심층 강화학습 기반의 무선랜 AP 선택 기법

정창현*, 김예림, 김지훈, 이상희, 이재욱

부경대학교 정보통신공학과

{ckdgus4123*, kgflafla02, kkkkkkk1379, dltkdgml4730}@pukyong.ac.kr, jlee0315@pknu.ac.kr

Deep Reinforcement Learning based WLAN AP selection scheme

Changhyun Jung*, Yerim Kim, Jihoon Kim, Sanghui Lee and Jaewook Lee

Dept Information and Communications Eng., Pukyong National University

요 약

실내 행사장과 같이 AP 간 커버리지가 겹치는 밀집 환경에서는 사용자 분배가 적절하지 않아 특정 AP에 사용자가 몰리면서 인터넷 품질이 저하되는 문제가 발생한다. 본 연구는 이러한 문제를 해결하기 위해 심층 강화학습의 하나인 DQN을 적용하여, 다수 AP가 존재하는 환경에서 단말의 요구 통신 속도를 만족시키는 최적의 단말-AP 선택 기법을 제안한다. 실험을 통해 제안하는 기법이 무선 네트워크 환경에서 단말들이 요구하는 데이터 속도를 공평하게 만족시키는 최적의 AP들을 선택하는 것을 확인하였다.

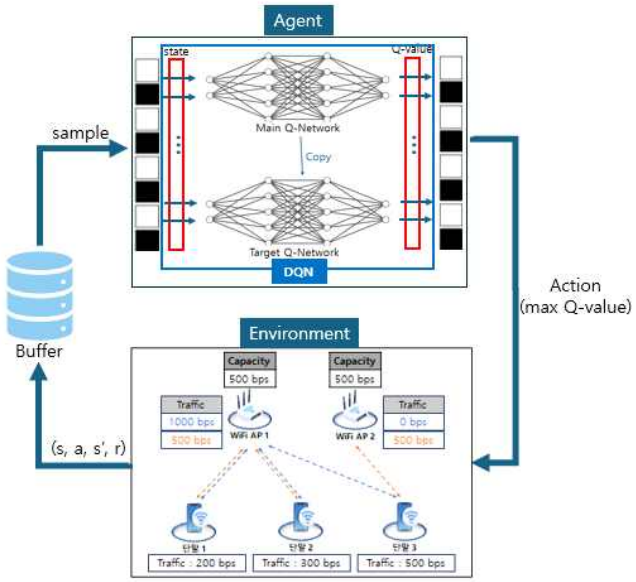
Ⅰ. 서 론

최근 다수의 AP를 분배하여 무선 네트워크의 커버리지를 넓히는 Mesh WiFi 기술이 활발히 연구 및 상용화되고 있다. 그러나 기존의 Mesh WiFi 기술은 AP들의 접속 반경이 겹치지 않게 배치하여 연결 서비스를 지원하는 방식이기 때문에, 실내 행사장과 같이 AP 간 커버리지가 겹치는 밀집된 환경에서는 사용자들을 적절히 분배하지 못하는 문제가 있다. 이로 인해 특정 AP에 많은 사용자가 몰리면서 인터넷 품질이 저하되는 문제가 발생한다. 따라서, 본 연구에서는 다수 AP가 존재하는 환경에서 단말들이 요구하는 통신 속도를 만족할 수 있는 단말과 AP 간의 연결 정책을 결정하는 심층 강화학습 기반의 AP 선택 기법을 제안한다. 본 논문에서 단말들이 특정 WiFi AP와 연결될 수 있도록 관련 연구[1] 에서 구현된 WiFi 시스템 컨트롤러가 존재하는 WiFi 시스템을 가정하였다. 최적의 단말, AP 연결 정책을 얻기 위해서 강화학습 알고리즘의 하나인 DQN (Deep Q-Network) 알고리즘[2]을 적용하였으며, 실험을 통해 제안하는 기법이 무선 네트워크 환경에서 단말들이 요구하는 데이터 속도를 공평하게 만족시키는 최적의 AP들을 선택하는 것을 확인하였다.

본 논문에서는 심층 강화학습 기반의 AP 선택 기법이 적용되는 시스템 모델을 기술하고, 심층 강화학습 문제를 정의한다. 그 후 실험을 통해 제안한 기법의 우수성을 검증한다.

Ⅱ. 심층 강화학습 기반의 AP 선택 기법

본 논문에서는 M개의 WiFi AP가 배치되어 있는 상황에서 N개의 단말이 존재하는 중앙 집중식 무선랜 시스템 환경을 가정한다[1]. 해당 환경에서 에이전트는 제안하는 기법을 통해 단말별로 연결되어야 하는 AP를 선택하고, 해당 정보를 각 단말들에게 알린다. 해당 단말들은 수신받은 AP 선택 정책에 따라 AP를 변경한다. 에이전트는 최적의 AP 선택하기 위해 [그림 1]과 같이 AI 모델을 DQN 기법을 통해 학습한다. 학습하는 동안 에이전트는 현재 단말이 연결되어야 할 AP들을 선택하고 단말들에게 알리고, 단말은 해당 AP를 연결한다. 그리고 단말은 현재 연결된 AP 정보와 요구하는 데이터 속도 정보를 에이전트에게 알림으로써 에이전트는 자신이 선택한 결정의 좋고 나쁨을 배우게 되고 지속적으로 상기 과정을 수행하여 최적의 선택을 AI 모델이 배울 수 있도록 한다.



〔그림 1〕 시스템 구조

최적의 사용자별 AP 선택 정책을 얻기 위해 DRL 문제(상태 공간(S)과 행위 공간(A) 그리고 보상 값(R))를 다음과 같이 정의하였다. 상태 공간 S는 각 단말별 현재 연결된 AP의 상태 공간과 단말별 요구하는 데이터 속도의 상태 공간으로 아래와 같이 정의된다.

$$S = \prod_{n=1}^{N} W_n \times \prod_{n=1}^{N} D_n$$

여기서 W_n 과 D_n 은 각각 단말 n이 연결될 수 있는 AP와 요구하는 데이터 속도의 상태 공간을 나타내며 W_n = {1, 2, ..., M} 과 D_n ={0, ..., $D_{\max}$} 로 정의할 수 있다. $D_{\max}$는 단말이 요구할 수 있는 최대 데이터 속도를 나타낸다.

행위 공간 A는 에이전트가 각 단말들이 연결되어야 할 AP들로 나타내며 다음과 같이 정의한다.

$$A = \prod_{n=1}^{N} A_n$$

여기서 A_n는 단말 n이 연결되어야 하는 AP를 의미하고 A_n의 상태 공간은 W_n과 같다.

끝으로 보상 값 R은 다음과 같이 정의하였다.

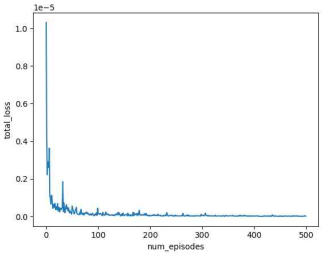
$$r(s,a) = \frac{\left[\sum_{m=1}^{M} \theta_m(r_n(s,a))\right]^2}{M\sum_{m=1}^{M}\left[\theta_m(r_n(s,a))\right]^2}$$

$r_n(s,a)$는 특정 상황 s에서 행위 a를 선택할 시 단말 n의 요구 사항 충족도를 나타내며 낮은 값일수록 요구 사항을 더 충족한다는 것을 의미한다. $r_n(s,a)$는 아래와 같이 정의된다.

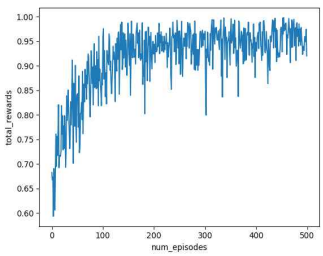
$$r_n(s,a) = \max\left\{(d_n - \frac{B_{a_n}}{N_{a_n}}(\delta(w_n = a_n) + p(1-\delta(w_n = a_n)), 0\right\}$$

d_n은 단말 n이 요구하는 데이터 속도이고, B_{a_n}는 단말 n과 연결되는 AP (a_n)가 제공하는 데이터 속도를 의미하며, N_{a_n}는 AP (a_n)와 연결된 단말의 수를 의미한다. 또한, δ 함수는 조건이 참이면 1을 출력하고 그렇지 않으면 0을 출력하는 함수로, 기존의 AP가 변경되지 않으면 1을 출력하고 AP가 변경된다면 0을 출력한다. p는 전체 서비스 시간 동안 단말 n의 AP 변경으로 인한 연결 시간을 제외한 서비스 받는 시간의 비율을 의미한다. $\theta_m()$ 함수는 AP m에 연결되어 있는 단말들의 $r_n(s,a)$ 값을 합하는 함수다. 정의된 보상 값들을 통해 동일한 AP에 다수의 단말이 연결되거나 AP가 변경될수록 단말들은 적은 데이터 속도로 서비스를 받는다. 단말들의 공평한 충족도를 보장하고자 Jain's fairness index를 활용하여 $r(s,a)$를 정의하였다. 에이전트는 정의된 상태, 행위 공간들과 보상 값을 토대로 최적의 보상 값을 보장할 수 있는 최적의 AP 선택 정책을 학습하기 위해 ε-greedy, 리플레이 버퍼, 타겟 네트워크 등의 DQN 기법을 적용한다[2].

Ⅲ. 모의실험 결과.



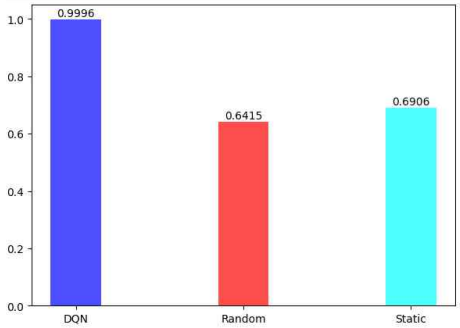
〔그림2〕 손실 그래프



〔그림3〕 보상 그래프

본 실험에서는 500개의 에피소드를 통해 에이전트를 학습시켰다. 각 에피소드는 20번의 스텝으로 구성되며, 각 에피소드에서의 평균 손실과 평균 보상을 측정하였다. 또한, 3개의 AP와 6개의 단말이 존재하고, 각 단말들은 서로 다른 데이터 속도를 요구하도록 가정하였다.

[그림 2]는 에피소드에 따른 평균 손실을 나타낸다. 평균 손실은 학습 초기에는 매우 높았으나, 약 100 에피소드 이후 급격히 감소하였으며, 이후 점진적으로 안정화되었다. 이는 에이전트가 환경에 적응하고 학습이



〔그림 4〕 case 별 성능 비교

진행됨에 따라 손실이 감소하는 것을 보여준다. [그림 3]은 에피소드에 따른 평균 보상을 나타낸다. 에이전트는 DQN 모델을 통해 지속적으로 학습을 진행하며, 초기에는 낮은 보상을 받다가 학습이 진행됨에 따라 평균 보상이 크게 증가하는 경향을 보인다. 약 300 에피소드 이후부터는 평균 보상이 안정화되는 모습을 보이는데, 이는 에이전트가 환경에서의 최적 행동 정책을 성공적으로 학습했음을 뜻한다.

[그림 4]는 제안하는 기법(DQN)과 임의의 AP들을 선택하는 Random 기법, 그리고 2개의 단말들을 각각의 AP에 공평하게 분배하는 기법인 Static 기법의 성능 비교 결과 그래프이다. 해당 결과를 통해 제안하는 기법의 평균 보상값 0.9996으로 우수한 성능을 보였다. 이는 Random 기법 경우 평균 보상값 0.6415와 비교할 때, 제안하는 기법이 무선 네트워크 환경에서 보다 효율적으로 AP 선택 및 트래픽 분배를 수행함을 나타낸다. 또한, Static 기법의 경우 평균 보상값이 0.6906으로, 이는 DQN 모델의 성능에 비해 낮은 값을 보였다. 제안하는 시스템은 지능적으로 단말들의 상황에 따라 네트워크를 최적화시킬 수 있기 때문이다. 이러한 결과는 DQN을 이용한 접근 방식이 단순 랜덤 선택보다 효과적이며, 무선 네트워크 성능 최적화에 유의미한 개선을 가져올 수 있음을 뜻한다.

Ⅳ. 결론

본 논문에서는 강화학습 기반의 무선랜 AP 선택 기법을 제안한다. 최적의 단말, AP 연결 정책을 얻기 위해서 강화학습 알고리즘의 하나인 DQN 알고리즘을 적용하였으며, 모의실험을 통해 해당 알고리즘의 성능을 확인하였다.

제안하는 기법은 다수의 AP가 존재하는 무선 네트워크 환경에서 단말들이 요구하는 데이터 속도를 공평하게 만족시키는 최적의 AP들을 선택함으로써 쾌적한 네트워크 환경을 제공한다. 향후 연구에서는 다양한 실시간 네트워크 환경과 사용자 이동성 시나리오를 고려한 확장된 강화학습 모델을 개발하고, 사용자 패턴 분석 및 예측을 기반으로 무선 네트워크의 효율성을 발전시킬 예정이다.

참 고 문 헌

[1] C. H. Jung, Y. R. Kim, J. H. Kim, S. H. Lee and J. W. Lee, (2024), "Intelligent Centralized Wireless LAN Network System", KICS Winter Conference 2023.

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller, (2013), "Playing Atari with Deep Reinforcement Learning", DeepMind Technologies.