Specification and Version
CPU	Intel(R) CoreTM i7-1065G7 CPU @ 1.30GHz
RAM	16GB
SSD	477GB
Python	3.12.0

4.2 평가 결과 및 분석

4.2.1 분류 정확도

Table 3은 종래 메타 학습 기반의 심층 강화학습 모델과 제안 모델의 K-shot, N-way 분류 정확도를 비교한 표이다. 실험 결과에 따르면 Reptile[12]의 분류 정확도가 가장 낮고, MAML[10]과 FOMAML[11]은 비슷한 성능을 보인다. 종래 모델은 보상 성형과 이전 작업에서 얻은 정책을 바탕으로 정책을 선택하기 때문에 다음 상태에서 가능한 행동 중 가장 큰 상태-행동 값을 선택하여 학습에 반영한다. 제안 모델을 종래 모델과 비교하였을 때 최대 5.5%의 분류 정확도가 향상되었다.

분류 정확도를 높이기 위해서 성능이 가장 높았던 5-shot 5-way의 환경에서 미니 배치 9개에 대해 기울기를 추출하였다. Iteration이 증가할 때 기울기의 파라미터 조합에 따른 분류 정확도를 비교한 결과는 Fig. 5와 같다. 이때, iteration은 20,000까지 설정하였다.

최적의 k값이 2로 도출됨에 따라, 제안 모델의 메타 학습 기울기 g는 0번째 단계와 1번째 단계의 기울기를 더한 값에서 최대 97.8%로 가장 좋은 성능을 보였다.

이에 따라 제안 모델의 기울기 조합은 수식 (4)와 같이 정의할 수 있다. 이 수식에서 처럼 $g(f(\theta^*))$는 현재 학습 데이터 α에 대한 나머지 두 기울기의 합으로 구할 수 있다.

Table 3. K-shot, N-way Classification Accuracy of the Proposed model, MAML, FOMAML, Reptile

Model	1-shot 5-way	5-shot 5-way	1-shot 20-way	5-shot 20-way
MAML [10]	96.13 ±0.62 %	97.67 ±0.3%	94.2 ±0.21%	95.02 ±0.4%
FOMAML [11]	95.81 ±0.7%	96.43 ±0.3%	89.24 ±0.68%	94.89 ±0.12%
Reptile [12]	93.4 ±0.0%	95,64 ±0.1%	86.33 ±0.23%	94.21 ±0.3%
Proposed Model	**96.72 ±0.04 %**	**97.8 ±0.03 %**	**91.42 ±0.18 %**	**95.8 ±0.38 %**

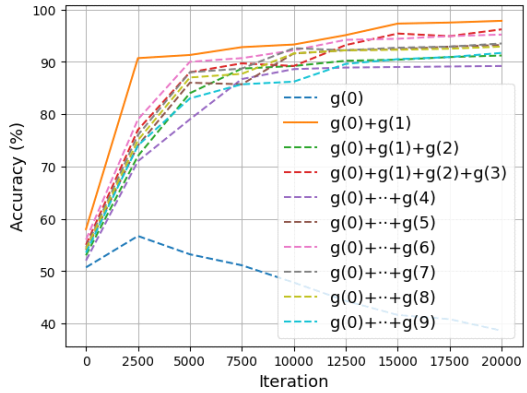

Fig. 5. Classification Accuracy of Different Gradient Combinations on 5-shot 5-way of the Proposed model

$$g(f(\theta)) = g(0) + g(1) \tag{4}$$

4.2.2 메모리 사용량

Fig. 6은 연합학습에 참여하는 노드를 25개까지 늘린 상황에서 제안 모델과 종래 모델의 메모리 사용량을 비교한 그래프이다.

종래 모델과 제안 모델 모두 노드 수가 많아질수록 메모리 사용량이 증가한다. FOMAML의 경우, 노드 1개의 환경에서 42.7MB, 노드가 25개인 환경에서 402.4MB의 메모리를 사용하며, 제안 모델은 각각 53.2MB, 546.8MB으로, 종래 모델에 비해 메모리 사용량이 약 1.3배 높다. 연합학습 기반의 제안 모델은 동시에 대용량의 메모리로 각 노드의 데이터를 병렬처리하므로 타 모델보다 더 많은 메모리

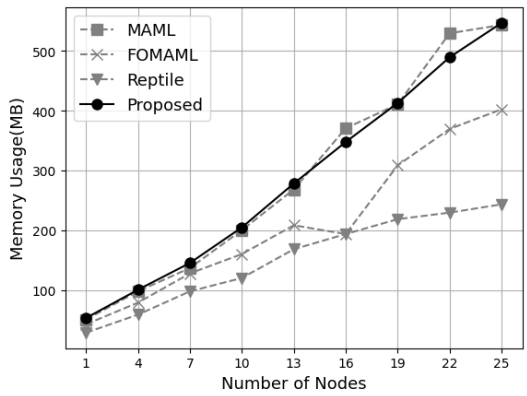

Fig. 6. Memory Usage of the Proposed model, MAML, FOMAML, Reptile

를 사용하지만, 더 높은 정확도를 보이는 것으로 분석된다.

4.2.3 지연시간

Fig. 7은 iteration을 20,000까지 늘리면서 제안 모델과 종래 모델의 지연시간을 비교한 그래프이다. 이때, 지연시간은 n번째 정책이 갱신되는 시간으로 설정하였다.

Iteration의 증가에 따라 종래 모델과 제안 모델의 지연시간은 공통적으로 감소하는 형태를 보인다. 본 실험에서 MAML, FOMAML, Reptile과 제안 모델의 지연시간은 각각 89.6s, 59.4s, 41.4s, 40.8s였다. 제안 모델은 종래 모델보다 평균 28.9% 적은 지연시간을 도출하면서, 비교적 안정적인 양상을 보인다.

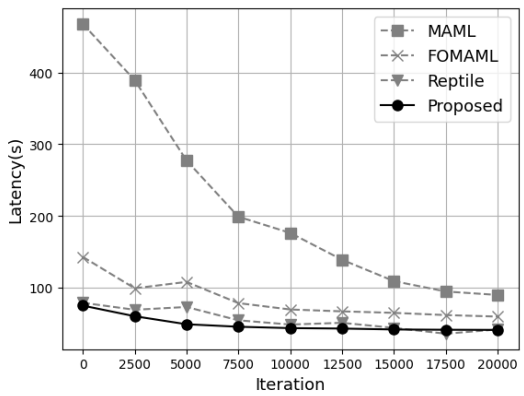

Fig. 7. Latency of the Proposed model, MAML, FOMAML, Reptile

4.2.4 보안성

제안 모델은 에이전트 별로 다른 신경망 학습 파라미터를 사용하여 학습함으로써 독립성을 유지하였다. 본 논문에서 보안성은 모델의 각 에이전트 학습 파라미터 간의 종속성 관계를 피어슨 상관 계수 r (Pearson Correlation Coefficient, PCC)[25]로 측정하였다. 각 에이전트의 신경망 학습 파라미터에 대한 큐행렬의 공분산을 M, M의 평균을 $\overline{M}$이라고 할 때, r은 수식 (5)로 정의한다[26]. 즉, 두 에이전트 간의 r은 에이전트 M의 공분산(covariance)을 각 표준편차를 곱한 값으로 나누어 계산할 수 있다.

$$r_{n,n+1} = \frac{\sum_{j=1}^{n}(M_j - \overline{M})(M_{j+1} - \overline{M})}{\sqrt{\sum_{j=1}^{n}(M_j - \overline{M})(M_{j+1} - \overline{M})^2}} \quad (5)$$

각 모델의 보안성을 비교한 결과는 Fig. 8과 같다. 실험에서 연합학습에 참여하는 노드는 4개로 설정하였고, 각 에이전트의 피어슨 상관 계수에 대한 독립성은 히트맵(heatmap)[27]으로 시각화하였다.

MAML, FOMAML과 Reptile의 피어슨 상관 계수는 각각 0.673, -0.634, -0.216으로, MAML은 에이전트 간 뚜렷한 양적 종속 관계를 가지며, FOMAML은 뚜렷한 음적 종속 관계, Reptile은 약한 음적 종속 관계를 나타냄을 알 수 있다. 즉, 산업용 사물 인터넷 환경에서 노드를 사용자라고 가정하였을 때, 사용자의 데이터가 유출된다면, 종래 모델에서는 개인의 중요 정보까지 유출될 가능성이 높음을 의미한다. 그러나, 제안 모델의 피어슨 상관 계수는 0.086으로, Liwen Qin 외 4명이 제안한 피어슨 상관 계수 관계 방법론에 따라 무시될 수 있는 선형 관계로 분류할 수 있다[28]. 제안 모델의 에이전트는 서로 독립적임을 알 수 있으며, 종래 모델 대비 평균 55.9%만큼의 보안성을 개선하였다.

V. 결 론

본 연구에서는 마르코프 결정 과정 기반의 연합-메타 학습을 통하여 산업용 사물 인터넷을 활용한 산업 현장에서 모델 학습의 분류 정확도와 보안성을 높였고, 탐색시간을 단축하였다. 평가를 통하여 제안 모델은 종래 모델 대비 5.5% 우수한 성능을 보였으며, 55.9%의 보안성이 향상됨을 확인하였다. 제안 모델은 사용자의 고유 가중치 값만 공유하기 때문에 산업용 사물 인터넷 환경에서 데이터 내의 정보를 파악하기 어렵고, 나아가 중요 정보를 예측하기도 힘들다. 평가 결과에서 정책은 실험 파라미터에 동조화(coupling)하는 경향을 보인다. 정책은 상태마다 파라미터와 가중치 간 거리의 값과 개수가 다르므로 고려할 요소들이 많고, 페널티(penalty)에 대한 변동이 크기 때문에, 보상에 직접적인 영향을 받는 상태-행동 값의 성능을 일관되게 유지할 필요성이 있다[29]. 후속 연구로 상태-행동 값의 과적합을 방지하면서 동시에 메모리 사용량을 개선할 방법에 관하여 연구할 계획이다.

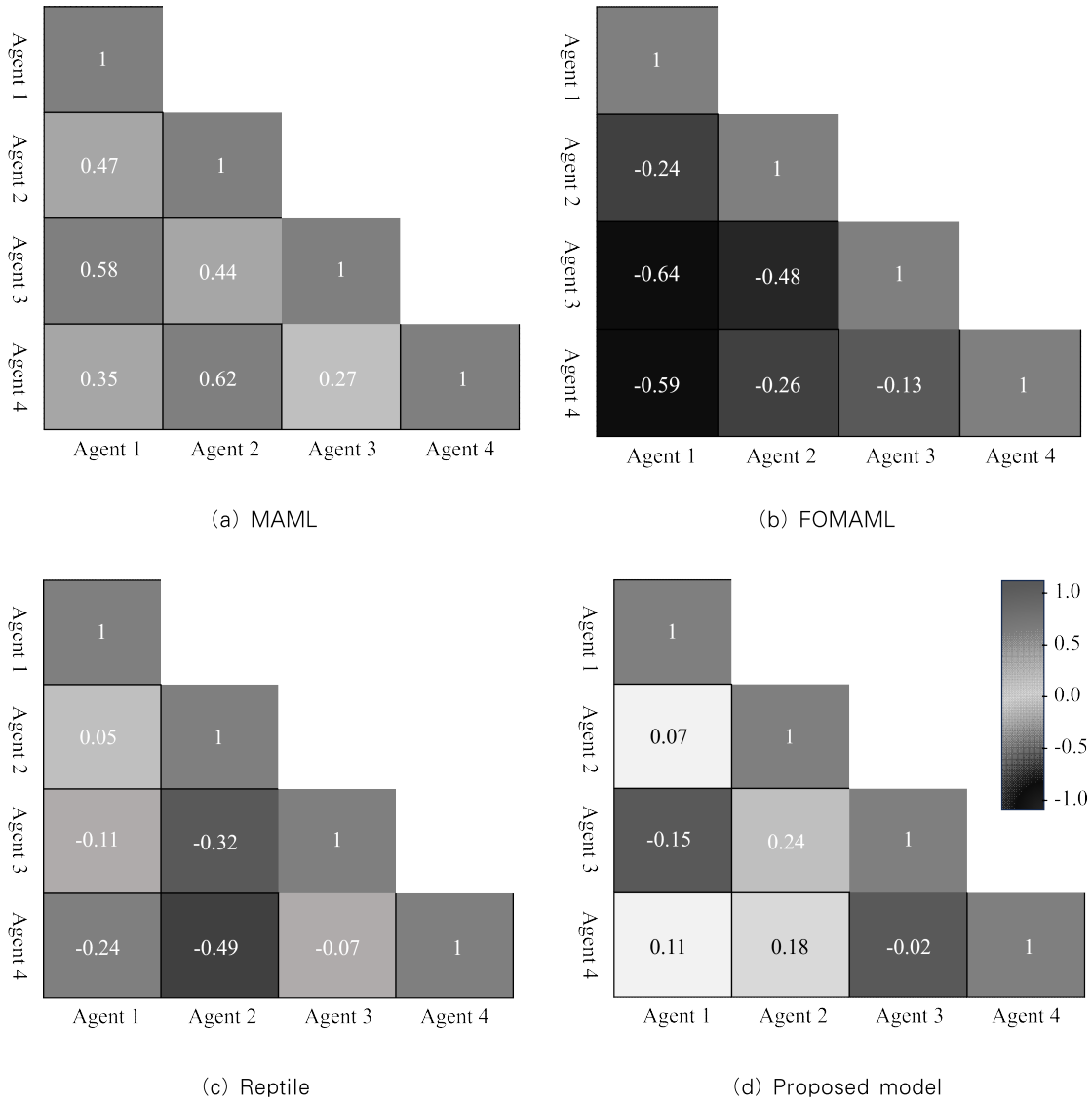

Fig. 8. Robustness of the Proposed model, MAML, FOMAML, Reptile through Independence between Agents

References

[1] Sun-jin Lee, Yu-rim Lee, So-Eun Jeon, and Il-Gu Lee, "Machine learning-based jamming attack classification and effective defense technique", Computers & Security(ELSEVIER), May. 2023.

[2] Badnava, Babak, et al. "A new potential-based reward shaping for reinforcement learning agent." 2023 IEEE 13th Annual Computing and Communication Workshop and Conference(CCWC). IEEE, Mar. 2023.

[3] Yang, Yulong, et al. "Reinforcement Learning with Reward Shaping and Hybrid Exploration in Sparse Reward Scenes." 2023 IEEE 6th International

Conference on Industrial Cyber-Physical Systems(ICPS). IEEE, May. 2023.

[4] Lu, Qin, and Georgios B. Giannakis. "Gaussian process temporal-difference learning with scalability and worst-case performance guarantees." ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE, Jun. 2021.

[5] Y. Hu, Y. Hua, W. Liu and J. Zhu, "Reward Shaping Based Federated Reinforcement Learning," in IEEE Access, vol. 9, pp. 67259-67267, Apr. 2021.

[6] Z. A. El Houda, D. Nabousli and G. Kaddoum, "Cost-efficient Federated Reinforcement Learning- Based Network Routing for Wireless Networks," 2022 IEEE Future Networks World Forum(FNWF), Montreal, QC, Canada, pp. 243-248, Oct. 2022.

[7] T. Gafni, M. Yemini and K. Cohen, "Restless Multi-Armed Bandits under Exogenous Global Markov Process," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), Singapore, Singapore. pp. 5218-5222, May. 2022.

[8] Abbas, Momin et al. "Sharp-MAML: Sharpness-Aware Model-Agnostic Meta Learning." International Conference on Machine Learning. Feb. 2022.

[9] Leibo, Joel Z., et al. "Scalable evaluation of multi-agent reinforcement learning with melting pot." International conference on machine learning. PMLR, Apr. 2021.

[10] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." International conference on machine learning. PMLR, Feb. 2017.

[11] Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. RL^2 : Fast Reinforcement Learning via Slow Reinforcement Learning, arXiv preprint arXiv:1611.02779, Nov. 2016.

[12] Nichol, Alex, Joshua Achiam, and John Schulman. "On first-order meta-learning algorithms." arXiv preprint arXiv:1803.02999. Oct. 2018.

[13] T. Yang, W. Zhang, Y. Bo, J. Sun and C. -X. Wang, "Dynamic Spectrum Sharing Based on Federated Learning and Multi-Agent Actor-Critic Reinforcement Learning," 2023 International Wireless Communications and Mobile Computing (IWCMC), Marrakesh, Morocco, pp. 947-952, Jun. 2023.

[14] W. Shin and J. Shin, "FedVar : Federated Learning Algorithm with Weight Variation in Clients," 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, pp. 1-4, Jul. 2022.

[15] J. K. Catapang, "Optimizing Speed and Accuracy Trade-off in Machine Learning Models via Stochastic Gradient Descent Approximation," 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI), Toronto, ON, Canada, pp. 124-128, Nov. 2022.

[16] McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, Feb. 2017.

[17] W. Yang, N. Wang, Z. Guan, L. Wu, X. Du and M. Guizani, "A Practical

Cross-Device Federated Learning Framework over 5G Networks," in IEEE Wireless Communications, vol. 29, no. 6, pp. 128-134, May. 2022.
[18] Richard S. Sutton and Andrew G.Barto, Reinforcement Learning, Second Edition : An Introduction, Cambridge : MIT Press, Mar. 2018.
[19] T. Deng, J. Huang and H. Li, "Extrapolation Method for Solving Fuzzy Volterra Integral Equations in Two Dimensions," 2022 7th International Conference on Computational Intelligence and Applications(ICCIA), Nanjing, China, pp. 22-26, Jun. 2022.
[20] G. Putro Dirgantoro, M. A. Soeleman and C. Supriyanto, "Smoothing Weight Distance to Solve Euclidean Distance Measurement Problems in K-Nearest Neighbor Algorithm," 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering(ICITISEE), Purwokerto, Indonesia, pp. 294-298, Nov. 2021.
[21] K. Beshara-Flynn and K. Adhikari, "Effects of Signal and Array Parameters on MSE and CRB in DOA Estimation," 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, NY, USA, pp. 0373-0379, Oct. 2022.
[22] Krizhevsky, Alex, and Geoff Hinton. "Convolutional deep belief networks on cifar-10." Unpublished manuscript 40.7 (2010), pp. 1-9, 2010.
[23] R. Ning, C. L. Philip Chen and T. Zhang, "Cross-Subject EEG Emotion Recognition Using Domain Adaptive Few-Shot Learning Networks," 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, pp. 1468-1472, Dec. 2021.
[24] https://github.com/juntang-zhuang/Adabelief-Optimizer. Oct. 2023.
[25] "Correlation Coefficient: Simple Definition, Formula, Easy Steps". Statistics How To. Oct. 2023.
[26] Weisstein, Eric W. "Statistical Correlation". mathworld.wolfram.com. Retrieved 22 Aug. 2020.
[27] Github, https://github.com/aleju/imgaug. Oct. 2023.
[28] L. Qin, W. Huang, Z. Guo, Y. Zhou and B. Zhang, "Topology Identification Method of Low-voltage Distribution Network Based on Improved Pearson Correlation Coefficient Method," 2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE), Chengdu, China, pp. 1-6. Dec. 2021.
[29] Li, Jiangeng, et al. "Generative Adversarial Imitation Learning from Human Behavior with Reward Shaping." 2022 34th Chinese Control and Decision Conference(CCDC). IEEE, Aug. 2022.

〈저 자 소 개〉

한 채 림 (Chae-Rim Han) 학생회원
2020년 3월~현재: 성신여자대학교 융합보안공학과 학사
2022년 3월~현재: 성신여자대학교 CSE LAB 연구원
〈관심분야〉 강화학습, 머신러닝, 정보보호

이 선 진 (Sun-Jin Lee) 학생회원
2021년 8월: 성신여자대학교 융합보안공학 학사
2023년 2월: 성신여자대학교 미래융합기술공학 석사
2023년 3월~현재: 성신여자대학교 미래융합기술공학 박사
〈관심분야〉 무선 네트워크, 네트워크 보안, 융합보안

이 일 구 (Il-Gu Lee) 종신회원
2003년 2월: 서강대학교 전자공학 학사
2005년 2월: KAIST 정보통신 석사
2016년 2월: KAIST 전산학부 박사
2005년 2월~2017년 2월: 한국전자통신연구원 선임연구원
2017년 3월~현재: 성신여자대학교 융합보안공학과/미래융합기술공학과 부교수
〈관심분야〉 융합보안, 정보보호, 정보통신