RF 충전 후방산란 CR 네트워크에서 효율적인 강화학습 기반 모드 최적화

오선애, 신요안*

숭실대학교 전자정보공학부

Email: {sunae0814@soongsil.ac.kr, yashin@ssu.ac.kr}

An Efficient Mode Optimization Based on Reinforcement Learning in RF-powered Backscatter CRNs

Shanai Wu, Yoan Shin*

School of Electronic Engineering, Soongsil University

(*Corresponding author)

요 약

본 논문은 RF 충전 후방산란 인지 무선 네트워크에서 2차 송신단말 (Secondary Transmitter; ST)이 1차 채널과 상호작용하면서 받는 보상을 통해 최적의 정책을 효율적으로 학습하는 방안을 제안한다. ST는 환경에서 변화되는 자신의 상태에 적합한 동작 모드를 수행하면서 주어진 시간 동안에 최대한 많은 데이터 패킷을 전송하는 것을 목표로 하며, 학습의 효율을 높이기 위해 Energy Outage가 발생하는 동작 모드를 선택한 경우에 Penalty를 부여하는 방안을 제안하였다. 랜덤하게 변화하고 예측이 어려운 환경에서 안정적인 학습이 가능한 Deep Q-network 알고리즘을 사용하였으며, ST가 학습된 인공신경망 모델로부터 모드를 선택하여 수행하는 모의실험을 통해 제안 기법의 학습 성능을 검증하였다.

Ⅰ. 서 론

주변의 무선 주파수 (Radio Frequency; RF) 신호로부터 에너지를 충전하는 RF 에너지 수집 기술이 센서 노드와 같은 저전력 단말의 자기 유지 가능한 (Self-sustainable) 에너지 공급 기술로 부상하고 있다. 주파수 이용 효율 극대화를 위한 인지 무선 (Cognitive Radio; CR) 기술과 결합한 RF 충전 CR 네트워크에서 2차 송신단말 (Secondary Transmitter; ST)은 주변의 1차 신호로부터 에너지를 수집하고, 1차 채널이 비어 있는 동안에 수집한 에너지를 사용하여 데이터를 전송하는 방식으로 동작할 수 있다. 따라서 ST의 전송 성능은 1차 시스템에 의해 결정되며, ST의 전송 성능을 개선하기 위해 주변 후방산란 통신 (Ambient Backscatter Communication; AmBC) 기술의 적용이 제안되었다[1]. AmBC는 주변 RF 신호를 반사하여 정보를 전송하는 통신 기술로서 전력소모가 적기 때문에 에너지 수집과 결합하여 효율적인 무선충전 통신 네트워크를 구성할 수 있다. 또한 RF 신호를 수신하는 단말로부터 7.2인치 이상 떨어져 있으면 후방산란 간섭을 발생하지 않는다는 연구결과가 보고된 바 있다[2].

RF 충전 후방산란 CR 네트워크에서 예측하기 어려운 1차 채널에 접근하여 점유 상태에 적합한 모드로 동작하면서 일정 시간 동안에 최대의 전송 성능을 얻기 위해, ST는 랜덤하게 변화하는 상태를 고려하여 순차적으로 동작 모드를 결정해야 한다. 이를 위해 본 논문에서는 강화학습의 활용 방안을 고려하였으며, 강화학습은 환경에서 시도한 행동과 그 결과로 나타나는 보상 사이의 상관관계를 시행착오를 통해 학습하는 방법이다.

Ⅱ. RF 충전 후방산란 CR 네트워크

본 논문에서는 1차 시스템과 2차 시스템이 각각 한 쌍의 송수신단으로 구성된 RF 충전 후방산란 CR 네트워크를 고려하였다. 1차 시스템에서 주사용자 (Primary User; PU)는 스펙트럼 대역에 접근할 수 있는 권한을 갖고 있는 사용자이며, 1차 채널의 상태는 PU의 전송 패턴에 따라 변화한다. 본 논문에서는 타임 슬롯 기반의 네트워크 모델을 고려하며, 따라서 임의 슬롯에서 1차 채널이 PU에 의해 사용될 확률이 p인 베르누이 분포에 따라 간단하게 모델링할 수 있다. 잔여 에너지가 충분한 경우에 ST는 1차 채널이 비어 있으면 Active 모드로 데이터를 전송하며, PU가 채널을 사용하고 있으면 RF 신호를 후방산란하여 정보를 전송하거나 에너지를 수집하여 배터리에 저장한다. ST는 타임 슬롯의 시작점에서 확률 λ로 발생하는 데이터를 저장장치에 보관하였다가 데이터가 저장된 순서에 따라 순차적으로 전송하며, 저장 공간이 부족한 경우에 가장 오래된 데이터부터 손실하게 된다. 전송 성능을 최대화하기 위해 ST는 1차 채널의 점유 상태에 적합한 모드로 동작해야 하며, 따라서 각 슬롯의 시작점에서 일정 시간 동안 스펙트럼을 센싱하는 방식으로 1차 채널의 점유 상태를 판단할 수 있다. 하지만 ST가 RF 신호로부터 수집 가능한 에너지가 제한적이기 때문에 부가적으로 소모되는 에너지를 최소화하기 위해, 본 논문에서는 에너지 수집 모드를 통해 1차 채널의 점유 상태를 판단하는 방안을 제안한다. 제안 기법은 슬롯을 두 개의 서브 슬롯으로 나누어 동작하며, 첫 번째 서브 슬롯에서 ST는 수집 모드를 통해 에너지 상태의 변화를 관찰하여 1차 채널의 점유 상태를 판단한다. 즉, 배터리의 잔여 에너지가 증가하면 PU가 1차 채널을 사용하고 있다고 판단하여 두 번째 서브 슬롯에서 주변 후방산란 모드로 동작하게 되고, 그렇지 않으면 1차 채널이 비어 있다고 판단하여 Active 전송 모드도 동작한다. ST는 에너지 수집을 통해 얻은 관찰 값으로 1차 채널의 점유 상태에 적합한 모드로 동작하면서 데이터를 성공적으로 전송하는 것도 중요하지만, 1차 채널과 상호작용 하면서 변화하게 될 자신의 상태에서 최적의 모드를 선택하면서 주어진 시간 동안에 최대한 많은 데이터를 전송하는 것도 중요한 문제이다. 따라서 ST는 순차적으로 동작 모드를 결정해야 하며, 본 논문에서는 강화학습을 통해 ST가 임의 상태에서 최적의 모드를 선택하는 정책을 학습하고자 한다.

Ⅲ. 제안하는 강화학습 기반 모드 최적화

ST가 순차적으로 동작 모드를 결정하는 문제에 접근할 수 있도록 하기 위해, 마르코프 결정 과정 (Markov Decision Process; MDP)을 통해 수학적으로 문제를 정의해야 한다. MDP를 구성하는 요소들로는 상태, 행동,

보상, 상태전이확률, 감가율 등이 있다. 따라서 본 논문에서는 첫 번째 서브 슬롯에서 에너지 수집 모드를 통해 얻은 관찰 값, 배터리의 잔여 에너지, 데이터 큐 (Queue) 등이 ST의 현재 상태 s가 되며, 이와 같은 상태들을 고려하여 ST가 두 번째 서브 슬롯에서 선택할 수 있는 행동 a들로는 Active 전송과 주변 후방산란뿐만 아니라 Idle 모드와 수집 모드가 있다. ST는 첫 번째 수집 모드를 통해 1차 채널이 비어 있다고 판단되어도 잔여 에너지가 부족하거나 저장된 데이터 패킷이 적은 경우에 Idle 모드로 동작하게 되며, 전송해야 하는 데이터가 적으면 1차 채널이 점유되었다고 판단하여도 Idle 모드로 동작하거나 배터리가 완전히 충전되지 않으면 에너지를 수집할 수 있다. ST는 Active 전송과 후방산란이라는 행동을 통해 데이터 전송에 따른 보상 (r)을 받게 되며, 상태에 적합한 행동을 선택하면 양의 보상을 받게 되고 상태에 적합하지 않은 행동을 선택하면 음의 보상을 받게 된다. 행동을 취한 후 다음 슬롯에 도달하게 될 상태 s'에는 확률적인 요인이 포함된다. ST는 에너지 상태와 데이터 큐의 변화를 스스로 관찰할 수 있는 반면에 환경에 해당하는 1차 채널의 상태 변화를 알 수 없다. 따라서 제안 기법은 에너지 수집 모드를 위한 서브 슬롯을 할당하여 1차 채널의 상태 변화를 관찰한다. 또한 ST가 행동을 결정하는 시점인 현재에 가까운 보상일수록 큰 가치를 갖도록 하기 위해 감가율을 사용하여 나중에 받게 될 보상의 가치를 감소할 수 있다.

본 논문에서는 Deep Q-network (DQN) 알고리즘을 사용하여 ST가 최적의 정책을 학습할 수 있도록 하였다. DQN은 Off Policy인 Q-learning 알고리즘이 Q값을 탐욕 (Greedy) 정책에 따라 업데이트하는 방식과 동일하게 경사하강법을 사용하여 오류함수를 최소화하도록 인공신경망의 가중치를 학습시킨다[3]. 반면에 환경에서 충분히 탐험하기 위해 행동 정책은 ϵ-탐욕 정책을 따른다. 또한 DQN은 환경에서 탐험하면서 얻은 (s, a, r, s') 샘플을 리플레이 메모리에 저장하였다가 학습에 사용하는 경험 리플레이 (Experience Replay)를 통해 샘플들의 상관관계가 학습에 주는 영향을 완화하였으며, 한 개의 샘플로 학습하는 것이 아니라 배치로 학습하기 때문에 학습이 안정적이다. 학습의 목표가 되는 정답이 타임 스텝마다 변하는 것을 방지하기 위해, DQN은 목표 인공신경망을 따로 구현하여 정답 역할을 하는 값을 제공하며 일정 시간 간격마다 학습하는 인공신경망의 가중치로 업데이트하여 준다. 리플레이 메모리는 사이즈가 정해져 있기 때문에 오래된 샘플들부터 삭제되며, 샘플들에 가중치를 할당할 수 없기 때문에 학습에 유리한 샘플들도 삭제된다. ST는 유한한 에피소드를 반복하면서 타임 스텝마다 인공신경망의 가중치를 업데이트하며, 본 논문에서는 Energy Outage (EO)가 발생하는 행동을 취하면 Penalty를 부여하는 동시에 에피소드를 강제로 종료시켜 나쁜 상황에서 신속하게 벗어나도록 하여 학습 효율을 향상시키는 방안을 제안하였다.

Ⅳ. 모의실험 결과 및 결론

제안 기법을 통해 ST가 달성 가능한 성능을 검증하기 위해 타임 스텝이 500인 에피소드를 반복하면서 학습을 수행하도록 하였으며, 학습 과정은 그림 1에서 도시한다. DQN 알고리즘의 하이퍼파라미터는 표 1에서 정리하였으며, 목표 신경망은 에피소드마다 업데이트하였다. 데이터 저장장치와 배터리의 용량은 모두 10으로 고려하였으며, 한 개의 슬롯을 기준으로 수집 가능한 평균 에너지와 Active 모드로 동작하면서 소모하는 에너지를 모두 1로 고려하였으며, Active 전송과 후방산란을 통해 각각 2개와 1개의 데이터 패킷을 전송할 수 있다고 가정하였다. 또한 데이터 발생 확률 λ와 1차 채널이 사용될 확률 p를 모두 0.5로 설정하였다. 그림 2는 EO Penalty를 적용하여 학습하는 과정을 도시하며, Penalty를 적용하지 않은 경우에 비해 학습 시간이 짧은 것을 확인할 수 있다. 그림 3은 ST가 두 가지 방식으로 학습된 DQN 모델로부터 탐욕 정책에 따라 행동하면서 얻을 수 있는 전송 성능을 도시하였다. 500개의 타임 스텝으로 구성된 유한한 에피소드를 500번 반복하면서 확인해 본 결과, 제안된 EO 패널티를 적용한 학습을 통해 DQN이 충분히 학습한 경우에 근접한 성능을 얻을 수 있음을 확인하였다.

표 1. DQN의 하이퍼파라미터

Hyper-parameter	Value
Number of hidden layers	2
Activation function	"ReLU" for hidden layers, "Linear" for output layer
Optimization	Adam Optimizer
Learning rate	0.0001
Discount factor	0.99
Epsilon	1 → 0.01
Batch size	64
Size of replay memory	2,000

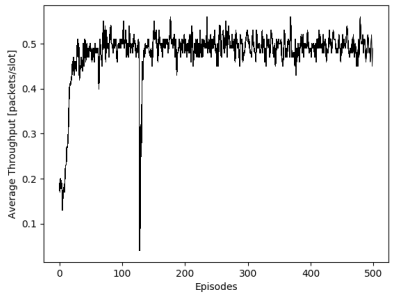

그림 1. DQN을 통한 제안 기법의 학습 과정 (Learning Steps: 250,000)

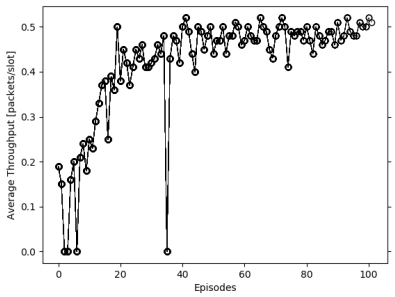

그림 2. 제안하는 EO Penalty를 적용한 학습하는 과정

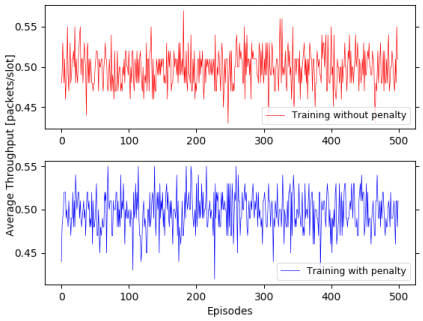

그림 3. 학습된 모델로부터 모드를 선택하여 수행한 성능 비교

ACKNOWLEDGMENT

본 논문은 2014년 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구결과임 (2014R1A5A1011478).

참 고 문 헌

[1] D. Hoang *et al.*, "Ambient backscatter: A new approach to improve network performance for RF-powered cognitive radio networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3659-3674, Sept. 2017.

[2] V. Liu *et al.*, "Ambient backscatter: Wireless communication out of thin air," *Proc. ACM SIGCOMM 2013*, Hong Kong, China, Aug. 2013.

[3] V. Mnih *et al.*, "Playing Atari with deep reinforcement learning," *Proc. NIPS 2013*, Lake Tahoe, USA, Dec. 2013.