https://doi.org/10.7840/kics.2022.47.10.1565

심층강화학습 기반 차량 대 차량 자원 할당 기법

문 지 훈◆, 심 병 효°

Deep Reinforcement Learning-Based Vehicle-to-Vehicle Resource Allocation

Jihoon Moon◆, Byonghyo Shim°

요 약

본 논문에서는 차량 간 통신에서 다양한 요구 조건을 만족하기 위해 심층강화학습을 사용하는 방안을 제시한다. 심층강화학습 중 actor-critic 알고리즘을 활용해 주변 채널 상황을 입력으로 받아 자원 할당을 출력해 차량 별 요구 조건을 만족하면서 네트워크 데이터 전송률 합을 최대화한다.

Key words : Vehicular communications, deep reinforcement learning, resource allocation, scheduling, mmWave/THz communications

ABSTRACT

In this paper, we propose a deep reinforcement learning (DRL)-based method to satisfy diverse requirements in vehicular communications. Using the channel as the input, the actor-critic algorithm outputs the resource allocation maximizing the network sum rate while ensuring the requirements.

Ⅰ. 서 론

지능형 운송 시스템 (ITS)의 급격한 발전과 함께, 많은 차량 간 통신 서비스는 우리의 일상 생활에 큰 변화를 가져올 것이다. 현재, 차량 군집 주행 및 원격 운전 등이 미래에 도입될 것으로 보이며, 이는 효율적인 차량 간 통신(차량 대 기지국, 차량 대 차량, 차량 대 네트워크 등) 기술의 발전 필요성을 증대시킨다[1]. 한편, 밀리미터파(millimeter wave) 및 테라헤르츠(terahertz) 대역 통신을 통해 다양한 도전적인 서비스 요구사항을 만족시키는 것이 통신의 미래 과제로 대두되고 있고, 차량 간 통신에서도 이와 같은 고주파 대역을 사용하여야 할 것으로 보인다. 그러나, 차량 환경에서의 높은 변동성 및 고주파 대역의 낮은 상관 시간(coherence time)은, 신호의 전송 및 수신에 어려움을 야기한다. 또한, 차량 환경의 특성상 극히 낮은 지연시간으로 정보를 처리해야 하는 상황이 발생할 수 있는데, 일반적인 통신 방식으로는 그러한 요구사항을 만족시키기 어렵다. 따라서, 차량 간 통신에서 낮은 지연시간으로 효율적인 자원 분배를 통해 고주파 대역 통신이 가능케 하는 기법의 개발이 필요하다.

본 논문에서는 심층강화학습을 활용해, 차량들의 요구조건을 만족하면서 네트워크 전송률을 최대화하는 자원 할당 기법을 제시하고자 한다. 송신단 차량에 위치한 심층강화학습 네트워크가 주변 수신단 차량의 채널 정보를 얻어 이를 네트워크 입력으로 이용해 차량별 자원 할당 방식을 정한다. 변동성이 있는 환경에서 주변 상황에 적응하여 의사 결정을 내리는 데 특화된 심층강화학습을 사용해 변동성이 높은 차량 통신 환경에서 효율적으로 수신 차량 별 자원 할당을 수행할 수 있다[2]. 또한, 심층신경망을 사용하기 때문에 학습이 끝난 이후, 적용하는 상황에서는 적은 복잡도로 문제를 해결할 수 있다. 따라서, 본 논문에서는 심층강화학습을 활용하여 적은 지연시간으로 자원 할당 문제를 해결하는 기법을 제시한다.

본 논문은 다음과 같이 구성된다. 서론에 이어 II장에서 시스템 모델을 설명하고, III장에서 제안하는 심층강화학습 기반 자원 할당 기법을 소개하며 IV장에서는 실험결과를 설명한고, V장에서 결론을 내린다.

Ⅱ. 시스템 모델

본 논문에서는 차량 대 차량 통신 환경에서 심층강화학습 기반 자원 할당 기법을 적용하기 위해, 정보를 송신하는 M대의 차량과 주변 K대의 수신 차량이 있

※ 본 연구는 방위사업청과 국방과학연구소가 지원하는 미래전투체계 네트워크기술 특화연구센터 사업의 일환으로 수행되었습니다 (UD190033ED).
◆ First Author : (ORCID:0000-0003-2937-5212)Seoul National University, INMC, jhmoon@islab.snu.ac.kr, 학생(박사), 정회원
° Corresponding Author : (ORCID:0000-0001-5051-1763)Seoul National University, INMC, bshim@snu.ac.kr, 정교수, 종신회원
논문번호 : 202209-213-B-LU, Received September 16, 2022; Revised October 1, 2022; Accepted October 1, 2022.

는 네트워크 상황을 가정한다. 송신 차량은 N개의 균등하게 배열된 선형 안테나를 가지며 수신 차량은 1개의 안테나를 가진다. 채널 모델은 다음과 같다.

$$\boldsymbol{h}_{m,k}=\sum_{l=1}^{L}\alpha_{m,k,l}\boldsymbol{a}(\theta_{m,k,l}). \quad (1)$$

여기서 L은 경로의 수, $\alpha_{m,k,l}\sim CN(0,1)$는 경로 이득이고 $\theta_{m,k,l}$는 송신 각도를 나타낸다. $\boldsymbol{a}(\theta)$는 안테나 배열 응답 벡터(antenna array response vector)로서 $\boldsymbol{a}(\theta_{m,k,l})=[1,e^{j\pi sin\theta_{m,k,l}},\cdots,e^{j(N-1)\pi sin\theta_{m,k,l}}]^{\top}$으로 정의된다. 심층강화학습은 얻은 데이터에 기반해 결정을 내리기 때문에, 다른 채널 모델이 주어져도 그에 적합하게 학습될 수 있다. m번째 송신 차량의 자원 할당 상황과 네트워크 전체 자원 할당 상황은 다음과 같이 행렬 $\boldsymbol{S}_m$ 및 $\boldsymbol{S}$로 정의할 수 있다.

$$\boldsymbol{S}_m=\begin{bmatrix} r_{m,1,1}\cdots r_{m,1,S}\\ \vdots\ \ddots\ \vdots\\ r_{m,K,1}\cdots r_{m,K,S}\end{bmatrix},\boldsymbol{S}=[\boldsymbol{S}_1,\ldots,\boldsymbol{S}_M]. \quad (2)$$

이때 $r_{m,k,s}$는 m번째 송신 차량의 k번째 수신 차량에 s번째 주파수 자원 할당 여부에 대한 변수(1: 할당, 0: 비할당)이다. 수신 차량별 데이터 전송률은

$$R_k(\boldsymbol{S})=\log_2\left(1+\frac{\sum_{m=1}^{M}\sum_{s=1}^{S}P_{m,k,s}|\boldsymbol{h}_{m,k}{}^{H}\boldsymbol{f}_{m,k}|^2}{\sum_{m=1}^{M}\sum_{s=1}^{S}\sum_{j\neq k}^{K}P_{m,j,s}|\boldsymbol{h}_{m,k}{}^{H}\boldsymbol{f}_{m,j}|^2+\sigma^2}\right) \quad (3)$$

이다. $\boldsymbol{f}_{m,k}$는 송신 차량의 빔포밍 벡터로 $\boldsymbol{f}_{m,k}=\frac{\boldsymbol{h}_{m,k}}{|\boldsymbol{h}_{m,k}|}$(maximum ratio transmission)이다.

Ⅲ. 심층강화학습 기반 자원 할당 기법

본 절에서는 제안하는 심층강화학습 기반 자원 할당 기법을 설명한다. 심층강화학습의 기본 개념에 대한 설명 이후 자원 할당에 적용하는 방식을 설명한다.

심층강화학습이란, 환경과의 정보 교환을 통해 점차적으로 좋은 순차적 의사 결정을 내릴 수 있는 기계학습의 한 종류인 강화학습과 심층신경망을 결합한 것이다[3]. 심층강화학습의 핵심은 상태(state), 행동(action), 보상(reward)이라는 세 가지 요소이며 이와 함께 심층신경망의 함수 근사 능력을 사용해 최적의 의사 결정 정책을 찾아간다. 무선통신에서는 이같이 최적의 의사 결정 정책을 찾는 심층강화학습을 여러 영역에서 사용할 수 있다[2].

심층강화학습의 목표는 보상의 누적 합인 Q값을 최대화하는 것인데, Q값은 다음과 같이 정의된다.

$$Q(\boldsymbol{s},\boldsymbol{a})=E\left[\sum_{t'=t}^{\infty}\gamma^{t'}r_{t'}|\boldsymbol{s}_t=\boldsymbol{s},\boldsymbol{a}_t=\boldsymbol{a}\right]. \quad (4)$$

최적의 Q값을 찾기 위해 다음 Bellman equation을 활용한다: $Q(\boldsymbol{s},\boldsymbol{a})=r+\gamma\sum_{s'\in S}P_{ss'}^{\boldsymbol{a}}E[Q(\boldsymbol{s}',\boldsymbol{a}')]$. 이때, $P_{ss'}^{\boldsymbol{a}}$은 상태 전환 확률이며 γ는 감쇄 인자이다.

제안하는 기법에서는 심층신경망을 통해 actor 네트워크와 critic 네트워크를 구현해, actor-critic 알고리즘으로 수신 차량들의 데이터 전송률 합을 최대화하는 자원 할당 방식을 구현한다.

1) 상태(state) $\boldsymbol{s}$: 송신단과 각 수신단 사이의 채널 이득 $\alpha_{m,k,l}$, 각도 $\theta_{m,k,l}$, 데이터 요구량 $R_{req,k}$.
2) 행동(action) $\boldsymbol{a}$: 자원 할당 행렬 $\boldsymbol{S}$.
3) 보상(reward) r: 전체 네트워크 데이터 전송률의 합인 $\sum_{k=1}^{K}R_k(\boldsymbol{S})$를 사용해 전체 네트워크의 데이터 전송률 합을 최대화하도록 한다. 또한, 수신 차량의 데이터 전송률 요구 조건을 만족하기 위해서, 보상 r에 요구 조건 달성을 실패한 차량마다 일정한 페널티(음의 실수)를 더해 최종 r은 다음과 같이 결정된다($I(x)$는 x가 참일 때 1, 아니면 0인 함수).

$$\sum_{k=1}^{K}[R_k(\boldsymbol{S})-I(R_k(\boldsymbol{S})<R_{req,k})]. \quad (5)$$

Critic 네트워크는 상태를 입력으로 받아 각 행동별 Q값을 예측한다. Bellman equation을 만족시키는 방향으로 critic 네트워크를 학습시킨다. 한편, actor 네트워크는 상태를 입력으로 받아 행동을 출력으로 낸다. 즉, 현재 네트워크 상황을 입력받아 각 수신 차량의 요구 조건을 만족하면서 데이터 전송률 합을 최대화하는 자원 할당 행렬 $\boldsymbol{S}$를 출력한다. 이때 행동은 Q값을 최대화하도록 학습시킨다. 따라서, actor와 critic 네트워크의 손실 함수는 각각 다음과 같다.

$$L_c=\frac{1}{2}(r+\gamma E_{\boldsymbol{a}'}[Q(\boldsymbol{s}',\boldsymbol{a}')]-Q(\boldsymbol{s},\boldsymbol{a}))^2. \quad (6)$$

$$L_a=-E_{\boldsymbol{a}}[Q(\boldsymbol{s},\boldsymbol{a})]. \quad (7)$$

Ⅳ. 실험 결과 및 논의

본 절에서는 제안하는 자원 할당 기법의 시뮬레이션 결과를 제시한다. 비교 기법은 기존의 round robin 방식과 proportional fair 기법이다[4]. 또한, 최적의 자원 할당(요구조건을 만족하는 상황 중 최대 데이터 전송률 합)을 수행했을 경우와도 데이터 전송률 합을 비교한다. 송신 차량 2대가 주변 3대의 차량에 4개의 주파수 자원을 할당하고, L은 1, N은 16인 상황을 가정하였다. 잡음의 크기는 -174dBm/Hz로 설정하였고 $R_{req,k}$는 [0,1]bps/Hz에서 랜덤 설정하였다. 또한 심층강화학습 네트워크는 actor와 critic 네트워크 각각 200의 폭을 가진 은닉층 3개이고, γ는 0.95이다.

그림 1은 송신 전력 대비 데이터 전송률 합 그래프이다. 송신 전력을 바꿔 가면서 성능을 비교했을 때, 제안하는 기법이 기존 기법들에 비해 우수한 성능을 보이는 것을 확인하였다. 제안하는 기법이 기존 기법들에 비해 평균적으로 각각 215.5%, 19.1% 이상의 데이터 전송률 이득이 있는 것을 볼 수 있다. 이러한 결과는 심층강화학습 에이전트가 수신 차량들의 데이터 전송률 요구 조건을 복합적으로 고려하면서 전체 네트워크 데이터 전송률 합을 최대화하는 결정을 내리는 것을 보여준다. 특히, proportional fair 기법에 비해, 채널이 좋은 수신 차량에 더 많은 자원을 할당해 더 높은 데이터 전송률 합을 얻는 것을 볼 수 있다.

또한, 본 기법은 최적의 자원 할당 방식과 비교해 약 5.7%의 성능 열화가 있는 것을 확인하였다. 본 심층강화학습 기반 기법은 심층신경망을 통해 자원 할당 방식을 결정하기 때문에, 최적화 기반 기법 등 복잡한 기존 기법에 비해 네트워크 규모에 덜 민감하다.

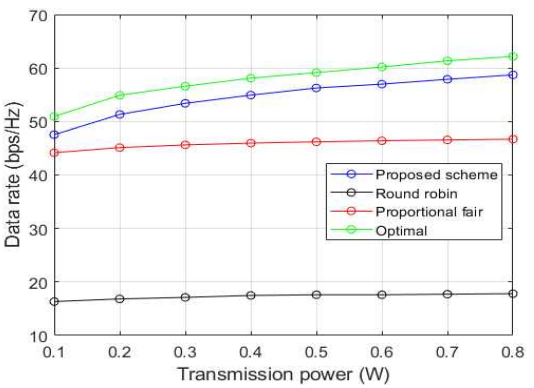

그림 1. 심층강화학습 기반 자원 할당 기법의 데이터 전송률 합 성능 그래프
Fig. 1. Sum rate performance of DRL-based resource allocation.

구체적으로, 심층신경망의 actor 네트워크 속 레이어(layer) 개수와 레이어 당 폭(width) 수가 일정하면, 차량 수나 자원의 수가 증가할 때 actor 네트워크의 연산량도 거의 동일하게 유지할 수 있다. 반대로, 최적 자원 할당 기법의 경우 $O(M^2K^2NS)$의 복잡도를 가져 M, K, N, S가 커지면서 복잡도가 증가한다. 따라서, 본 심층강화학습 기반 기법은 최적 기법에 준하는 성능을 보이면서 차량과 자원의 수가 많아지는 경우에 자원 할당 처리 지연시간이 급증하는 것을 막아, 차량 네트워크의 빠른 환경 변화에 대응할 수 있다.

Ⅴ. 결 론

본 논문에서는 심층강화학습을 사용하여 차량 간 통신에서 자원 할당 문제를 해결하는 기법을 제시하였다.. 제안하는 기법은 기존 자원 할당 기법 대비 높은 데이터 전송률을 달성하는 것을 확인하였다.

References

[1] P. K. Singh, S. K. Nandi, and S. Nandi, “A tutorial survey on vehicular communication state of the art, and future research directions,” *Veh. Commun.*, vol. 18, no. 100164, 2019. (https://doi.org/10.1016/j.vehco m.2019.100164)

[2] H. Ju and B. Shim, “데이터 전송 시간 단축을 위한 심층강화학습 기반 자원할당 기법 연구,” *KICS Winter Conf.*, pp. 1230-1231, 2021.

[3] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529-533, 2015. (https://doi.org/10.1038/nature14236)

[4] M. T. Kawser, H. M. A. B. Farid, A. R. Hasin, A. M. J. Sadik, and I. K. Razu, “Performance comparison between round robin and proportional fair scheduling methods for LTE,” *Int. J. Inf. Electron. Eng.*, vol. 2, no. 5, pp. 678-681, 2012. (https://doi.org/10.7763/ IJIEE.2012.V2.186)