

부분 작업 오프로딩을 위한 심층 강화학습 알고리즘 연구

Study on Deep Reinforcement Learning for Partial Task Offloading in Edge Cloud Network

석 영 준(YeongjunSeok), 임 현교(Hyun-kyo Lim), 올라이산(IhsanUllah), 한 연 희(Youn-Hee Han)

Advanced Technology Research Center
Korea University of Technology and Education
{dsb04163, glenn89, ihsan, yhhan}@koreatech.ac.kr

요 약

IoT 및 5G네트워크 서비스가 발전하며 자원이 제한된 로컬 장치에서 요구되는 에너지 및 지연시간을 만족하기 위해 부분 작업 오프로딩을 사용할 수 있다. 하지만 부분 작업 오프로딩에서 오프로딩 비율을 결정은 조합 최적화 문제와 같아 해결하기 어렵다. 본 논문에서는 이러한 조합 최적화 문제 해결을 위해 에지 클라우드 컴퓨팅 체계에서 작업 오프로딩 방식에 계층적 심층 강화학습(Hierarchical Deep Reinforcement Learning)을 적용하고 성능을 확인한다. 또한 계층적 심층 강화학습 기반 작업 분배 최적화 문제에 대한 Markov Decision Process (MDP) 및 목적함수를 정의한다. 계층적 심층 강화학습에서의 각 계층은 서로 다른 알고리즘을 활용하며 오프로딩 시뮬레이션 결과는 제안하는 방법이 다른 자원 소모와 지연 시간에서 다른 방법보다 우수하다는 것을 입증한다.

키워드: IoT, 5G 네트워크, 부분 작업 오프로딩, 자원 소모, 지연 시간, 에지 클라우드 컴퓨팅 체계, 계층적 심층 강화학습, 작업 분배 최적화

Abstract

"As IoT and 5G network services continue to evolve, partial task offloading is becoming an important solution to meet energy and delay requirements on resource-constrained local devices. However, determining the offloading ratio in partial task offloading is a combinatorial optimization problem, which is difficult to solve. In this paper, we propose using hierarchical deep reinforcement learning in edge cloud computing systems to solve this combinatorial optimization problem for task offloading. We also define a Markov Decision Process (MDP) and objective function for the hierarchical deep reinforcement learning-based task allocation optimization problem. Each layer in the hierarchical deep reinforcement learning encourages the use of different algorithms, and simulation results show that the proposed method is superior in terms of resource consumption and delay time compared to other methods."

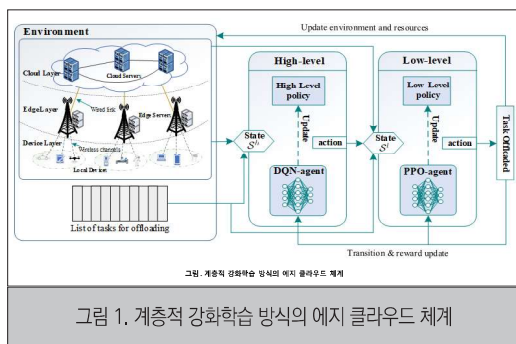
Key words: IoT, 5G Network, Partial Task Off-Loading, Resource Consumption, Delay Time, Edge Cloud Computing Systems, Hierarchical Deep Reinforcement Learning, Task Allocation Optimization

1. 서론

IoT 및 5G 무선 네트워크는 증강/가상 현실과 같은 새로운 애플리케이션의 사용을 가능하게 했다. 하지만, 자원이 제한된 로컬 장치가 지연에 민감하고 계산 집약적인 애플리케이션을 지원하기는 어렵다[1][2]. 따라서 에지 클라우드 컴퓨팅 체계에 작업을 오프로딩하여 로컬 장치에서 이러한 애플리케이션을 지원하도록 할 수 있다[3].

계층적 심층 강화학습(Hierarchical Deep Reinforcement Learning)은 복잡한 문제를 상위 레벨과 하위 레벨로 분할하여 정복하는 접근법을 제안함으로써 강화학습을 확장한다[4][5]. 이 접근법에서, 문제의 복잡성은 여러 개의 작은 문제로 분할되며 이러한 문제들은 일반적으로 해결하기가 더 쉽고, 그 해결책들은 다른 문제들을 해결하기 위해 재사용될 수 있다.

부분 작업(Partial Task Off-loading) 오프로딩 비율에 따라 지연시간 및 에너지 소모가 다양해진다. 따라서 최적의 오프로딩 비율을 선택해야 하는 조합 최적화 문제가 발생한다. 본 논문에서는 위 문제를 해결하기 위해 에지 클라우드 컴퓨팅 체계에서 계층적 심층 강화학습 기반 부분 작업 오프로딩을 제안하고, 다양한 오프로딩 방법과 비교하여 에너지 소모 최소화 및 지연시간 만족 측면에서 우수하다는 것을 증명한다.



2. 시스템 모델 및 문제 정의

본 장에서는 에지 클라우드 컴퓨팅 체계의 시스템 모델 및 문제 정의를 설명한다. 그림 1은 에지 클라

우드 체계에서 작업 오프로딩을 위해 제안된 계층적 심층 강화학습 방식을 보여준다. 시스템 모델은 작업 모델, 에너지 모델, 지연시간 모델로 구성된다.

2.1. 작업 모델(Task Model)

제안하는 모델에서 Base Station은 M 개 존재하며 각 Base Station에 N 개의 장치가 있다. 각 장치에는 처리할 작업이 하나 존재하며 작업은 데이터의 사이즈 σ (MB) 및 지연시간 λ (s)으로 구성되어 있다. 부분 작업 오프로딩을 통한 분할 비율은 다음과 같이 표시된다. $V = \{x, y, z\}$. 순서대로 로컬, 에지, 클라우드 별 작업의 오프로딩 비율을 나타낸다. 오프로딩 비율의 총합은 1로 제한된다.

2.2. 에너지 모델(Energy Model)

작업 당 소요되는 에너지 E 는 전송 에너지 E^{tr} 와 처리 에너지 E^{ex} 의 합이며 정규화 되어 0에서 1사이의 값을 가진다. 처리 에너지 E^{ex} 는 다음과 같이 계산된다. $E^{ex} = p^{ex} \sigma V$. p^{ex} (J/bit)는 계층의 컴퓨팅 에너지를 나타내며 V 는 계층의 작업 비율을 나타낸다. 전송 에너지 E^{tr} 는 다음과 같이 계산된다. $E^{tr} = P^{tr} (\sigma V + \alpha)$. P^{tr} (J/bit)는 전송 컴퓨팅 에너지를 나타내며 α 는 처리 후 반환되는 데이터의 사이즈이다.

2.3. 지연시간 모델(Delay Model)

지연 시간 T 는 전송 지연 시간 T^{tr} 와 처리 지연 시간 T^{ex} 의 합이다. 작업 당 전송 지연시간 T^{tr} 은 다음과 같이 계산된다. $T^{tr} = \sigma V / \mu$. μ (Gbps)는 오프로딩 계층의 전송률이다. 작업 당 계층에서 처리 지연시간 T^{ex} 은 다음과 같이 계산된다. $T^{ex} = \sigma V / F$. F (CPU cycle/s)는 계층의 컴퓨팅 사이클을 나타낸다. 지연시간은 각 계층의 지연시간 중 가장 큰 값을 사용한다.

2.4. 문제 정의

작업을 분할하여 오프로딩 한다면 분할 비율 $V = \{x, y, z\}$ 에 따라 작업 처리에 소모되는 에너

지 및 지연시간이 다양해진다. 모든 Base Station M 의 모든 작업 N 을 효율적으로 오프로딩하여 소모되는 에너지를 최소화하고 지연시간을 만족하기 위해서 본 논문에서 해결하고자 하는 문제를 다음과 같이 정의한다.

3. 계층적 심층 강화 학습

본 장에서는 시스템 모델에서 작업 오프로딩을 위한 계층적 심층 강화학습의 상위 레벨 MDP(Markov Decision Process)와 하위 레벨 MDP를 설명한다.

3.1. 상위 레벨 MDP(High-Level MDP)

상위 레벨 에이전트는 상태로 네트워크 정보를 입력받아 행동으로 작업의 분할 오프로딩 위치를 출력한다. 행동은 [로컬, 에지, 클라우드] 같은 형태를 가진다. 예시로 로컬, 에지, 클라우드에 작업을 분할해서 할당한다면 행동의 형태는 [1, 1, 1] 같으며, 로컬 및 에지에만 작업을 분할해서 할당한다면 행동의 형태는 [1, 1, 0]과 같다. 이러한 상위 레벨의 행동 형태는 하위 레벨의 상태로 전송된다. 보상은 하위 레벨 에이전트의 행동 후 계산된다.

3.2. 하위 레벨 MDP(Low-Level MDP)

하위 레벨 에이전트는 상태로 네트워크 정보 및 상위 레벨의 행동을 입력받아 행동으로 오프로딩 비율을 출력한다. 행동의 형태는 상위 레벨과 같으며 예

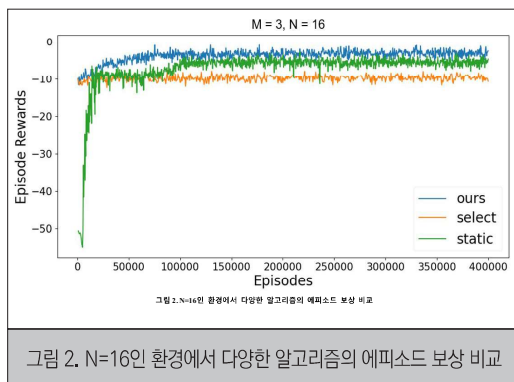


그림 2. $N=16$ 인 환경에서 다양한 알고리즘의 에피소드 보상 비교

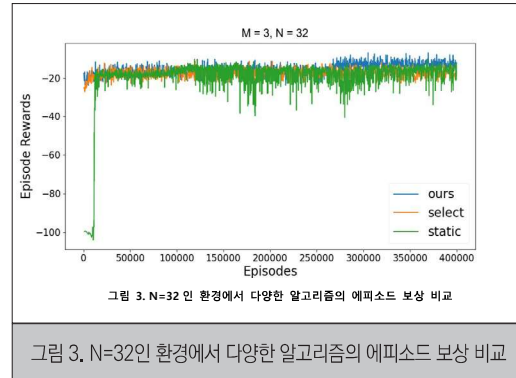


그림 3. $N=32$ 인 환경에서 다양한 알고리즘의 에피소드 보상 비교

시는 [0.4, 0.4, 0.2] 같다. 보상은 에너지 모델을 통해 계산된 작업의 컴퓨팅 에너지와 지연시간 모델을 통해 계산된 작업의 지연시간을 통해 계산된다. 리워드 R 은 $T < \lambda$ 라면 $1 - E$, $T > \lambda$ 라면 $-E$ 이다. 계산된 보상은 상위 레벨에도 전송된다.

4. 실험 및 결과

비교 실험에서 계층적 심층 강화학습 알고리즘은 상위 레벨은 DQN, 하위 레벨은 PPO를 사용했다. static은 심층 강화학습 PPO만을 사용했고 select는 static과 알고리즘은 같지만 작업을 분할하여 오프로딩 하지 않는다. 결과는 3회 학습의 평균을 사용했으며 [그림 2]와 같다. 환경은 $M=3$ 이며 $N=16$, $N=32$ 으로 2가지 경우를 테스트했다. X축은 에피소드를 나타내며, Y축은 학습 후 테스트에서 받은 보상을 나타낸다. 제시한 기법이 $T < \lambda$ 를 만족하며 E 는 최소화하는 것을 알 수 있다.

5. 결론

본 연구에서는 에지 클라우드 네트워크에서 작업의 지연시간은 만족하며 에너지 소모는 최소화하는 계층적 심층 강화학습 방법을 제시하고 효과를 검증했다. 본 연구는 현재 작업의 오프로딩 순서는 선택하지 않는다. 향후 연구에서 오프로딩 순서 또한 강화학습을 통해 결정하는 것으로 더욱 효율적으로 오프로딩 할 수 있도록 기대한다.

참고 문헌

- [1] A. Singh, S. C. Satapathy, A. Roy, and A. Gutub, "AI-Based Mobile Edge Computing for IoT: Applications, Challenges, and Future Scope," *Arab. J. Sci. Eng.*, pp. 1-31, 2022.
- [2] A. S. Mohammed, K. Venkatachalam, S. Hub-lovský, P. Trojovský, and P. Prabu, "Smart Edge Computing for 5 g/6 g Satellite IOT for Reducing Inter Transmission Delay," *Mob. Netw. Appl.*, pp. 1-10, 2022.
- [3] B. Dai, J. Niu, T. Ren, and M. Atiquzzaman, "Towards Mobility- Aware Computation Offloading and Resource Allocation in End-Edge-Cloud Orchestrated Computing," *IEEE Internet Things J.*, 2022.
- [4] T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum, "Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation," *Advances in neural information processing systems*, vol. 29, 2016.
- [5] M. M. Botvinick, "Hierarchical Reinforcement Learning and Decision Making," *Current opinion in neurobiology*, vol. 22, no. 6, pp. 956-962, 2012.