https://doi.org/10.7840/kics.2025.50.8.1265

강화 학습을 사용한 무선 네트워크 MAC 프로토콜의 조사

박 준 영◆, 나 웅 수°

Survey of Wireless Network MAC Protocols Using Reinforcement Learning

Jun-young Park◆, Woongsoo Na°

요 약

증가하는 수요와 다양한 환경의 무선 네트워크에서는 충돌, 지연, 공정성, 전력 소모 등과 같은 많은 문제들이 발행하고 있다. 이러한 문제들을 인공지능 기술을 활용하여 해결하거나 최소화시켜 프로토콜의 성능을 향상시키는 연구들이 논의되고 있다. 본 논문에서는 무선 네트워크 환경의 MAC 프로토콜에서 위에서 서술한 문제를 해결하기 위해 강화 학습 기술을 사용한 연구들에 대해서 조사했다. 무선 네트워크 MAC 프로토콜에서 나타나는 문제점을 살펴보고, 강화 학습이 어떻게 활용되었는지 알아본다. 마지막으로 향후 해결해야 하는 문제점을 서술한다

키워드 : 무선 네트워크, MAC 프로토콜, 기계학습, 강화학습, 5G & 6G 네트워크
Key Words : Wireless Network MAC Protocol, Machine Learning, Reinforcement Learning. 5G & 6G Network

ABSTRACT

In wireless networks with increasing demand and diverse environments, numerous issues such as collisions, delays, fairness, and power consumption are arising. Research is being discussed to enhance protocol performance by resolving or minimizing these issues using artificial intelligence technologies. This paper investigates studies that have applied reinforcement learning techniques to address the aforementioned problems in the MAC protocols of wireless network environments. It examines the issues present in wireless network MAC protocols and explores how reinforcement learning has been utilized. Finally, it outlines the challenges that need to be addressed in the future.

Ⅰ. 서 론

무선 네트워크 기술들이 발전함에 따라 스마트폰, 노트북, 태블릿 PC, IoT 장비 등과 같은 기기들의 사용자가 늘어나고 있고, 무선 네트워크에 접속하는 노드들의 수들 또한 늘어가고 있는 추세이다. 무선 네트워크의 성능을 높이기 위해 하드웨어와 소프트웨어적으로 많은 연구가 진행되고 있지만, Medium Access Control(MAC) 프로토콜에서 발생하는 문제들을 해결하는 것은 어려운 일이다. 노드의 수가 늘어남에 따라 충돌이 발생해 네트워크 처리량에 부정적인 영향이 미치는 경우[1], 지연이 발생해 네트워크 처리량에 부정적인 영향이 미치는 경우[2,3], 채널 접근과 데이터 전송 기회가 불공평한 공정성 문제[4], 지연 및 재전송으로

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2023-RS-2022-00156353) 및 2025년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(RS-2019-NR040074)
◆ First Author : Kongju National University Department of Software, jjunyoung99@gmail.com, 학생회원
° Corresponding Author : Kongju National University Department of Software, wsna@kongju.ac.kr, 종신회원
논문번호 : 202501-013-C-RN, Received January 8, 2025; Revised February 26, 2025; Accepted March 18, 2025

인해 지속적으로 에너지를 소모해야 하는 전력 소모 문제[5] 등의 많은 문제점들이 존재하고 이러한 문제점들은 네트워크 성능에 악영향을 끼치고 있다. 이러한 문제점을 해결하기 위해 인공지능 기술을 활용하는 연구가 계속되고 있다.

6G 네트워크에서는 높은 밀도의 환경에서 낮은 지연 시간과 높은 사용자 경험 데이터 전송속도를 필요로 하기 때문에 인공지능과 통합하여 네트워크 관리와 최적화를 자율적으로 수행하는 초지능 네트워크에 대한 연구가 계속되고 있다[45]. MAC 프로토콜도 네트워크 자원 관리를 통해 높은 처리량과 낮은 지연 시간을 목표로 하기 때문에 초지능 네트워크와 융합한다면 6G 네트워크의 요구사항을 지키면서 네트워크 성능을 향상시킬 수 있는 좋은 방법이라고 생각된다. 특히 강화 학습은 에이전트가 환경과 상호작용하며 보상을 최대화하는 방향으로 학습하기 때문에, 네트워크 트래픽 패턴이나 사용자의 요구사항이 지속적으로 변화하는 6G 환경에서 적응력 있는 자원 관리 및 최적화가 가능하다. 그리고 강화 학습은 사전 모델링 없이 실시간 데이터를 기반으로 정책을 개선할 수 있어 복잡한 네트워크 상황에서도 처리량과 지연 시간을 개선하는 데 효과적이다. 이러한 장점들은 MAC 프로토콜에 강화 학습을 적용함으로써 효율적인 네트워크 자원 관리를 할 수 있게 하며 6G 네트워크의 성능을 극대화 하는데 중요한 역할을 할 것이다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 MAC 프로토콜에 강화학습을 사용한 기술들에 대해 살펴본 후 장단점을 비교한다. 마지막으로 Ⅲ 장에서는 결론을 맺으며 마무리한다.

Ⅱ. 강화학습 연구

본 장에서는 강화학습이 무선 네트워크 환경에서의 MAC 프로토콜에 어떻게 사용되는지 알아보고자 한다. 가치 기반 알고리즘, 정책 기반 알고리즘, 하이브리드 알고리즘, 다중 에이전트 강화학습으로 구분하여 소개하고자 한다.

2.1 가치 기반 알고리즘

2.1.1 신경망 없는 가치 기반 알고리즘

논문 [6]에서는 의료용 IoT 네트워크에서 MDP로 모델링하여 MAC 계층의 매개변수를 조정하는 기법을 제안했다. 노드의 배터리 잔량, E2E 지연, 패킷 손실 비율을 상태 S로 가지고 경쟁 윈도우를 조정, 프레임 우선순

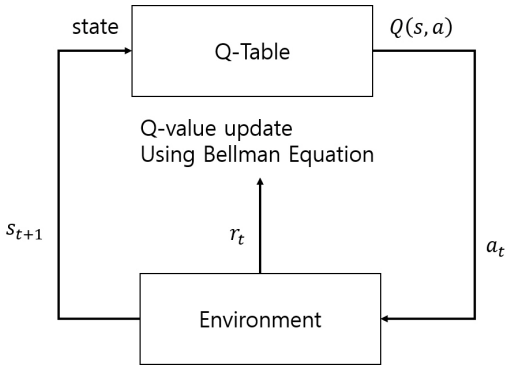

그림 1. 신경망 없는 가치기반 강화학습
Fig. 1. Value-Based Reinforcement Learning Schematic Without Neural Networks

위 조정, 프레임 전송 기회 조정, 슬립/기상 모드 전환을 행동 A로 취한다. 이때 에너지 소비 감소, E2E 지연 감소, 충돌 감소, 패킷 손실 비율 감소의 합을 보상 R로 평가하게 된다. 시뮬레이션 결과 전자 의무 기록 시스템에서 Rank-Based-All-Nodes 알고리즘은 에너지 효율을 최대 26.5% 개선했으며, E2E 지연은 15.87ms, 패킷 손실 비율은 2.4%로 감소시켰다. 심전도와 뇌전도 측정 시스템에서 Rank-Based-Extremum-Nodes 알고리즘은 에너지 효율을 각각 23.5%, 16.2% 증가했다. 그러나 알고리즘이 의료용 IoT 환경에서 최적화 되어있어 다른 환경에 적용 가능한 확장성 문제가 존재할 수 있다.

논문 [7]은 인지 라디오 네트워크 환경에서 단순 강화학습과 SARSA 알고리즘을 사용하여 채널 선택을 최적화하는 기법을 제안했다. 단순 강화학습에서는 모든 채널에 대해 동일한 초기 가중치를 부여하고 채널이 사용 중인지 유휴 상태인지에 따라 가중치를 증감시키고 가장 높은 가중치를 가진 채널을 선택하는 방식으로 학습이 진행된다. SARSA 알고리즘은 모든 상태-행동 쌍에 동일한 초기 Q값을 설정하고 단순 강화학습과 동일하게 채널의 상태에 따라 보상을 부여한다. 보상에 따라 Q값을 업데이트 하는 방식으로 학습을 진행한다. 실험 결과 트래픽 부하가 높아졌을 때 단순 강화학습은 30 Erlangs까지 안정적으로 처리량을 유지할 수 있었고, SARSA는 35 Erlangs까지 유지할 수 있었다. 그리고 SARSA가 단순 강화학습에 비해 더 낮은 충돌률을 기록했으며, 지연 시간 또한 SARSA가 더 낮은 시간을 기록했다. 그러나 단순 강화학습과 SARSA 모두 높은 계산 비용을 가지고 있어 에너지 소비 관련 문제가 발생할 수 있다.

논문 [8]은 차세대 고효율 무선랜 환경에서 Q러닝을 사용하여 백오프 윈도우의 크기를 조정하는 기법인

iQRA 메커니즘을 제안했다. 스테이션의 백오프 단계를 상태 S로 가지며 백오프 단계에 따라 백오프 윈도우의 크기를 증감하는 행동 A를 수행한다. 보상 R은 채널 충돌 확률에 따라 주어진다. 실험 결과 스테이션이 증가하는 밀집 무선 랜 환경에서 iQRA 메커니즘은 BEB 알고리즘은 60~90% 높은 처리량 및 COSB 알고리즘에 비해 13~28% 높은 처리량을 유지했다. iQRA 메커니즘이 COSB보다 약간 높은 채널 접근 지연을 보였지만 네트워크가 밀집될수록 iQRA의 접근 지연이 더 효율적임을 보였다. 그러나 상태공간 S가 백오프 단계로 단순화 되어있기 때문에 다양한 네트워크 환경 요소들에 대한 영향을 반영하지 못한다.

논문 [9]는 무선 메시 네트워크에서 Q러닝을 사용하여 전송속도를 증가, 감소, 유지하는 알고리즘인 RARE를 제안했다. 이 알고리즘에서는 평균 대기열 길이, 전송 성공률과 실패율을 기반으로 한 링크 품질, 연속 성공 및 실패 횟수를 상태 S로 가진다. 상태에 따라 전송속도를 증가, 감소, 유지하는 행동 A를 수행하고, 그에 따라 음과 양의 보상 R을 받게 된다. 실험 결과 RARE 알고리즘은 17개의 노드가 동시에 전송할 때 ONOE, ARF, AARF 보다 약 40~60% 높은 처리량을 보여주었고, 3.5Mbps 부하에서도 ONOE, ARF, AARF 보다 약 50~65% 높은 처리량을 유지했다. 그러나 단순한 보상함수를 가지고 있기 때문에 네트워크에 세부적인 변화에 대응하지 못할 가능성이 있다.

논문 [10]은 무선 센서 네트워크에서 슬로티드 ALOHA방식에 Q러닝을 적용하여 슬롯을 선택하는 방식을 학습하는 ALOHA-Q 기법을 고안했다. 특정 슬롯에서 성공적으로 전송을 했는지 충돌이 발생했는지를 상태 S로 받아서 Q값이 가장 높은 슬롯을 선택하는 행동 A를 취한다. 전송 결과에 따라 양의 보상과 음의 보상을 받으면서 학습을 진행한다. 실험 결과 ALOHA-Q는 슬로티드 ALOHA에 보다 약 3.5배 높은 처리량을 보였고, 높은 트래픽 부하에서도 3초 이하의 낮은 지연을 유지하는 것을 보여주었다. 그러나 슬로티드 ALOHA 기반의 알고리즘이기 때문에 트래픽이 높다면 충돌 가능성이 증가한다.

논문 [11]은 인지 라디오 네트워크 환경에서 기존 CSMA 방식을 Q러닝과 결합한 M-CSMA-RL 기법을 고안했다. M-CSMA-RL에서는 각 채널이 사용 가능한지 사용 중인지를 상태 S로 가진다. 사용 가능한 채널 중 하나를 선택하여 데이터 전송을 시도하는 행동 A를 취하고, 전송이 성공적이면 양의 보상을 받고 전송에 실패하면 음의 보상을 받게 된다. 실험 결과 충돌률이 랜덤 채널 선택 방식과 일반 CSMA방식보다 낮았으며, 데이터 전송 성공률은 랜덤 채널 선택 방식에서는 60%~70% 정도였고, M-CSMA-RL 방식에서는 85%~90%를 보였다. 그러나 Q러닝은 상태-행동 값 계산 및 생신에 많은 계산 리소스를 요구하기 때문에 에너지 소비 관련 문제가 발행할 수 있다.

논문 [12]는 Wi-SUN 환경에서 언슬로티드 CSMA/CA 프로토콜의 성능을 최적화 하기위해 Q러닝을 사용한 기법인 QUC@64를 제안했다. 네트워크 채널의 유휴 시간과 사용시간에 의해 상태 S를 가지고, 백오프 값의 범위에서 백오프 지연 값을 선택하는 행동 A를 취한다. 이때 전송이 성공하면 양의 보상을 받고 실패하거나 채널이 사용중이면 음의 보상을 받는다. 실험 결과 제안된 QUC@64가 기존 방식인 UC@128보다 패킷 전달 비율에서 20% 성능향상을 보였고, 지연시간 또한 20% 낮음을 보여주었다. 그러나 네트워크 규모가 커지거나 다양한 패턴의 트래픽이 발생하면 학습 효율이 떨어질 수 있다.

논문 [13]은 CSMA/CA 기반 무선 네트워크 환경에서 Q러닝을 사용하여 데이터 전송 속도를 제어하는 기법을 제안했다. 경쟁 윈도우의 크기를 기반으로 연속적인 전송 실패 횟수를 계산한 값을 상태 S로 가진다. 행동 A는 변조 및 부호화 방식(MCS) 수준 8개 중에 선택한다. 이때 각 시간 단계에서 성공적으로 수신한 ACK 패킷의 수를 보상으로 받게 된다. 실험 결과 정지 및 비정지 시나리오에서 높은 처리량을 달성하는 것을 보여주었고, 기존 방식인 Minstrel 및 CARA와 비교해서 유사한 성능을 보여주었다. 그러나 Minstrel 및 CARA와 비교했을 때 처리량 변동성이 더 큰 것을 보여주었다. 이는 안정성이 중요한 환경에서는 단점으로 작용할 수 있다.

논문 [14]는 IoT 네트워크에서 Q러닝을 사용하여 경쟁 기간의 길이를 동적으로 조정하는 기법을 제안했다. 경쟁 기간의 길이를 상태 S로 가지고, 경쟁 기간의 길이를 늘리거나 줄이거나 유지하는 행동 A를 취한다. 보상은 경쟁 결과에 따라 받게 되는데 전송에 성공하면 높은 보상을 받고 실패하거나 유휴 슬롯은 낮은 보상을 받게 된다. 실험결과 Q러닝 기반 MAC이 하이브리드 MAC 대비 최대 101.7% 높은 처리량을 보였고, 토큰 기반 MAC 대비 최대 344.3% 높은 처리량을 달성했다. 그리고 하이브리드 MAC와 토큰 기반 MAC 보다 최대 87.3%감소된 종단 간 지연을 달성했다. 그러나 Q러닝 기반 MAC는 실패한 전송의 에너지 소비는 줄이지만, 성공적인 전송의 에너지 소비는 토큰 기반 MAC 보다 높다는 문제가 존재한다.

논문 [15]는 IoT 네트워크에서 Q러닝을 사용하여 경

쟁 윈도우의 크기를 동적으로 조정하는 알고리즘을 제안했다. 현재 경쟁 윈도우의 크기를 상태 S로 가지고, 경쟁 윈도우의 크기를 조정하는 행동 A를 취한다. 만약 처리량이 증가하면 양의 보상을, 감소하면 음의 보상을 받는다. 연구에서는 경쟁 윈도우 조정에 보수적인 모드 1과 공격적인 모드 2라는 2가지의 모드를 제안했다. 실험 결과 모드 2가지 모두 기존 규칙 기반 경쟁 윈도우 조정 방식보다 18~26% 높은 처리량을 기록했고, 모드1과 2의 충돌률이 각각 13.7%, 17.2로 규칙 기반 방법의 22.4%보다 낮은 것을 보였다. 그러나 이 연구에서는 숨겨진 노드를 고려하지 않고 실험을 진행했기 때문에 실제 네트워크에서는 숨겨진 노드 문제가 발생할 수 있다.

논문 [16]은 사물용 IoT 네트워크에서 Q러닝을 사용하여 트래픽 패턴을 학습하는 QMA기법을 제안했다. 현재 슬롯 ID를 상태 S로 가진다. 대기, 채널이 사용가능한지 확인(CCA) 후 판단하여 전송, 확인하지 않고 전송 3가지를 행동 A로 취한다. 전송에 성공하거나 유휴 상태에서 패킷 수신이 확인되면 양의 보상, 전송에 실패한다면 음의 보상을 얻는다. 실험결과 10 packets/s 이상인 숨겨진 노드가 존재하는 상황에서 97%의 높은 패킷 전달률을 달성했다. 그러나 대규모 네트워크에서 슬롯의 수가 증가하는 경우 Q 테이블의 크기가 커져 계산비용이 커질 수 있는 문제가 존재한다.

논문 [17]은 LoRa 무선랜 네트워크에서 Q러닝을 사용하여 주파수 채널 자원을 동적으로 할당하는 기법을 제안했다. 모든 노드에 대한 주파수 할당 상태를 상태 S로 가진다. 노드에 할당할 수 있는 주파수 채널 선택을 행동 A로 취하고, 성공적으로 수신된 패킷의 수와 네트워크 내 다른 노드들의 패킷 수의 비율에 따라 보상을 받게 된다. 실험 결과 제안된 방식이 무작위 자원 할당에 비해 평균 패킷 전달률을 주파수 채널 수가 8일때는 20%이상, 16일때는 13% 개선했다. 또한 충돌률도 소폭 개선한 것을 보였다. 그러나 네트워크의 규모가 커진다면 상태 공간과 행동 공간의 크기가 커지므로 계산비용이 커질 수 있는 문제가 존재한다.

논문 [18]은 무선 네트워크에서 Q러닝을 사용하여 최적의 백오프 넘버를 선택하는 기법을 제안했다. 현재 스테이션이 사용하는 백오프 넘버를상태 S로 가진다. 경쟁 윈도우 범위 내의 가능한 백오프 넘버를 선택하는 행동 A를 취하고 패킷 전송 결과에 따라 양의 보상과 음의 보상을 얻는다. 실험결과 기존 CSMA/CA 방식과 비교했을 때 100 스테이션 환경에서 약 30% 높은 처리량을 보였고, 최적의 백오프 번호를 학습하여 충돌률 또한 감소 되었다. 그러나 단순한 보상 함수를 가지고 있기 때문에 다양한 트래픽 패턴에서의 학습이 어려울 수 있는 문제가 발생할 수 있다.

논문 [19]는 무선 네트워크 환경에서 Q러닝을 사용하여 동적으로 CSMA/CA와 TDMA를 전환하는 SOMAC 알고리즘을 제안했다. 이 알고리즘에서는 현재 네트워크에서 사용중인 MAC 프로토콜을 상태 S로 가진다. MAC 프로토콜을 유지하거나 전환하는 행동 A를 취하고, 네트워크 성능이 향상되었다면 0의 보상을, 성능이 동일하거나 감소했다면 음의 보상을 받게 된다. 실험결과 동적 네트워크에서는 SOMAC이 CSMA/CA에 비해 20% 향상된 처리량을 보였고, 3배 더 낮은 지연 시간을 보였다. 그러나 두가지의 MAC 프로토콜만 고려하여 확장성이 낮다는 한계가 존재한다.

논문 [20]은 무선 센서 네트워크 환경에서 Q러닝을 사용하여 노드의 활성 시간을 동적으로 결정하는 기법인 RL-MAC을 제안했다. 노드의 전송 대기열에 있는 패킷 수를 상태 S로 가지고, 노드가 활성 시간을 결정하는 것을 행동 A로 취한다. 활성 시간 동안 성공적으로 송수신된 패킷 수와 에너지 효율성에 따라 받는 즉각적 보상과 패킷 전송 실패를 패널티로 반영하는 지연 보상을 합친 값을 최종 보상으로 얻는다. 실험 결과 스타 토폴로지에서 RL-MAC이 S-MAC 대비 50% 더 높은 처리량을 달성했고, 최대 50% 높은 에너지 효율성을 달성했다. 지연시간 또한 최대 44.4% 감소한 것을 보였다. 그러나 RL-MAC은 독립적으로 학습하고 협력 메커니즘이 없기 때문에 네트워크의 전역적인 최적화를 달성하기 어렵다는 문제점이 존재할 수 있다.

논문 [33]은 수중 무선 센서 네트워크 환경에서 다음 홉 중계 노드 선택을 최적화하는 LIBRA 알고리즘을 제안했다. 중계 노드의 물리적 위치, 송신 노드와의 거리, 최적 거리의 정보를 상태 S로 받고, 다음 홉 중계 노드 선택을 행동 A로 취한다. 보상은 최적 중계 간 거리, 현재 중계 노드에서 목적지까지의 잔여 거리에 따라 다르게 부여된다. 실험 결과 LIBRA는 최적 에너지 효율의 약 90%를 달성했고, ALOHA와 SFAMA보다 최대 80배 더 높은 처리량을 달성했다. 그러나 보상 함수가 거리를 기반으로 설계되어 특정 중계 노드가 과부하 상태가 될 수 있는 가능성이 존재한다.

표 1에는 신경망 없는 가치 기반 알고리즘을 사용한 연구들에 대한 인공지능 알고리즘, 네트워크 도메인, 장점, 단점이 정리되어 있다. 이러한 연구들의 공통적인 장점은 네트워크 처리량을 향상시키고 지연 시간을 줄이는 데 기여한다는 점이 있다. 그러나 계산 자원 문제와 단순한 상태나 보상 설계 문제라는 단점도 함께 나타

표 1. 신경망 없는 가치 기반 알고리즘 요약
Table 1. Value-Based Algorithm Without Neural Network Summary

Machine Learning Algorithm	Network Domain	Disadvantage	Advantage
MDP[6]	IoT	Scalability	Energy reduction
SARSA[7]	CRN	Energy consumption	Enhanced throughput Reduction collision rate
Q-Learning[8]	Dense WLAN	Simple state space	Enhanced throughput Reduction delay
Q-Learning[9]	WMN	Simple reward	Enhanced throughput
Q-Learning[10]	WSN	Collision possibility	Enhanced throughput
Q-Learning[11]	CRN	High energy consumption	Enhanced throughput
Q-Learning[12]	Wi-SUN	Low learning efficiency	Enhanced transmission
Q-Learning[13]	WLAN	Low stability	Enhanced PDR
Q-Learning[14]	IoT	High success energy	Enhanced throughput Reduction delay
Q-Learning[15]	IoT	Hidden node	Enhanced throughput
Q-Learning[16]	IoT	High Computation Cost	Enhanced PDR
Q-Learning[17]	LoRaWAN	High Computation Cost	Enhanced PDR
Q-Learning[18]	WLAN	Simple Reward Function	Enhanced throughput
Q-Learning[19]	WLAN	Fewer MAC Protocols	Enhanced throughput
Q-Learning[20]	WSN	Network Optimization Difficulties	Enhanced throughput
LIBRA[33]	UWSN	Overburden a specific node	Enhanced throughput

난다. 신경망 없는 가치 기반 알고리즘을 사용한 MAC 프로토콜 연구 또한 주로 저전력 및 저비용 노드로 구성된 네트워크 환경에서 성능을 최적화하고 향상시키는 것을 주요 목표로 하고 있다. 최대한 적은 에너지를 사용하여 최대의 성능 향상을 이루는 것을 목표로 설계되고 있다.

2.1.2 신경망 있는 가치 기반 알고리즘

논문 [21]은 무선 센서 네트워크 환경에서 Q러닝과 DQN을 사용하여 CSMA/CA의 매개변수인 경쟁 윈도우, 최대 백오프 지수, 최대 백오프 횟수를 동적으로 조정하는 기법인 QL-CSMA/CA와 DQN-CSMA/CA를 제안했다. 성공적으로 전송된 패킷 수, 실패한 패킷 수, 충돌 발생 횟수를 상태 S로 가진다. 경쟁 윈도우, 최대 백오프 지수, 최대 백오프 횟수를 선택하는 행동 A를 취한다. 보상은 처리량, 패킷 손실률, 지연 시간에 따라 가중치를 주어 받게 된다. 실험 결과 CSMA/CA 알고리즘과 비교했을 때 DQN-CSMA/CA가 107% 높은 처리량을 기록했고, QL-CSMA/CA 알고리즘보다 7% 높은 처리량을 기록했다. 지연 시간 또한 DQN-CSMA/CA, QL-CSMA/CA순으로 낮은 것을 확인했다. 그러나 보상 함수에 사용하는 가중치의 값이 고정된 값이 아니라 동적으로 변할 수 있다면 더 좋은 성능을 보일 수 있다고 생각된다.

논문 [22]는 OFDMA기반 802.11ax 네트워크 환경에서 Q러닝과 DQN을 사용하여 경쟁 윈도우 값을 동적으로 조정하는 기법인 QL-MAC과 DQN-MAC을 제안했다. QL-MAC에서는 경쟁 윈도우의 현재 값을 상태

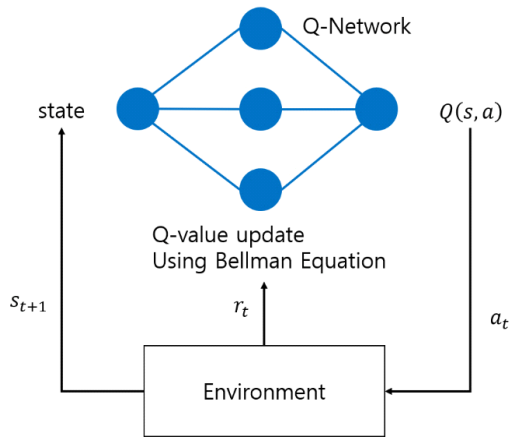

그림 2. 신경망있는 가치기반 강화학습 도식
Fig. 2. Value-Based Reinforcement Learning Schematic With Neural Networks

S로 가지고, 경쟁 윈도우를 증가, 유지, 감소하는 행동 A를 취한다. 보상은 네트워크 처리량을 기반으로 받게 된다. DQN-MAC에서는 네트워크의 활성 스테이션의 수를 상태 S로 받고, 경쟁 윈도우의 최대와 최솟값 사이에서 경쟁 윈도우를 선택하는 행동 A를 취한다. QL-MAC과 동일하게 네트워크 처리량에 비례해 보상을 받게 된다. QL-MAC은 정적 네트워크에서 실험을 진행했는데 적응형 BTM 그룹화 알고리즘보다 9RU 환경에서 14.3%, 18RU 환경에서 10.3% 높은 처리량을 기록했고, 더 낮은 지연 시간 또한 기록했다. DQN-MAC은 동적 네트워크에서 실험을 진행했는데 QL-MAC과 적응형 BTM 그룹화 알고리즘보다 평균 5~10% 높은 처리량을 달성했고, 더 낮은 지연 시간 또한 보였다. 그러나 Q러닝과 DQN 모두 네트워크가 복잡해질수록 계산 비용이 늘어난다는 문제점을 가지고 있다.

논문 [23]은 IoT 네트워크 환경에서 DQN을 사용하여 초기 백오프 윈도우의 크기를 동적으로 선택하는 기법을 제안했다. 현재 채널 계수, 현재 에너지 수준, 이전 주기 동안 수집된 에너지 3가지를 상태 S로 가진다. 초기 백오프 윈도우 크기를 5개 중에 선택하는 행동 A를 취한다. 에너지 부족 이벤트와 전송된 패킷 수에 따라 양의 보상과 0의 보상을 얻게 된다. 실험결과 DQN 기반 알고리즘은 전통적인 802.11 방식보다 약 44.4%, EIED,/LILD 알고리즘보다 약 30% 처리량이 높고 에너지 부족 확률이 낮은 것을 확인할 수 있다. 그러나 초기 2시간동안 성능이 낮다는 문제점이 존재한다.

논문 [24]는 이기종 무선 네트워크 환경에서 DQN을 사용하여 에이전트가 각 시간 슬롯에서 수행할 행동을 선택하는 기법인 DLMA를 제안하였다. 과거부터의 에이전트의 행동과 채널 상태 쌍을 상태 S로 가지고, 특정 시간 슬롯에서 전송을 시도할지 대기할지 행동 A를 취한다. 전송이 성공적이라면 양의 보상을 받고, 그렇지 않다면 0의 보상을 받는다. 실험 결과 TDMA 노드와 q-ALOHA 노드가 공존하는 환경에서 TDMA 노드가 점유하지 않은 슬롯과 q-ALOHA 노드의 전송 확률을 고려하여 최적에 근접한 약 98%의 총 처리량을 달성했다. 그러나 CSMA/CA와 같은 복잡한 프로토콜이 공존하는 환경에 대해 고려하지 않아 더 다양한 프로토콜이 존재하는 환경에서의 실험이 필요할 것으로 보인다.

논문 [25]는 무선네트워크 환경에서 DQN을 사용하여 MAC 프로토콜의 매개변수를 동적으로 선택하는 기법을 제안했다. MAC 계층의 수신 전력, 재전송 빈도, 현재 매개변수, 채널 사용률, 애플리케이션 계층의 버퍼 크기 및 시간, 비트레이드 이력, QoE 이력을 상태 S로 받는다. MAC 계층의 매개변수 설정과 애플리케이션 계층의 노드의 비트레이트 선택 범위와 품질 변화 허용도를 행동 A로 취한다. MAC 계층의 네트워크 평균 처리량과 애플리케이션 계층의 QoE에 따라 보상을 받게 된다. 실험 결과 제안된 방법이 노드의 수가 19~25개인 환경에서 CSMA/CA보다 평균 처리량이 최대 14.7% 향상 된 것을 확인했다. 그러나 중앙 집중식 게이트웨이에서 모든 학습과 제어를 수행하는데 게이트웨이 과부하가 발생할 수 있다는 문제가 존재한다.

논문 [26]은 이기종 무선 네트워크 환경에서 DQN을 사용하여 패킷 전송 여부를 결정하는 기법인 CS-DLMA를 제안했다. 과거 채널 상태와 동작 쌍을 상태 S로 가지고, 패킷을 전송할지 채널 상태를 감지할지를 선택하는 행동 A를 취한다. 전송에 성공하면 양의 보상을, 충돌이 발생하면 0의 보상을 얻는다. 실험 결과 WiFi와 CS-DLMA의 성능을 비교했을 때 CS-DLMA가 최적 처리량에 99% 수준의 성능을 달성했으며, TDMA와 ALOHA 노드 개별 처리량 또한 WiFi 대비 각각 10~15% 증가했다. 그러나 에너지 소비나 지연과 같은 다른 요소는 고려하지 않고 전송 성공 여부와 패킷 길이에만 의존하는 단순한 보상 함수를 가지고 있다.

논문 [27]은 무선 네트워크 환경에서 DQN을 사용하여 최적의 프로토콜 블록을 선택하는 DeepMAC 프레임워크를 제안했다. 블록 조합과 네트워크 평균 처리량의 이력 데이터, 각 블록의 활성화 여부를 상태 S로 가진다. 다음 상태로 가기 위한 블록 조합 선택을 행동 A로 취한다. 처리량과 소비 에너지를 기반으로 보상을 받게 된다. 실험 결과 DeepMAC이 CSMA/CA보다 노드 수가 5~10개인 저 부하 네트워크에서는 30%이상의 처리량 증가, 노드의 수가 20~50개인 고 부하 네트워크에서는 최대 50%이상의 처리량을 보여주었다. 그러나 모든 노드가 슈퍼 노드의 제어를 받아야 하는데 이는 특정 노드에서 공정성 문제가 발생할 수 있다고 생각된다.

논문 [28]은 다중 채널 이종 네트워크 환경에서 DQN을 사용하여 특정 채널로의 전송 여부를 결정하는 MC-DLMA를 제안했다. 각 채널에서의 관측 값(전송 성공, 채널 비어 있음, 충돌 발생)을 상태 S로 가진다. 특정 채널로 전송을 하거나 대기하는 행동 A를 취하고, 전송이 성공했다면 1의 보상을 받고 아니라면 0의 보상을 받는다. 단일 MC-DLMA 노드와 TDMA와 q-ALOHA 노드와 공존하는 실험에서는 랜덤 접근 방식 대비 40%이상의 처리량 향상이 있었고, TDMA와 FW-ALOHA 노드와 공존하는 환경에서는 42% 이상의 처리량 향상이 있었다. 다중 MC-DLMA 노드,

q-ALOHA, FW-ALOHA가 공존하는 환경에서는 52% 이상의 처리량 향상이 있었다. 그러나 DQN은 신경망을 사용하여 계산 비용이 높기 때문에 에너지 소비 문제가 발생할 수 있다.

논문 [29]는 무선 네트워크 환경에서 DQN과 연합 학습을 같이 사용하여 전송 여부를 결정하는 FRMA 기법을 제안했다. 시간 슬롯 단위로 채널 상태와 과거 행동 데이터를 상태 S로 가진다. 데이터 전송을 시도하거나 대기하는 행동 A를 취한다. 전송에 성공하면 양의 보상을 얻고, 실패하면 음의 보상을 얻는다. 일정 주기마다 각 스테이션에서 학습한 가중치를 중앙 스테이션으로 모아 평균 가중치를 생성한 후 다시 각 스테이션으로 전송한다. 실험 결과 FRMA는 RTS/CTS 보다 5%, DCF Basic 보다 20% 높은 처리량을 달성했고, 공정성 또한 높은 것을 보였다. 그러나 각 스테이션에 DQN을 적용하여 주기적으로 연합 학습을 진행해야 하기 때문에 계산 비용이 높다는 문제가 존재한다.

논문 [30]은 IoT용 무선 애드혹 네트워크 환경에서 DQN과 DDQN을 사용하여 데이터 전송 여부를 결정하는 방향성 MAC 프로토콜인 DDMAC, DDDMAC를 제안했다. 특정 시간 슬롯에서 노드의 패킷 대기열, 채널 상태, 재전송 횟수를 상태 S로 가지고 데이터를 전송하거나 채널 상태를 확인하는 행동 A를 취한다. 데이터 전송에 성공했다면 양의 보상을, 실패한다면 음의 보상을 얻는다. 실험 결과 DDDMAC는 CSMA와 AL-DMAC에 비해 각각 최대 54.1%, 72.7% 높은 처리량을 기록했고 58.2%, 58.1% 낮은 지연 시간을 기록했다. 그러나 노드의 이동성이 높은 네트워크에서는 학습이 어렵다는 문제가 존재한다.

논문 [31]은 클라우드 라디오 접속 네트워크 환경에서 DDQN을 사용하여 RRH의 온/오프 상태를 결정하는 기법을 제안했다. 사용자의 데이터 요구 속도, RRH의 상태, UE와 RRH 사이 채널 상태를 상태 S로 가지고, RRH를 켜거나 끄는 행동 A를 취한다. UE의 처리량, 전송 대역폭, RRH의 소비 전력을 기반으로 보상을 얻는다. 실험 결과 기존 FA 방식 대비 DDQN 방식이 최대 22% 에너지 소비 감소를 달성했고, DQN 대비 13% 절감을 보였다. 그러나 보상 함수의 설계가 에너지 효율만을 고려하고 처리량이나 지연과 같은 항목을 사용하지 않아 처리량과 에너지 소비 관계에 대해서 고려하지 않았다는 문제점이 존재한다.

논문 [32]는 IoT 네트워크 환경에서 DDQN을 사용하여 어떤 백스캐터 장치를 서비스할지 결정하는 DRL-MAC를 제안했다. 현재 선택한 백스캐터 장치, 전송할 데이터의 크기와 지속 시간, eACK 신호 수신 여부를 상태 S로 가지고, 백스캐터 장치를 선택하는 행동 A를 취한다. eACK 신호를 성공적으로 수신하면 1의 보상, 수신하지 못하거나 충돌이 발생하면 0의 보상

표 2. 신경망 있는 가치 기반 알고리즘 요약
Table 2. Value-Based Algorithm with Neural Network Summary

Machine Learning Algorithm	Network Domain	Disadvantage	Advantage
Q-Learning, DQN[21]	WSN	Fixed Reward Function weight	Enhanced throughput
Q-Learning, DQN[22]	Wi-Fi 6	High Computation Cost	Enhanced throughput
DQN[23]	IoT	Low initial performance	Enhanced throughput Energy efficiency
DQN[24]	Heterogeneous Wireless Networks	Simple MAC Protocols	Enhanced throughput
DQN[25]	WLAN	Gateway overload	Enhanced throughput
DQN[26]	Heterogeneous Wireless Networks	Simple Reward Function	Enhanced throughput
DQN[27]	WLAN	Low fairness	Enhanced throughput
DQN[28]	Multi-channel Heterogeneous Network	High energy consumption	Enhanced throughput
DQN, FL[29]	WLAN	High Computation Cost	Enhanced throughput
DQN, DDQN[30]	IoT	Hard mobility node learning	Enhanced throughput
DDQN[31]	CRAMN	Simple Reward Function	Energy reduction
DDQN[32]	IoT	High Computation Cost	Enhanced throughput Energy efficiency

을 얻는다. 실험 결과 DRL-MAC은 슬로티트 AHOLA 방식에 비해 30~35% 더 높은 채널 활용도를 달성했다. 그리고 WiFi 통신에 비해 최대 4배 높은 처리량 또한 달성했다. 그러나 백스캐터 장치의 수와 예약 단계가 증가하면 상태 및 행동 공간의 수가 기하급수적으로 증가하여 계산 비용이 커진다는 문제점이 발생할 수 있다.

표 2에는 신경망을 사용하는 가치 기반 알고리즘을 활용한 연구에서 적용된 인공지능 알고리즘, 네트워크 도메인, 장점, 단점이 정리되어 있다. 이러한 연구들의 공통된 장점은 신경망을 활용함으로써 더욱 복잡한 계산을 수행할 수 있다는 점이며, 이를 통해 네트워크 처리량을 향상시키고 지연 시간을 줄이는 데 기여하고 있다. 그러나 신경망 기반 학습은 높은 계산 비용이 요구된다는 공통적인 한계를 가지며, 이는 에너지 소비 증가로 이어져 제원 제약이 존재하는 네트워크 환경에서 성능 저하를 초래할 수 있다. 이러한 문제를 해결하기 위해 많은 연구들이 계산 비용과 에너지 소비를 줄여 자원 제약이 있는 네트워크에서도 효율적으로 성능을 향상시킬 수 있는 방안을 모색하고 있다.

2.2 정책 기반 알고리즘

논문 [34]는 무선 네트워크 환경에서 MAB를 사용하여 TDMA 슬롯 할당, CSMA 접근을 선택하는 기법을 제안했다. TDMA 정보와 CSMA 단계에서의 관찰 내용을 상태 S로 가진다. TDMA 슬롯 할당, CSMA 접근을 행동 A로 취하고, 네트워크 처리량과 슬라이스 간 격리에 따라 보상을 받게 된다. 실험 결과 제안된 방법은 기존 CSMA 기반 방식 대비 10~20% 높은 처리량을 보였다. 그러나 새로운 프로토콜이 아니라 프로토콜의 전환이기 때문에 기존 프로토콜의 단점을 가지고 있다는 문제점이 존재한다.

논문 [35]는 무선 네트워크 환경에서 POMDP 기반으로 모델링하여 대역폭-채널 구성을 선택하는 PoBA 알고리즘을 고안했다. 대역폭-채널 구성과 간섭 여부를 상태 S로 가지고 가능한 대역폭-채널 구성 중에서 선택하는 행동 A를 취한다. 대역폭과 처리량에 따라 보상을 받게 된다. 실험 결과 체인형 토폴로지와 랜덤 토폴로지 환경에서 중앙 집중식 채널 할당 방식보다 PoBA 알고리즘이 10%-30% 더 높은 처리량을 달성했다. 그러나 채널 수가 많아지면 상태 공간이 기하급수적으로 커져 계산 비용이 증가한다는 문제점이 존재한다.

논문 [36]은 IoT 네트워크 환경에서 REINFORCE를 사용하여 보조 송신기의 전송을 제어하는 기법인 RL 기반 DSA 시스템을 제안했다. 각 보조 송신기에서 수집된 총 간섭 값을 상태 S로 받고, 보조 송신기에서 전송을 제어하는 행동 A를 취한다. 보상은 영역 스펙트럼 효율성을 기반으로 주어진다. 실험 결과 RL기반 DSA 시스템이 CSMA/CA 보다 평균 영역 스펙트럼 효율성이 2배 이상 우수한 성능을 보였고 더 나은 적응성과 유연성을 제공하는 것을 보였다. 그러나 학습 시간이 길어질수록 기본 사용자 네트워크의 성능이 감소하는 문제점이 존재한다.

논문 [37]은 무선 네트워크 환경에서 PPO를 사용하여 시퀀스에서 두 개의 비트를 선택하여 전환하는 TMUI 프로토콜 시퀀스를 제안했다. 시퀀스 집합과 어느 시퀀스에서 비트가 전환될 지 지정하는 측면 지표를 상태 S로 가지고, 두 개의 비트를 선택하여 비트를 전환하는 행동 A를 취한다. 최적의 시퀀스 집합이 발견되면 1의 보상을, 그렇지 않다면 목표 처리량 달성횟수와 목표 처리량 최댓값에 따른 보상을 얻는다. 실험 결과 TMUI 프로토콜 시퀀스가 슬로티트 ALOHA에 비해 처리량 분산이 0으로 사용자 간 공정성을 보장하고 모든 슬롯에서 0보다 큰 처리량을 제공하는 것을 보였다. 그러나 실험 환경이 두 사용자 환경으로 진행되어 다수 사용자 환경에서도 성능을 낼 수 있는지 검증이 필요하다.

표 3에는 정책 기반 알고리즘을 사용한 연구들에 대한 인공지능 알고리즘, 네트워크 도메인, 장점, 단점이 정리되어 있다. 이러한 연구들의 공통된 장점은 네트워크 처리량을 향상시켰다는 기여가 존재한다. 그러나 대부분의 연구에서 에너지 소비와 관련된 실험 결과가 부족하여 에너지 소비가 전체적인 네트워크 성능에 미치는 영향을 명확히 알 수 없다는 한계가 있다. 따라서

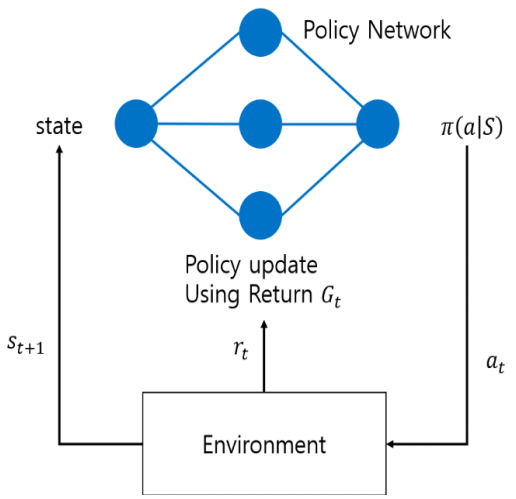

그림 3. 정책기반 강화학습 도식
Fig. 3. Policy-Based Reinforcement Learning Schematic

표 3. 정책 기반 알고리즘 요약
Table 3. Policy-based algorithm summary

Machine Learning Algorithm	Network Domain	Disadvantage	Advantage
MAB[34]	WLAN	Disadvantages of existing protocols	Enhanced throughput
POMDP[35]	WLAN	High Computation Cost	Enhanced throughput
REINFORCE[36]	IoT	Less Performance as learning gets longer	Enhanced adaptability Enhanced scalability
PPO[37]	WLAN	Small experimental setting	Enhanced throughput

향후 연구에서는 에너지 소비 관련 시험 내용을 추가하여 네트워크 성능을 보다 다방면으로 평가할 필요성이 보인다.

2.3 다중 에이전트 강화학습

논문 [38]은 IoT 네트워크 환경에서 Hysteretic Q러닝을 사용하여 패킷 전송 확률을 동적으로 조정하는 기법인 DRLI-MAC를 제안했다. 노드가 경험한 패킷 충돌 확률을 [0, 1] 범위에서 24개의 이산화 된 값을 상태 S로 가진다. [0, 1] 범위에서 20개의 이산화 된 값을 행동 A로 취한다. 각 노드의 자체 처리량 및 이웃 노드의 처리량을 기반으로 양과 음의 보상을 얻는다. 실험 결과 5, 8, 12개의 노드를 가진 네트워크에서는 DRLI-MAC는 ALOHA보다 30~35% 더 높은 처리량을 보였고, DRLI-MAC은 ALOHA 대비 노드 간 처리량 편차를 최대 50%를 줄여 공정한 자원 할당을 보장했다. 그러나 네트워크 밀도가 증가했을 때 DRLI-MAC의 성능이 감소했다는 한계가 존재한다. 추후 연구에서 상태 없는 Q러닝을 사용하여 해결할 것이라고 말했다

논문 [39]는 저비용 트랜시버를 사용하는 IoT 및 무선 센서 네트워크에서 환경에서 Hysteretic Q러닝을 사용하여 전송 확률을 동적으로 조정하는 기법인 DMLLI를 제안했다. 노드의 충돌 확률을 [0, 1] 범위에서 5개 구간으로 나눈 값을 상태 S로 가진다. 고정된 수로 분할된 전송 확률을 행동 A로 취한다. 처리량 변화와 공정성 변화 모두 임계값보다 큰 경우 50의 보상을 받고, 그렇지 않은 경우 -50의 보상을 얻게 된다. 실험 결과 노드의 수가 2~25개인 네트워크 환경에서 DMLLI는 ALOHA보다 55~62% 낮은 충돌률을 보였고, 37~79% 높은 공정성을 보였다. 그러나 네트워크의 크기가 커질수록 계산 및 통신 오버헤드가 증가하여 IoT 환경에서 에너지 소비 문제가 발생할 수 있다.

논문 [40]은 무선 센서 네트워크 및 TSCH 기반의 IoT 네트워크 환경에서 다중 에이전트 Q러닝을 사용하여 전송 슬롯을 동적으로 선택하는 기법인 QL-TSCH를 제안했다. 슬롯의 전송 성공 또는 실패 정보를 상태 S로 가지고, 전송 슬롯을 선택하는 행동 A를 취한다. 전송에 성공하면 양의 보상, 실패하면 음의 보상을 얻게 된다. 실험 결과 QL-TSCH는 다중 홉 네트워크와 고트래픽 환경에서 97%의 높은 패킷 전달률을 유지했고, 낮은 지연 시간을 보였다. 그러나 IoT 네트워크를 환경으로 가지고 있지만, 에너지 소비 관련 내용이 없다는 점이 아쉬운 점이다.

논문 [41]은 차세대 무선 네트워크 환경에서 QMIX를 사용하여 전송 여부를 결정하는 기법인 Q-LBT를 제안했다. 과거 모든 에이전트의 행동과 각 에이전트의 D2LT를 정규화한 값을 글로벌 상태로 가지고 에이전트의 전송 여부, 에이전트 i의 D2LT, 에이전트 i를 제외한 모든 에이전트들의 D2LT를 관찰 값으로 가진다. 현재 슬롯에서 데이터 전송 여부를 행동 A로 가진다. 전송에 성공하면 양의 보상을 받고, 충돌이 발생하면 음의 보상, 그 외의 경우에는 0의 보상을 받게 된다. 실험 결과 Q-LBT가 기존 WiFi 프로토콜 대비 27% 높은

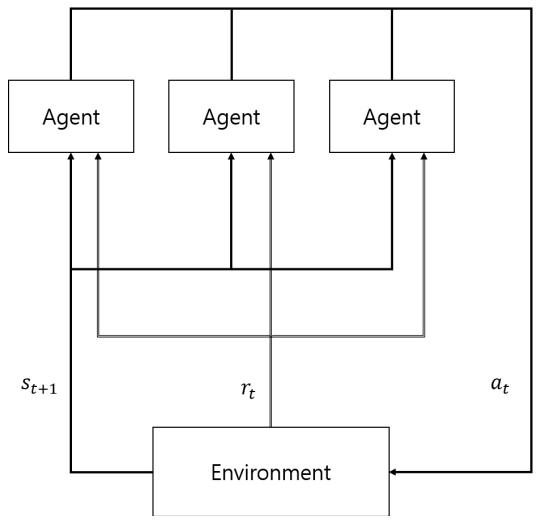

그림 4. 다중 에이전트 강화학습 도식
Fig. 4. Multi-agent Reinforcement Learning Schematic

처리량 및 43% 낮은 지연 시간, CSMA/CA 기반 접근 방식 보다 12% 높은 처리량 및 35% 낮은 지연 시간을 달성했다. 그러나 에너지 효율성에 대한 고려 부족이라는 문제점이 있다.

논문 [42]는 무선 네트워크 환경에서 다중 에이전트 PPO를 사용하여 MAC 프로토콜의 매개변수를 조정하는 기법인 DTDE를 제안했다. 현재 MAC 프로토콜 동작, 가시 노드 수, 수신 신호 강도, 처리량, 지연, 에어타임, 트래픽 특성을 상태 S로 가진다. 백오프 함수 유형, 백오프 크기 및 시간, 감지 슬롯 크기, 에너지 감지 임계값, 대 채널 점유 시간, 전송 전력, 변조 및 코딩 방식을 행동 A로 취한다. 처리량과 트래픽 도착률, 에어타임에 따라 보상을 다르게 받는다. 실험 결과 고밀도 트래픽 환경에서 DTDE는 표준 NR-U 대비 약 33% 높은 처리량을 기록했고, 지연 시간 또한 최대 35.5% 감소했다. 그러나 에이전트가 독립적으로 신경망을 학습하기 때문에 계산 비용이 급격히 증가할 수 있는 가능성이 있다.

논문 [43]은 무선 네트워크 환경에서 다중 에이전트 PPO를 사용하여 패킷 전송 여부를 동적으로 결정하는 기법인 MAPPO-MAC을 제안했다. 스테이션의 관측값, 현재 행동이 지속된 슬롯 수, 스테이션의 지수 가중 평균 처리량과 나머지 스테이션들의 지수 가중 평균 처리량을 로컬 상태로 가지고 모든 스테이션의 행동, 네트워크 전체 관측값, 지속 슬롯 수, 모든 스테이션의 지수 가중 평균 처리량을 글로벌 상태로 가진다. 패킷 전송 여부를 결정하는 행동 A를 취한다. 스테이션이 최적 행동을 선택했다면 1의 보상을 받고, 그 외의 경우는 -1의 보상을 받게 된다. 실험 결과 MAPPO-MAC은 CSMA/CA와 IPPO-MAC보다 노드 수에 따라 최대 2배 높은 처리량을 달성했다. 그러나 중앙 집중식 학습을 기반으로 하기 때문에 AP가 모든 스테이션으로부터 데이터를 수집하고 학습하는 과정에서 계산 부담이 커질 수 있다는 문제점이 있다.

논문 [44]는 무선 네트워크 환경에서 MADDPG를 사용하여 데이터 전송 여부를 동적으로 결정하는 기법을 제안했다. 사용자 장비의 상태는 버퍼 상태와 이전 메시지 및 행동 이력으로 구성되고, 기지국의 상태는 업링크 채널 상태와 수신한 메시지 이력으로 구성된다. 사용자 장비의 행동은 데이터를 전송하거나 대기, 삭제 중 하나를 선택하고 기지국의 행동은 각 사용자 장비에게 보낼 제어 메시지를 선택한다. 데이터가 성공적으로 수신된 경우 양의 보상을 받고, 데이터를 불필요하게 삭제한 경우에는 음의 보상을 받는다. 아무런 데이터도 전송되지 않으면 -1의 보상을 받는다. 실험 결과 시간당 성공적으로 전송된 데이터의 개수가 제안된 MADDPG를 사용한 알고리즘이 Contention-Free와 DDPG 방식보다 각각 16.6%, 31.25% 더 높은 값을 달성했다. 그러나 널리 사용되는 CSMA/CA, TDMA, FDMA와 같은 프로토콜과 비교가 이루어지지 않아서 실질적인 성능 개선 정도를 비교하기 어렵다.

표 4에는 다중 에이전트 강화학습 알고리즘을 사용한 연구들에 대한 인공지능 알고리즘, 네트워크 도메인, 장점, 단점이 정리되어 있다. 이러한 연구의 전체적인 장점으로는 네트워크 처리량을 향상시키고, 지연 시간을 줄이고, 공정성을 보장한다는 점이 있다. 그러나 실

표 4. 에이전트 강화학습 알고리즘 요약
Table 4. Multi-agent Reinforcement Learning Algorithm Summary

Machine Learning Algorithm	Network Domain	Disadvantage	Advantage
Hysteretic Q-learning[38]	IoT	Performance decreases with increasing density	Enhanced throughput Enhanced fairness
Hysteretic Q-learning[39]	IoT, WSN	Energy consumption	Enhanced fairness Reduce collision
MARL(QL)[40]	IoT, WSN	Not considered energy	Enhanced PDR Reduce delay
Q-Mix[41]	NGWN	Not considered energy	Enhanced throughput Reduce delay
MAPPO[42]	WLAN	High Computation Cost	Enhanced throughput Reduce delay
MAPPO[43]	WLAN	High Computation Cost	Enhanced throughput Enhanced fairness
MADDPG[44]	WLAN	Fewer MAC Protocols	Enhanced Goodput

험 결과에 에너지 소비를 고려하지 않고, 계산 비용이 높다는 단점이 존재한다. 무선 네트워크에서 에너지 소비와 계산 비용 문제는 성능에 직접적인 영향을 미치는 요소이기 때문에 향후 실험에서는 두 가지 문제를 고려하여 실험을 진행하는 것이 필요해 보인다.

Ⅲ. 결 론

본 논문에서는 MAC 프로토콜에 강화학습을 사용하여 네트워크 자원을 최적화하거나 MAC 프로토콜을 전환하여 전체적인 네트워크의 성능을 향상시킨 연구들에 대해 살펴보았다. 가치 기반 학습과 정책 기반 학습 그리고 다중 에이전트 강화학습으로 세분화하여 살펴보았다. 대부분의 연구들이 저전력 및 저비용 노드를 사용하는 환경에서 네트워크 처리량, 지연 시간 측면에서 우수한 성능을 보였다. 그러나 공통적으로 계산 비용의 증가와 노드의 에너지 소비라는 문제점이 존재했다. 특히, 저전력 및 저비용 노드를 사용하는 네트워크 환경에서는 이러한 두가지 문제점이 성능에 큰 영향을 미칠 수 있다. 따라서 향후 연구에서는 계산 비용을 줄이고 에너지 소비를 최소화하는 방안을 모색함으로써 자원 제약이 있는 네트워크 환경에서도 효율적인 성능 향상을 달성할 수 있는 새로운 접근 방법을 개발하는 것이 필요하다고 보인다.

References

[1] C. Kai, S. Zhang and L. Wang, “Impacts of packet collisions on link throughput in CSMA wireless networks,” in *China Commun.*, vol. 15, no. 3, pp. 1-14, Mar. 2018. (https://doi.org/10.1109/CC.2018.8331987)

[2] G. Wang, Y. Shu, L. Zhang, and O. W. W. Yang, “Delay analysis of the IEEE 802.11 DCF,” *14th IEEE Proc. PIMRC 2003*, vol. 2, pp. 1737-1741, Beijing, China, 2003. (https://doi.org/10.1109/PIMRC.2003.1260412)

[3] T.-S. Ho and K.-C. Chen, “Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol,” in *Proc. PIMRC ’96 - 7th Int. Symp. Personal, Indoor, and Mobile Commun.*, vol. 2, pp. 407-411, Taipei, Taiwan, 1996. (https://doi.org/10.1109/PIMRC.1996.567426)

[4] S. Lu, V. Bharghavan, and R. Srikant, “Fair scheduling in wireless packet networks,” in *IEEE/ACM Trans. Netw.*, vol. 7, no. 4, pp. 473-489, Aug. 1999. (https://doi.org/10.1109/90.793003)

[5] A. Maatouk, M. Assaad, and A. Ephremides, “Energy efficient and throughput optimal CSMA scheme,” in *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 316-329, Feb. 2019. (https://doi.org/10.1109/TNET.2019.2891018)

[6] G. Famitafreshi, M. S. Afaqui, and J. Melià-Seguí, “Introducing reinforcement learning in the wi-fi MAC layer to support sustainable communications in e-Health scenarios,” in *IEEE Access*, vol. 11, pp. 126705-126723, 2023. (https://doi.org/10.1109/ACCESS.2023.3331950)

[7] Y. Tang, D. Grace, T. Clarke, and J. Wei, “Multichannel non-persistent CSMA MAC schemes with reinforcement learning for cognitive radio networks,” *2011 11th ISCIT*, pp. 502-506, Hangzhou, China, 2011. (https://doi.org/10.1109/ISCIT.2011.6092159)

[8] R. Ali, N. Shahin, Y. B. Zikria, B.-S. Kim, and S. W. Kim, “Deep reinforcement learning paradigm for performance optimization of channel observation‐based MAC protocols in dense WLANs,” in *IEEE Access*, vol. 7, pp. 3500-3511, 2019. (https://doi.org/10.1109/ACCESS.2018.2886216)

[9] A. Al-Saadi, R. Setchi, Y. Hicks, and S. M. Allen, “Multi-rate medium access protocol based on reinforcement learning,” *2014 IEEE Int. Conf. SMC*, pp. 2875-2880, San Diego, CA, USA, 2014. (https://doi.org/10.1109/SMC.2014.6974366)

[10] Y. Chu, S. Kosunalp, P. D. Mitchell, D. Grace, and T. Clarke, Application of reinforcement learning to medium access control for wireless sensor networks, *Eng. Appl. Artificial Intell.*, vol. 46, Part A, pp. 23-32, 2015. (https://doi.org/10.1016/j.engappai.2015.08.004)

[11] H. Li, D. Grace, and P. D. Mitchell, "Collision reduction in cognitive radio using multichannel 1-persistent CSMA combined with reinforcement learning," *2010 Proc. Fifth Int. Conf. Cognitive Radio Oriented Wireless Netw. and Commun.*, pp. 1-5, Cannes, France, 2010. (http://dx.doi.org/10.4108/ICST.CROWNCOM2010.9294)

[12] S. Lee and S.-H. Chung, "Unslotted CSMA/CA mechanism with reinforcement learning of Wi-SUN MAC layer," *2022 Thirteenth ICUFN*, pp. 202-204, Barcelona, Spain, 2022. (https://doi.org/10.1109/ICUFN55119.2022.9829651)

[13] S. Cho, "Rate adaptation with q-learning in csma/ca wireless networks," *J. Inf. Process. Syst.*, vol. 16, no. 5, pp. 1048-1063, 2020. (https://doi.org/10.3745/JIPS.03.0148)

[15] C.-M. Wu, Y.-C. Kao, K.-F. Chang, C.-T. Tsai, and C.-C. Hou, "A q-learning-based adaptive MAC protocol for internet of things networks," in *IEEE Access*, vol. 9, pp. 128905-128918, 2021. (https://doi.org/10.1109/ACCESS.2021.3103718)

[15] Y.-W. Chen and K.-C. Kao, "Study of contention window adjustment for CSMA/CA by using machine learning," *2021 22nd APNOMS*, pp. 206-209, Tainan, Taiwan, 2021. (https://doi.org/10.23919/APNOMS52696.2021.9562498)

[16] F. Meyer and V. Turau, "QMA: A resource-efficient, q-learning-based multiple access scheme for the IIoT," *2021 IEEE 41st ICDCS*, pp. 864-874, DC, USA, 2021. (https://doi.org/10.1109/ICDCS51616.2021.00087)

[17] N. Aihara, K. Adachi, O. Takyu, M. Ohta, and T. Fujii, "Q-learning aided resource allocation and environment recognition in LoRaWAN with CSMA/CA," in *IEEE Access*, vol. 7, pp. 152126-152137, 2019. (https://doi.org/10.1109/ACCESS.2019.2948111)

[18] T.-W. Kim and G.-H. Hwang "Performance enhancement of CSMA/CA MAC protocol based on reinforcement learning," *J. Inf. and Commun. Convergence Eng.*, vol. 19, no. 1, 2021. (https://doi.org/10.6109/jicce.2021.19.1.1)

[19] A. Gomes, D. F. Macedo, and L. F. M. Vieira, "Automatic MAC protocol selection in wireless networks based on reinforcement learning," *Computer Commun.*, vol. 149, pp. 312-323, 2020. (https://doi.org/10.1016/j.comcom.2019.10.023)

[20] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," *Int. J. Sensor. Netw.*, vol. 1, no. 3/4, pp. 117-124, Sep. 2006. (https://doi.org/10.1109/ICNSC.2006.1673243)

[21] J. Lei, D. Tan, X. Ma, and Y. Wang, "Reinforcement learning based multiparameter joint optimization in dense multi-hop wireless networks," *Ad Hoc Netw.*, vol. 154, no. 103357, Mar. 2024. (https://doi.org/10.1016/j.adhoc.2023.103357)

[22] J. Lei, L. Li, and Y. Wang, "QoS-oriented media access control using reinforcement learning for next-generation WLANs," *Comput. Netw.*, vol. 219, p. 109426, 2022.

[23] Y. Zhao, J. Hu, K. Yang, and S. Cui, "Deep reinforcement learning aided intelligent access control in energy harvesting based WLAN," in *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14078-14082, Nov. 2020. (https://doi.org/10.1016/j.comnet.2022.109426)

[24] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," in *IEEE J. Sel. Areas in Commun.*, vol. 37, no. 6, pp. 1277-1290, Jun. 2019. (https://doi.org/10.1109/JSAC.2019.2904329)

[25] K. Aruga and T. Fujii, "MAC protocol adaptation method in coordination with application," *2024 ICAIIC*, pp. 463-466, Osaka, Japan, 2024.

(https://doi.org/10.1109/ICAIIC60209.2024.10463518)

[26] Y. Yu, S. C. Liew, and T. Wang, "Carrier-sense multiple access for heterogeneous wireless networks using deep reinforcement learning," *2019 IEEE WCNCW*, pp. 1-7, Marrakech, Morocco, 2019. (https://doi.org/10.1109/WCNCW.2019.8902705)

[27] H. B. Pasandi and T. Nadeem, "MAC protocol design optimization using deep learning," *2020 ICAIIC*, pp. 709-715, Fukuoka, Japan, 2020. (https://doi.org/10.1109/ICAIIC48513.2020.9065254)

[28] X. Ye, Y. Yu, and L. Fu, "MAC protocol for mMulti-channel heterogeneous networks based on deep reinforcement learning," *GLOBECOM 2020*, pp. 1-6, Taipei, Taiwan, 2020. (https://doi.org/10.1109/GLOBECOM42002.2020.9347989)

[29] L. Zhang, H. Yin, Z. Zhou, S. Roy, and Y. Sun, "Enhancing wifi multiple access performance with federated deep reinforcement learning," *2020 IEEE 92nd VTC2020-Fall*, pp. 1-6, Victoria, BC, Canada, 2020. (https://doi.org/10.1109/VTC2020-Fall49728.2020.9348485)

[30] N. Kim, W. Na, D. S. Lakew, N.-N. Dao, and S. Cho, "DQN-based directional MAC protocol in wireless ad hoc network in internet of things," in *IEEE Int. Things J.*, vol. 11, no. 7, pp. 12918-12928, Apr. 2024. (https://doi.org/10.1109/JIOT.2023.3338562)

[31] A. Iqbal, M.-L. Tham, and Y. C. Chang, "Double deep Q-network-based energy-efficient resource allocation in cloud radio access network," in *IEEE Access*, vol. 9, pp. 20440-20449, 2021. (https://doi.org/10.1109/ACCESS.2021.3054909)

[32] X. Cao, Z. Song, B. Yang, X. Du, L. Qian, and Z. Han, "Deep reinforcement learning MAC for backscatter communications relying on wi-fi architecture," *2019 IEEE GLOBECOM*, pp. 1-6, Waikoloa, HI, USA, 2019. (https://doi.org/10.1109/GLOBECOM38437.2019.9013445)

[33] D. Dugaev, Z. Peng, Y. Luo, and L. Pu, "Reinforcement-learning based dynamic transmission range adjustment in medium access control for underwater wireless sensor networks," *Electr.*, vol. 9, no. 10, 2020.

[34] A. D. Shoaei, M. Derakhshani, S. Parsaeefard, and T. Le-Ngoc, "Learning-based hybrid TDMA-CSMA MAC protocol for virtualized 802.11 WLANs," *2015 IEEE 26th Annual Int. Symp. PIMRC*, pp. 1861-1866, Hong Kong, China, 2015. (https://doi.org/10.3390/electronics9101727)

[35] S. Jang, K. G. Shin, and S. Bahk, "Post-CCA and reinforcement learning based bandwidth adaptation in 802.11ac networks," in *IEEE Trans. Mobile Computing*, vol. 17, no. 2, pp. 419-432, Feb. 2018. (https://doi.org/10.1109/TMC.2017.2709309)

[36] H. Cha and S.-L. Kim, "A reinforcement learning approach to dynamic spectrum access in internet-of-things networks," *2019 IEEE ICC*, pp. 1-6, Shanghai, China, 2019. (https://doi.org/10.1109/ICC.2019.8762091)

[37] C. Adjih, C. S. Chen, C. S. Gobin, and I. Hmedoush, "Designing medium access control protocol sequences through deep reinforcement learning," *2023 EuCNC/6G Summit*, Gothenburg, Sweden, 2023. (https://doi.org/10.1109/EuCNC/6GSummit58263.2023.10188299)

[38] H. Dutta and S. Biswas, "Medium access using distributed reinforcement learning for IoTs with low-complexity wireless transceivers," *2021 IEEE 7th WF-IoT*, pp. 356-361, New Orleans, LA, USA, 2021. (https://doi.org/10.1109/WF-IoT51360.2021.9595062)

[39] H. Dutta and S. Biswas, "Distributed reinforcement learning for scalable wireless medium access in IoTs and sensor networks,"

Computer Netw., vol. 202, p. 108662, 2022. (https://doi.org/10.1016/j.comnet.2021.108662)

[40] H. Park, H. Kim, S. -T. Kim, and P. Mah, "Multi-agent reinforcement-learning-based time-slotted channel hopping medium access control scheduling scheme," in *IEEE Access,* vol. 8, pp. 139727-139736, 2020. (https://doi.org/10.1109/ACCESS.2020.3010575)

[41] Z. Guo, Z. Chen, P. Liu, J. Luo, X. Yang, and X. Sun, "Multi-agent reinforcement learning-based distributed channel access for next generation wireless networks," in *IEEE J. Sel. Areas in Commun.*, vol. 40, no. 5, pp. 1587-1599, May 2022. (https://doi.org/10.1109/JSAC.2022.3143251)

[42] N. Keshtiarast, O. Renaldi, and M. Petrova, "Wireless MAC protocol synthesis and optimization with multi-agent distributed reinforcement learning," in *IEEE Networking Lett.,* vol. 6, no. 4, Dec. 2024. (https://doi.org/10.1109/LNET.2024.3503289)

[43] J. Xiao, Z. Chen, X. Sun, W. Zhan, X. Wang, and X. Chen, "Online multi-agent reinforcement learning for multiple access in wireless networks," in *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3250-3254, Dec. 2023. (https://doi.org/10.1109/LCOMM.2023.3326267)

[44] M. P. Mota, A. Valcarce, J. -M. Gorce, and J. Hoydis, "The emergence of wireless MAC protocols with multi-agent reinforcement learning," *2021 IEEE GC Wkshps*, pp. 1-6, Madrid, Spain, 2021. (https://doi.org/10.1109/GCWkshps52748.2021.9681991)

[45] Z. Zhang, et al., "6G wireless networks: Vision, requirements, architecture, and key technologies," in *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28-41, Sep. 2019. (https://doi.org/10.1109/MVT.2019.2921208)

박 준 영 (Jun-young Park)

2024년 2월 : 국립공주대학교 컴퓨터공학부 소프트웨어전공 졸업

2024년 3월~현재 : 국립공주대학교 소프트웨어학과 석사과정

<관심분야> MAC 프로토콜, 강화학습, 인공지능

[ORCID:0009-0005-2580-0498]

나 웅 수 (Woongsoo Na)

2010년 2월 : 중앙대학교 공학사

2012년 2월 : 중앙대학교 공학석사

2017년 2월 : 중앙대학교 공학박사

2018년~2019년 : 한국전자통신연구원 통신미디어 연구소

2020년~2024년 : 국립공주대학교 소프트웨어학과 조교수

2024년~현재 : 국립공주대학교 소프트웨어학과 부교수

<관심분야> MAC 프로토콜, 인공지능, 5G/6G 네트워크, 자원 최적화

[ORCID:0000-0003-3861-8001]