(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년05월13일
(11) 등록번호 10-2806734
(24) 등록일자 2025년05월08일

(51) 국제특허분류(Int. Cl.)
H04B 7/06 (2017.01) *G06N 3/092* (2023.01)
H04B 17/336 (2015.01) *H04B 7/0413* (2017.01)
H04W 84/12 (2009.01)
(52) CPC특허분류
H04B 7/0617 (2013.01)
G06N 3/092 (2023.01)
(21) 출원번호 10-2024-0119971
(22) 출원일자 2024년09월04일
심사청구일자 2024년09월04일
(56) 선행기술조사문헌
KR102508071 B1*
조부성, 김문석, "IEEE 802.11ay 표준 무선랜의 MU-MIMO 빔포밍 훈련 데이터 딥러닝 적용 방안", 2023년도 한국통신학회 추계종합학술발표회(2023.11.22) 1부.*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
김문석
서울특별시 광진구 능동로 209 (군자동)
조부성
서울특별시 광진구 능동로 209 (군자동)
(74) 대리인
수안특허법인

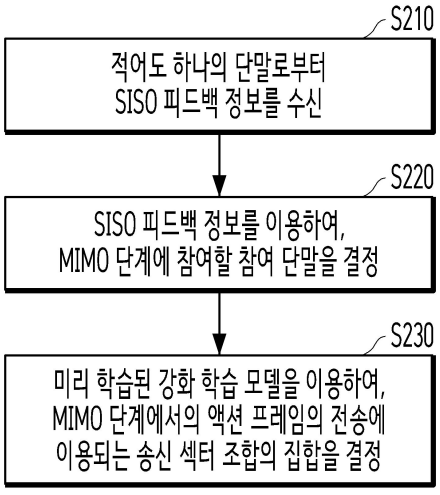
전체 청구항 수 : 총 10 항 심사관 : 구영회

(54) 발명의 명칭 **강화 학습 기반의 빔포밍 훈련 방법 및 장치**

(57) 요 약

무선랜 시스템에서, 강화 학습 기반으로 빔포밍 훈련을 수행하는 방법과 장치가 개시된다. 개시된 강화 학습 기반의 빔포밍 훈련 방법은 적어도 하나의 단말로부터 SISO 피드백 정보를 수신하는 단계; 상기 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하는 단계; 및 미리 학습된 강화 학습 모델을 이용하여, 상기 MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정하는 단계를 포함하며, 상기 강화 학습 모델의 상태 정보는, 상기 SISO 피드백 정보, SNR 임계값 및 상기 참여 단말에 대한 정보를 포함하며, 상기 강화 학습 모델의 액션 정보는, 상기 송신 섹터 조합에 대한 정보 및 상기 액션 프레임의 추가 전송 여부에 대한 정보를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
H04B 17/336 (2015.01)
H04B 7/0413 (2013.01)
H04W 84/12 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 2710016156
과제번호 2022R1G1A1012117
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 딥러닝 기반 차세대 밀리미터파 무선랜 통신 및 센싱 기술 연구
과제수행기관명 세종대학교
연구기간 2024.03.01 ~ 2025.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 2710008709
과제번호 II211816
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성(R&D)
연구과제명 메타버스 자율트윈 핵심기술 연구
과제수행기관명 세종대학교 산학협력단
연구기간 2024.01.01 ~ 2024.12.31
공지예외적용 : 있음

명 세 서

청구범위

청구항 1

적어도 하나의 단말로부터 SISO 피드백 정보를 수신하는 단계;

상기 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하는 단계; 및

미리 학습된 강화 학습 모델을 이용하여, 상기 MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정하는 단계를 포함하며,

상기 강화 학습 모델의 상태 정보는,

상기 SISO 피드백 정보, SNR 임계값 및 상기 참여 단말에 대한 정보를 포함하며,

상기 강화 학습 모델의 액션 정보는,

상기 송신 섹터 조합에 대한 정보 및 상기 액션 프레임의 추가 전송 여부에 대한 정보를 포함하며,

상기 강화 학습 모델을 위한 보상값은

상기 액션 프레임을 수신한 참여 단말의 개수와, 상기 결정된 참여 단말의 개수에 따라서 결정되며,

상기 강화 학습 모델의 보상값은

하기 [수학식 1]의 보상 함수에 따라서 결정되는

강화 학습 기반의 빔포밍 훈련 방법.

[수학식 1]

$$\beta \times \left(\frac{N_{STA}}{|u_{tot}| - 1}\right) \times (|u_{tot}| - N_{Transmit}) \quad (|u_{c_{setup}}| = |u_{tot}|)$$
$$-(|u_{tot}| - |u_{c_{setup}}|) \quad (|u_{c_{setup}}| < |u_{tot}|)$$

여기서, $|u_{c_{setup}}|$ 은 상기 액션 프레임을 수신한 참여 단말의 개수, $|u_{tot}|$ 는 상기 결정된 참여 단말의 개수, β 는 튜닝 파라미터, $N_{Transmit}$ 는 상기 액션 프레임의 전송 횟수, N_{STA} 는 상기 단말의 개수를 나타냄.

청구항 2

제 1항에 있어서,

상기 상태 정보에 포함된 SISO 피드백 정보는

송신 섹터에 대한 SISO 피드백 정보 및 상기 송신 섹터 조합에 대한 SISO 피드백 정보

를 포함하는 강화 학습 기반의 빔포밍 훈련 방법.

청구항 3

제 1항에 있어서,

상기 참여 단말에 대한 정보는

상기 단말의 참여 단말 여부를 나타내는 정보 및 상기 액션 프레임에 대한 상기 참여 단말의 수신 여부를 나타내는 정보

를 포함하는 강화 학습 기반의 빔포밍 훈련 방법.

청구항 4

삭제

청구항 5

삭제

청구항 6

제 1항에 있어서,

상기 송신 섹터 조합의 집합을 결정하는 단계는

상기 액션 프레임이 추가로 전송되지 않는 제1조건 및 상기 액션 정보의 생성 횟수가 상기 결정된 참여 단말의 개수에 대응되는 제2조건 중, 적어도 하나가 만족될때까지 상기 액션 정보를 생성하는

강화 학습 기반의 빔포밍 훈련 방법.

청구항 7

제 1항에 있어서,

상기 참여 단말은

상기 단말 중, 상기 SNR 임계값 이상의 SNR값을 포함하는 상기 SISO 피드백 정보를 전송한 단말인,

강화 학습 기반의 빔포밍 훈련 방법.

청구항 8

제 1항에 있어서,

상기 송신 섹터 조합의 집합을 결정하는 단계는

상기 상태 정보를 입력받아, 상기 액션 정보를 출력하는 인공 신경망을 이용하여, 상기 송신 섹터 조합의 집합을 결정하는

강화 학습 기반의 빔포밍 훈련 방법.

청구항 9

제 8항에 있어서,

상기 인공 신경망은

상기 상태 정보를 입력받는 입력층;

복수의 은닉층;

출력층;

상기 은닉층 사이 및 상기 출력층과 마지막 은닉층의 사이에 배치되는 제1활성화 함수 레이어; 및

상기 출력층의 출력값을 입력받아 상기 액션 정보를 출력하는 제2활성화 함수 레이어

를 포함하는 강화 학습 기반의 빔포밍 훈련 방법.

청구항 10

단말로부터 SISO 피드백 정보를 수신하고, 액션 프레임을 전송하는 위상 배열 안테나;

메모리; 및

프로세서를 포함하며,

상기 프로세서는

상기 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하고,

미리 학습된 강화 학습 모델을 이용하여, 상기 MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정하며,

상기 강화 학습 모델의 상태 정보는,

상기 SISO 피드백 정보, SNR 임계값 및 상기 참여 단말에 대한 정보를 포함하며,

상기 강화 학습 모델의 액션 정보는,

상기 송신 섹터 조합에 대한 정보 및 상기 액션 프레임의 추가 전송 여부에 대한 정보를 포함하며,

상기 강화 학습 모델을 위한 보상값은

상기 액션 프레임을 수신한 참여 단말의 개수와, 상기 결정된 참여 단말의 개수에 따라서 결정되며,

상기 강화 학습 모델의 보상값은

하기 [수학식 1]의 보상 함수에 따라서 결정되는

액세스 포인트.

[수학식 1]

$$\beta \times \left(\frac{N_{STA}}{|u_{tot}| - 1}\right) \times (|u_{tot}| - N_{Transmit}) \quad (|u_{c_{setup}}| = |u_{tot}|)$$

$$-(|u_{tot}| - |u_{c_{setup}}|) \quad (|u_{c_{setup}}| < |u_{tot}|)$$

여기서, $|u_{c_{setup}}|$ 은 상기 액션 프레임을 수신한 참여 단말의 개수, $|u_{tot}|$ 는 상기 결정된 참여 단말의 개수, β 는 튜닝 파라미터, $N_{Transmit}$ 는 상기 액션 프레임의 전송 횟수, N_{STA} 는 상기 단말의 개수를 나타냄.

청구항 11

제 10항에 있어서,

상기 상태 정보에 포함된 SISO 피드백 정보는

송신 섹터에 대한 SISO 피드백 정보 및 상기 송신 섹터 조합에 대한 SISO 피드백 정보

를 포함하는 액세스 포인트.

청구항 12

제 10항에 있어서,

상기 참여 단말에 대한 정보는

상기 단말의 참여 단말 여부를 나타내는 정보 및 상기 액션 프레임에 대한 상기 참여 단말의 수신 여부를 나타내는 정보

를 포함하는 액세스 포인트.

청구항 13

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 무선랜 시스템에서의 빔포밍 훈련 방법 및 장치에 관한 것으로서, 더욱 상세하게는 강화 학습 기반의 빔포밍 훈련 방법 및 장치에 관한 것이다.

배 경 기 술

[0003] IEEE 802.11 표준에 따르면, MU-MIMO 빔포밍 훈련은 SISO 단계와 MIMO 단계로 수행된다.

[0004] SISO 단계에서 액세스 포인트는 송신 섹터별로 짧은 섹터 스윕 프레임(short sector sweep frame)을 전송하고, 단말로부터 SISO 피드백 정보를 수신한다. SISO 피드백 정보는 송신 섹터별로 전송된 짧은 섹터 스윕 프레임에 대해서, 단말이 측정한 SNR값을 포함한다. 액세스 포인트는 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 단말을 결정한다.

[0005] 그리고 MIMO 단계는 세부적으로 BF 셋업(setup) 하위 단계, BF 훈련(training) 하위 단계, BF 피드백(feedback) 하위 단계 및 BF 선택(selection) 하위 단계로 이루어진다. 각 하위 단계에서 액세스 포인트는 빔포밍 훈련을 위한 액션 프레임을 단말로 전송한다.

[0006] BF 셋업 하위 단계에서 전송되는 액션 프레임은 MIMO 단계에 참여하는 단말들의 리스트를 포함한다. BF 훈련 하위 단계에서 전송되는 액션 프레임들은, beam refinement protocol-receive/transmit(BRP-RX/TX) 프레임들이며, 이러한 프레임들은 후보 전송 안테나 가중치 벡터(antenna weight vector, AWV들)을 테스트하기 위한 트레이닝 필드를 포함한다. BF 피드백 하위 단계에서 액세스 포인트는 각 단말들로부터 BF 훈련 하위 단계 동안 수행된 테스트에 대한 피드백 프레임들을 수집한다. BF 선택 하위 단계에서 전송되는 액션 프레임에는 DL MU-MIMO 전송에 사용할 수신 안테나 가중치 벡터들이 포함된다.

[0007] 관련 선행문헌으로 대한민국 등록특허 제10-2648565호, 제10-2508071호, 제10-2154481호, 제10-2153923호, 대한민국 공개특허 제2022-0036494호가 있다.

발명의 내용

해결하려는 과제

[0009] 본 발명은, 빔포밍 훈련을 위한 액션 프레임이 효율적으로 전송될 수 있도록, 송신 섹터를 위상 배열 안테나에 할당하는 빔포밍 훈련 방법을 제공하기 위한 것이다.

[0010] 특히 본 발명은, BF 셋업 하위 단계 및 BF 선택 하위 단계에서 액션 프레임이 효율적으로 전송될 수 있는 송신 섹터 조합을 결정하여 빔포밍 훈련을 수행하는 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0012] 상기한 목적을 달성하기 위한 본 발명의 일 실시예에 따르면, 적어도 하나의 단말로부터 SISO 피드백 정보를 수신하는 단계; 상기 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하는 단계; 및 미리 학

습된 강화 학습 모델을 이용하여, 상기 MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정하는 단계를 포함하며, 상기 강화 학습 모델의 상태 정보는, 상기 SISO 피드백 정보, SNR 임계값 및 상기 참여 단말에 대한 정보를 포함하며, 상기 강화 학습 모델의 액션 정보는, 상기 송신 섹터 조합에 대한 정보 및 상기 액션 프레임의 추가 전송 여부에 대한 정보를 포함하는 강화 학습 기반의 빔포밍 훈련 방법이 제공된다.

[0013] 또한 상기한 목적을 달성하기 위한 본 발명의 다른 실시예에 따르면, 단말로부터 SISO 피드백 정보를 수신하고, 액션 프레임을 전송하는 위상 배열 안테나; 메모리; 및 프로세서를 포함하며, 상기 프로세서는 상기 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하고, 미리 학습된 강화 학습 모델을 이용하여, 상기 MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정하며, 상기 강화 학습 모델의 상태 정보는, 상기 SISO 피드백 정보, SNR 임계값 및 상기 참여 단말에 대한 정보를 포함하며, 상기 강화 학습 모델의 액션 정보는, 상기 송신 섹터 조합에 대한 정보 및 상기 액션 프레임의 추가 전송 여부에 대한 정보를 포함하는 액세스 포인트가 제공된다.

발명의 효과

[0015] 본 발명의 일실시예에 따르면, 강화 학습 모델을 이용함으로써, 빔포밍 훈련을 위한 액션 프레임의 불필요한 전송이 감소하고, 빔포밍 훈련 시간이 줄어들 수 있다.

도면의 간단한 설명

[0017] 도 1은 무선랜 시스템에서의 일반적인 빔포밍 훈련 방법을 설명하기 위한 도면이다.

도 2는 본 발명의 일실시예에 따른 무선랜 시스템에서 빔포밍 훈련 방법을 설명하기 위한 도면이다.

도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 도시하는 도면이다.

도 4는 본 발명의 일실시예에 따른 강화 학습 모델을 위한 인공신경망을 설명하기 위한 도면이다.

도 5는 본 발명의 일실시예에 따른 강화 학습 기반의 빔포밍 훈련 장치를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0019] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0021] 도 1은 무선랜 시스템에서의 일반적인 빔포밍 훈련 방법을 설명하기 위한 도면으로서, 도 1에는 송신 섹터가 6개(TS1 내지 TS6)이고, 액세스 포인트의 위상 배열 안테나가 2개이며, 제1위상 배열 안테나(110)에는 제1 내지 제3송신 섹터(TS1, TS2, TS3)가 할당되고, 제2위상 배열 안테나(120)에는 제4 내지 제6송신 섹터(TS4, TS5, TS6)가 할당된 예시가 도시된다.

[0022] 액세스 포인트는 SISO 단계에서 각 송신 섹터 별로 짧은 섹터 스윕 프레임을 전송한다. 제1위상 배열 안테나(110)는 제1 내지 제3송신 섹터(TS1, TS2, TS3)로 짧은 섹터 스윕 프레임을 전송하고, 제2위상 배열 안테나(120)는 제4 내지 제6송신 섹터(TS4, TS5, TS6)로 짧은 섹터 스윕 프레임을 전송한다. 액세스 포인트에 접속된 단말은 짧은 섹터 스윕 프레임을 수신하여 SNR을 측정하고, 측정된 SNR값을 포함하는 SISO 피드백 정보를 액세스 포인트로 전송한다. 즉 하나의 무선 단말은, 제1 내지 제6송신 섹터(TS1 내지 TS6)에 대해 측정된 6개의 SNR 값을 포함하는 SISO 피드백 정보를 액세스 포인트로 전송한다. 예컨대, 6개의 송신 섹터에 4개의 단말이 위치한 경우, 4개의 단말은 각각 6개의 송신 섹터에 대해 측정된 SNR값을 포함하는 SISO 피드백 정보를 액세스 포인트로 전송할 수 있다.

[0023] 액세스 포인트는 측정된 SNR값을 이용하여 MIMO 단계에서 빔포밍 훈련을 수행한다. 액세스 포인트는 BF 셋업 하위 단계, BF 훈련 하위 단계, BF 피드백 하위 단계 및 BF 선택 하위 단계 각각에서 액션 프레임을 전송하여, 빔포밍 훈련을 수행한다.

[0024] 액세스 포인트는, MIMO 단계에서 액세스 포인트에 접속한 단말이 모두 수신할 수 있도록 액션 프레임을 전송하

며, 이를 위해 복수의 송신 섹터의 일부 송신 섹터를 포함하는 송신 섹터 조합 중 하나를 선택하여, 액션 프레임을 전송할 수 있다. 송신 섹터 조합은, 위상 배열 안테나 각각에 할당된 송신 섹터 중에서 하나씩 선택되어 결정될 수 있다. 도 1과 같은 예시에서, 송신 섹터 조합은 (TS1, TS4), (TS1, TS5), (TS1, TS6), (TS2, TS4), (TS2, TS5), (TS2, TS6), (TS3, TS4), (TS3, TS5), (TS3, TS6)와 같이, 9개로 결정될 수 있으며, 액세스 포인트는 이러한 송신 섹터 조합 중에 하나를 선택하여, 선택된 송신 섹터 조합에 포함된 복수의 송신 섹터로 액션 프레임을 전송할 수 있다.

[0025] 모든 송신 섹터 조합으로 액션 프레임을 전송하는 것은 빔포밍 훈련 시간의 증가를 초래하여 비효율적이므로, 모든 참여 단말이 액션 프레임을 수신하면서도 액션 프레임의 전송 횟수가 최소가 될 수 있는 효율적인 송신 섹터 조합을 결정하는 것이 필요하다. 특히, 모든 참여 단말이 액션 프레임을 수신하도록 액션 프레임이 전송되는 BF 셋업 하위 단계, BF 선택 하위 단계에서, 액션 프레임의 전송 횟수가 최소가 될 수 있도록, 효율적인 송신 섹터 조합의 결정이 필요하다. 이에 본 발명은 강화 학습을 이용하여 최적의 송신 섹터 조합을 결정하고, 이를 통해 빔포밍 훈련을 수행하는 방법을 제안한다.

[0026] 본 발명의 일실시예에 따른 빔포밍 훈련 방법은 무선랜 시스템의 액세스 포인트에서 수행된다.

[0028] 도 2는 본 발명의 일실시예에 따른 무선랜 시스템에서 빔포밍 훈련 방법을 설명하기 위한 도면이며, 도 3은 본 발명의 일실시예에 따른 강화 학습 모델을 도시하는 도면이다.

[0029] 도 2를 참조하면 본 발명의 일실시예에 따른 액세스 포인트는 적어도 하나의 단말로부터 SISO 피드백 정보를 수신(S210)하고, SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정(S220)한다. 액세스 포인트는 SISO 피드백 정보에 포함된 SNR값과 미리 설정된 SNR 임계값을 비교하여, 참여 단말을 결정할 수 있으며, SNR 임계값 이상의 SNR값을 포함하는 SISO 피드백 정보를 전송한 단말을 참여 단말로 결정할 수 있다. SNR 임계값은 사용자에 의해 설정될 수 있으며, 후술되는 강화 학습 모델의 상태 정보로도 이용된다.

[0030] 액세스 포인트는 미리 학습된 강화 학습 모델을 이용하여, MIMO 단계에서의 액션 프레임의 전송에 이용되는 송신 섹터 조합의 집합을 결정(S230)한다. 여기서, 송신 섹터 조합의 집합이란, 위상 배열 안테나 각각에 대해 할당된 송신 섹터 중에서, 액션 프레임의 전송에 이용되는 송신 섹터 조합을 포함하는 집합을 의미한다.

[0031] 강화 학습 모델은 일실시예로서 도 3의 어드밴티지 액터-크리틱(advantage actor critic, A2C) 모델일 수 있다. 강화 학습 모델에서 에이전트(RL Agent)는 액세스 포인트에 대응되며, 환경(Environment)은 빔포밍 훈련에 대응된다. 그리고 에이전트는 액터와 크리틱으로 구성될 수 있다. 에이전트, 즉 액세스 포인트는 현재 빔포밍 환경의 상태(state, S_t)를 관찰(observation)하며, 관찰된 상태에서 보상(reward, R_t)이 최대가 되는 액션(action, A_t)을 수행한다. 여기서 액션은 현재 송신 섹터 조합의 집합을 결정하는 것에 대응되며, 액터와 크리틱은, TD(Temporal Difference) 에러를 통해 업데이트될 수 있다.

[0032] 강화 학습 모델의 상태 정보는, SISO 피드백 정보, SNR 임계값 및 참여 단말에 대한 정보를 포함한다. 여기서, SISO 피드백 정보는 송신 섹터에 대한 SISO 피드백 정보 및 송신 섹터 조합에 대한 SISO 피드백 정보를 포함하며, 참여 단말에 대한 정보는 단말의 참여 단말 여부를 나타내는 정보 및 액션 프레임에 대한 참여 단말의 수신 여부를 나타내는 정보를 포함할 수 있다.

[0033] 그리고 강화 학습 모델의 액션 정보는 송신 섹터 조합에 대한 정보 및 액션 프레임의 추가 전송 여부에 대한 정보를 포함한다. 여기서, 송신 섹터 조합에 대한 정보는, 강화 학습 모델의 액션에 의해 결정된 송신 섹터 조합을 나타낸다.

[0034] 일실시예로서, 상태 정보(S_t)는 [수학식 1]과 같이 표현될 수 있으며, 다차원의 연속적인 값들로 이루어질 수 있다.

수학식 1

[0035]
$$\mathbf{S_t} = (\mathbf{F}, \mathbf{F_{A_{t-1}}}, \delta_{\mathbf{mu}}, \mathbf{P_{u_{tot}}})$$

[0036] 여기서, $\mathbf{F}$는 송신 섹터에 대한 SISO 피드백 정보, $\mathbf{F_{A_{t-1}}}$는 송신 섹터 조합에 대한 SISO 피드백 정보, $\delta_{\mathbf{mu}}$는

SNR 임계값, $\mathbf{P_{u_{tot}}}$ 는 참여 단말에 대한 정보를 나타낸다.

[0037] 도 1의 예시에서와 같이, 6개의 송신 섹터(TS1 내지 TS6)가 이용되는 경우, 송신 섹터에 대한 SISO 피드백 정보($\mathbf{F}$)는 $(\mathbf{F_{TS1}}, \cdots, \mathbf{F_{TS6}})$ 로 표현될 수 있으며, $\mathbf{F_{TS1}}$은 제1송신 섹터(TS1)에 대한 SISO 피드백 정보를 나타낸다. 그리고 4개의 단말(STA1 내지 STA4)이 송신 섹터에 위치할 때, 4개의 단말(STA1 내지 STA4)이 각각 SISO 피드백 정보를 전송하므로, 제1송신 섹터에 대한 SISO 피드백 정보($\mathbf{F_{TS1}}$)는, $(f_{TS1}^{STA1}, \cdots, f_{TS1}^{STA4})$ 로 표현될 수 있다. f_{TS1}^{STA1} 는 제1단말(STA1)이 전송하는 제1송신 섹터(TS1)에 대한 SISO 피드백 정보를 나타낸다.

[0038] $\mathbf{F_{A_{t-1}}}$는 1번째부터 t-1번째까지 강화 학습 모델의 액션에 의해 결정된 송신 섹터 조합에 대한 SISO 피드백 정보를 나타내며, $(\mathbf{F_{c_1}}, \cdots, \mathbf{F_{c_{t-1}}}, 0, \cdots, 0)$ 와 같이 표현될 수 있다. 여기서, $\mathbf{F_{c_{t-1}}}$은 t-1번째에 강화 학습 모델의 액션에 의해 결정된 송신 섹터 조합에 대한 SISO 피드백 정보를 나타낸다. 만일, 제1 및 제4송신 섹터(TS1, TS4)가 송신 섹터 조합으로 결정된 경우, 송신 섹터 조합에 대한 SISO 피드백 정보는 제1 및 제4송신 섹터(TS1, TS4)에 대해 측정된 SNR값을 포함할 수 있다.

[0039] SNR 임계값은 단말들이 프레임들을 성공적으로 수신하기 위한 임계값으로서, 사용자에 의해 설정될 수 있다.

[0040] 참여 단말에 대한 정보는 송신 섹터에 위치한 단말별로 할당되는 정보로서, 예컨대 4개의 단말(STA1 내지 STA4)이 송신 섹터에 위치할 때, $(\mathrm{pSTA1}, \mathrm{pSTA2}, \mathrm{pSTA3}, \mathrm{pSTA4})$ 으로 표현될 수 있다. $\mathrm{pSTA1}$은 제1 단말(STA1)에 대한 정보를 나타낸다. 단말이 참여 단말로 결정되지 않은 경우, 즉 참여 단말 그룹($\mathbf{u_{tot}}$)에 포함되지 않은 경우 -1, 단말이 참여 단말로 결정되었지만 액션 프레임을 수신하지 못한 경우 0, 단말이 참여 단말로 결정되고 액션 프레임을 최소 한번 수신한 경우 1이, 참여 단말에 대한 정보로 할당될 수 있다.

[0041] $\mathbf{s_t}$ 에 포함되는 각 요소들의 크기는 $|\mathbf{F}| = \mathrm{N_{STA}} \times \mathrm{N_{PAA}} \times \mathrm{N_{TS}}$, $|\mathbf{F_{A_{t-1}}}| = \mathrm{N_{PAA}} \times \mathrm{N_{STA}} \times \mathrm{N_{STA}}$, $|\mathbf{P_{u_{tot}}}| = \mathrm{N_{STA}}$ 와 같으며, 이러한 크기에 의해 후술되는 인공신경망의 입력층의 노드 개수가 결정될 수 있다. 여기서, $\mathrm{N_{STA}}$ 는 송신 섹터에 포함되는 단말의 개수, $\mathrm{N_{PAA}}$ 는 위상 배열 안테나의 개수, $\mathrm{N_{TS}}$ 는 송신 섹터의 개수를 나타낸다.

[0042] 일실시예로서, 액션 정보($\mathbf{A_t}$)는 [수학식 2]와 같이 표현될 수 있으며, $(\mathrm{N_{PAA}} + 1)$ 차원의 이산적인 값들로 이루어질 수 있다.

수학식 2

[0043] $$\mathbf{A_t} = (\mathbf{c_t}, \omega_t)$$

[0044] 여기서, $\mathbf{c_t}$ 는 t번째에 강화 학습 모델의 액션에 의해 결정된 송신 섹터 조합을 나타내며, $(\mathrm{TS}_{\mathrm{PAA}_1}^{\mathrm{t}}, \cdots, \mathrm{TS}_{\mathrm{PAA}_{\mathrm{N_{PAA}}}}^{\mathrm{t}})$ 로 표현될 수 있다. ω_t 는 t번째의 액션 프레임 전송 이후, 추가로 액션 프레임을 전송할지 여부를 나타내는 값으로서, 0 또는 1로 결정될 수 있다. 0은 액션 프레임을 추가로 전송하지 않는 것을 나타내며, 1은 액션 프레임을 추가로 전송하는 것을 나타낼 수 있다.

[0045] 단계 S230에서 액세스 포인트는 강화 학습 모델의 타임 스텝(t) 증가에 따라서 상태 정보와 액션 정보를 갱신하면서, 송신 섹터 조합의 집합을 결정한다. 타임 스텝은 1부터 1씩 증가할 수 있으며, 액세스 포인트는 미리 설정된 제1 및 제2조건 중 적어도 하나가 만족될 때까지 액션 정보를 생성하여, 송신 섹터 조합의 집합을 결정한다.

[0046] 여기서, 제1조건은 액션 프레임이 추가로 전송되지 않는 조건이며, 제2조건은 액션 정보의 생성 횟수가 결정된 참여 단말의 개수에 대응되는 조건일 수 있다. 다시 말해 제1조건은, $\omega_t = 0$ 인 조건이며, 제2조건은 타임 스텝의 값이, 참여 단말의 개수에 대응되는 조건일 수 있다.

[0047] 액세스 포인트는 제1 및 제2조건 중 적어도 하나가 만족될 때까지 강화 학습 모델에 의해 생성된 적어도 하나의 송신 섹터 조합을 송신 섹터 조합의 집합으로 결정할 수 있으며, BF 셋업 하위 단계, BF 선택 하위 단계에서, 송신 섹터 조합의 집합에 포함된 송신 섹터 조합으로, 액션 프레임을 전송할 수 있다.

[0048] 또한 빔포밍 훈련 과정에서 획득된 데이터 예컨대, 상태 정보와, 액션 정보 등은 강화 학습 모델의 학습에 이용될 수 있다.

[0050] **보상 함수**

[0051] 강화 학습 모델을 위한 보상값은 액션 프레임을 수신한 참여 단말의 개수와, 단계 S220에서 결정된 참여 단말의 개수에 따라서 결정될 수 있다.

[0052] 이러한 보상을 위한 보상 함수(R_t)는 일실시예로서, [수학식 3]과 같이 정의될 수 있다.

수학식 3

[0053]
$$R_t = \begin{cases} 0 & (S_{t+1} \neq S_L) \\ \beta \times \left(\frac{N_{STA}}{|u_{tot}| - 1}\right) \times (|u_{tot}| - N_{Transmit}) & (|u_{c_{setup}}| = |u_{tot}|) \\ -(|u_{tot}| - |u_{c_{setup}}|) & (|u_{c_{setup}}| < |u_{tot}|) \end{cases}$$

[0054] 여기서, $S_{t+1} \neq S_L$ 은 제1 및 제2조건 중 어느 하나도 만족되지 않은 상태를 나타낸다. 그리고 $|u_{c_{setup}}| = |u_{tot}|$ 및 $|u_{c_{setup}}| < |u_{tot}|$ 각각은 제1 및 제2조건 중 적어도 하나가 만족된 상태에서, 액션 프레임을 수신한 참여 단말의 개수가, 단계 S220에서 결정된 참여 단말의 개수에 대응되는 경우 및 액션 프레임을 수신한 참여 단말의 개수가, 단계 S220에서 결정된 참여 단말의 개수보다 적은 경우를 나타낸다. 그리고 $|u_{c_{setup}}|$ 은 액션 프레임을 수신한 참여 단말의 개수, $|u_{tot}|$ 는 결정된 참여 단말의 개수, β 는 튜닝 파라미터, $N_{Transmit}$ 는 액션 프레임의 전송 횟수, N_{STA} 는 단말의 개수를 나타낸다. 튜닝 파라미터는 액세스 포인트의 환경과 좌표, SNR 임계값에 따라 달라질 수 있는 임의의 상수값으로서 실험적으로 결정될 수 있다.

[0055] 여기서, $\left(\frac{N_{STA}}{|u_{tot}| - 1}\right)$ 는 빔포밍 훈련 과정에서 얻을 수 있는 최대 보상값이, 결정된 참여 단말의 개수에 따라 달라지지 않도록 하는 역할을 한다.

[0057] 도 4는 본 발명의 일실시예에 따른 강화 학습 모델을 위한 인공신경망을 설명하기 위한 도면이다.

[0058] 본 발명의 일실시예에 따른 액세스 포인트는, 상태 정보를 입력받아, 액션 정보를 출력하는 인공신경망을 이용

하여, 송신 섹터 조합의 집합을 결정한다. 이러한 인공신경망은, 강화 학습 모델의 가치 함수($V_{\phi}(S)$)와 정책 함수($\pi^{\theta}(S,A)$)에 대한 인공신경망으로서, 액세스 포인트는 강화 학습 모델을 이용해 송신 섹터 조합의 집합을 결정하면서, 수집된 데이터를 이용하여 강화 학습 모델에 대한 학습을 수행하여, 인공신경망을 업데이트 할 수 있다.

[0059] 본 발명의 일실시예에 따른 강화 학습 모델에서 에이전트의 액터는 정책 함수를 이용하여 액션, 즉 송신 섹터 조합 집합의 결정을 수행한다. 그리고 에이전트의 크리틱(비평가)는 가치 함수를 이용하여 액터의 액션, 즉 송신 섹터 조합 집합의 결정이 적절한지 여부를 평가한다. 가치 함수의 평가 결과와 보상에 따라, 인공신경망의 파라미터(가중치)가 업데이트된다.

[0060] 도 4를 참조하면, 본 발명의 일실시예에 따른 입력층(410), 복수의 은닉층(420), 출력층(430), 제1활성화 함수 레이어(440), 제2활성화 함수 레이어(450)를 포함한다.

[0061] 입력층(410)은 상태 정보를 입력받으며, 제1활성화 함수 레이어(440)는 은닉층 사이에 배치되고, 출력층(430)과 마지막 은닉층 사이에 배치될 수 있다. 일실시예로서 제1활성화 함수 레이어(440)는 정류 선형 유닛(rectified linear unit, ReLU) 활성화 함수를 이용하는 레이어일 수 있다.

[0062] 제2활성화 함수 레이어(450)는 출력층(430)의 출력값을 입력받아 액션 정보를 출력하며, 일실시예로서 softmax 활성화 함수를 이용하는 레이어일 수 있다.

[0063] 출력층(430)의 노드는 $N_{PAA}+1$ 개의 노드 그룹(431, 432, 433)으로 분할될 수 있으며, 각 노드 그룹(431, 432, 433)은 액션 정보의 각 차원을 나타낸다.

[0064] 출력층(430)의 노드 그룹(431, 432, 433)들의 출력값에 softmax 활성화 함수를 적용시켜, 노드 그룹(431, 432, 433) 각각에서의 노드들의 출력값이 이산확률분포를 이루도록 하면, 출력층(430)의 노드의 출력값은 해당 액션을 선택할 확률을 나타낸다. 이 때, 각 노드에 할당된 액션이 선택될 확률은 [수학식 4]와 같이 표현될 수 있다.

수학식 4

[0065] $$\pi_{\theta}(\mathbf{S},\mathbf{A}) = \prod_{i=0}^{N_{PAA}} \pi_{\theta}(\mathbf{S},\mathbf{A}[i])$$

[0066] 가치 함수($V_{\phi}(\mathbf{S})$)에 대한 출력층은 1개의 노드(434)를 포함하며, [수학식 5]와 같이, 상태 정보로부터 획득되는 보상에 대한 기대값(V(S))을 출력한다.

수학식 5

[0067] $$V(\mathbf{S}) = E\{G_t|\mathbf{S}_t = \mathbf{S}\}$$

[0068] 액터는, 타임 스텝 t로부터 미래에 받을 감쇠된 보상값의 합인, G_t가 최대가되는 액션을 결정하며, G_t는 [수학식 6]과 같이 정의된다.

수학식 6

[0069] $$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots$$

[0070] 여기서, γ 은 감가율(Discount Factor)을 나타내며, R_t는 타임 스텝 t에서의 보상값을 나타낸다.

[0071] 액세스 포인트는 일실시예로서, 시간차 학습(Temporal Difference Learning)을 이용해, 인공신경망의 가중치를 업데이트할 수 있으며, $V_\phi(S_t)$에 대한 추정치 TD 타겟 y_t^ϕ 과, TD 에러 Δ_t^ϕ는 [수학식 7]과 같이 계산될 수 있다.

수학식 7

[0072] $$y_t^\phi = R_t + \gamma \times V_\phi(S_{t+1}) \times D_t$$
$$\Delta_t^\phi = y_t^\phi - V_\phi(S_t)$$

[0073] 여기서, D_t는 전술된 제1 및 제2조건 중 적어도 하나가 만족되는지 여부를 나타내는 값으로서, 만족된 경우 0, 만족되지 않은 경우 1로 설정될 수 있다.

[0074] 정책 함수의 손실 함수($loss_\pi$)와, 가치 함수의 손실 함수($loss_V$)는 [수학식 8]과 같이 정의되며, 인공신경망은 손실값이 최소가 되도록 업데이트될 수 있다.

수학식 8

[0075] $$loss_\pi = -\log\pi_\theta(S_t, A_t) \times \Delta_t$$
$$loss_V = lossfunction(V_\phi(S_t), y_t)$$

[0076] 한편 본 발명의 일실시예에 따른 인공신경망은 정책 함수와 가치 함수가 통합된 네트워크이므로, [수학식 9]와 같은 손실 함수의 손실값이 최소가 되도록 인공신경망의 가중치(파라미터)가 업데이트될 수 있다.

수학식 9

[0077] $$loss = -\log\pi_\phi(S_t, A_t) \times \Delta_t + smooth_l1_loss(V_\phi(S_t), y_t)$$

[0078] 가치 함수의 손실 함수는 smooth l1 손실 함수로 변환되어, 정책 함수의 손실 함수와 더해질 수 있다.

[0080] 도 5는 본 발명의 일실시예에 따른 강화 학습 기반의 빔포밍 훈련 장치를 설명하기 위한 도면으로서, 도 5에서는 액세스 포인트가 일실시예로서 도시된다.

[0081] 도 5를 참조하면, 본 발명의 일실시예에 따른 액세스 포인트는 위상 배열 안테나(510), 메모리(520) 및 프로세

서(530)를 포함한다.

[0082] 위상 배열 안테나(510)는 단말로부터 SISO 피드백 정보를 수신하고, 액션 프레임을 전송한다.

[0083] 프로세서(530)는 메모리(520)와 전기적으로 연결되며, 전술된 빔포밍 훈련 방법을 수행하기 위한 일련의 과정을 수행한다. 프로세서(530)는 수신된 SISO 피드백 정보를 이용하여, MIMO 단계에 참여할 참여 단말을 결정하고, 미리 학습된 강화 학습 모델을 이용하여, MIMO 단계에서의 액션 프레임의 전송에 이용될 송신 섹터 조합의 집합을 결정한다.

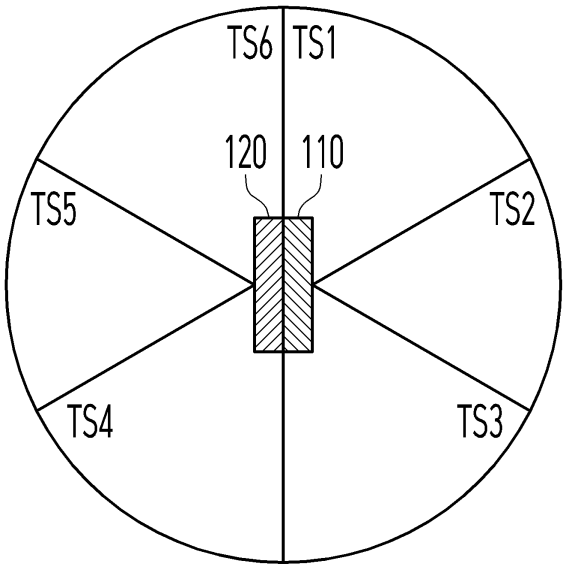
[0084] 여기서 강화 학습 모델은 전술된 강화 학습 모델로서, 강화 학습 모델의 상태 정보는, SISO 피드백 정보, SNR 임계값 및 참여 단말에 대한 정보를 포함하며, 강화 학습 모델의 액션 정보는, 송신 섹터 조합에 대한 정보 및 액션 프레임의 추가 전송 여부에 대한 정보를 포함한다.

[0086] 앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

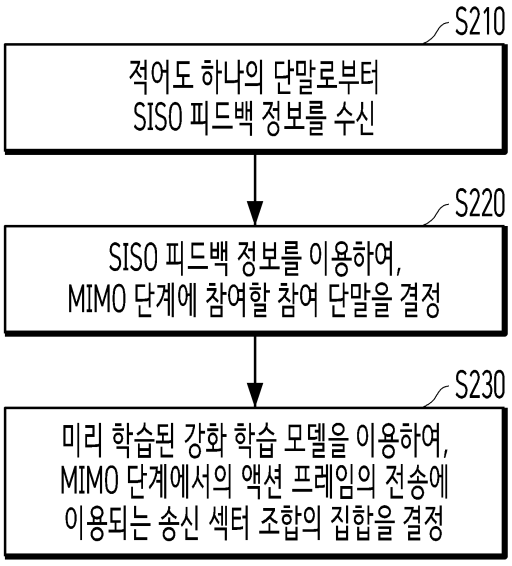
[0088] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

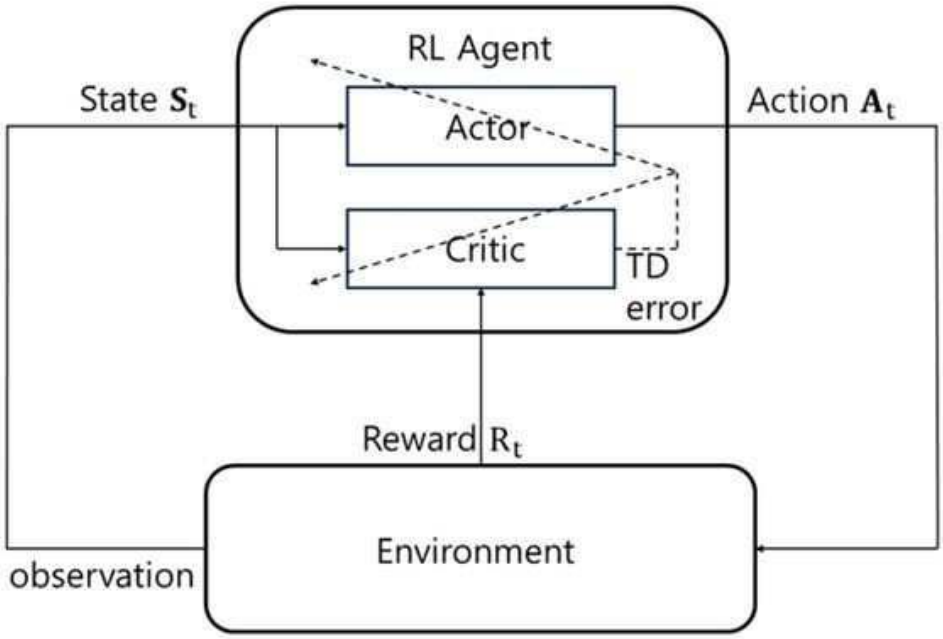
도면1



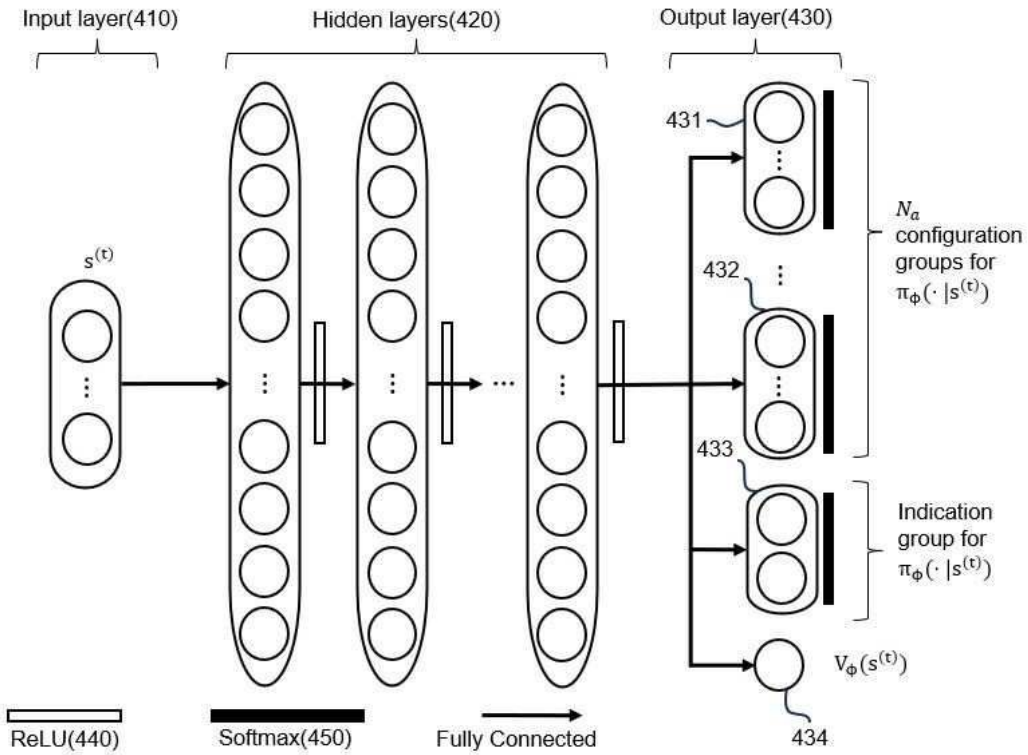
도면2


S210
적어도 하나의 단말로부터
SISO 피드백 정보를 수신
S220
SISO 피드백 정보를 이용하여,
MIMO 단계에 참여할 참여 단말을 결정
S230
미리 학습된 강화 학습 모델을 이용하여,
MIMO 단계에서의 액션 프레임의 전송에
이용되는 송신 섹터 조합의 집합을 결정

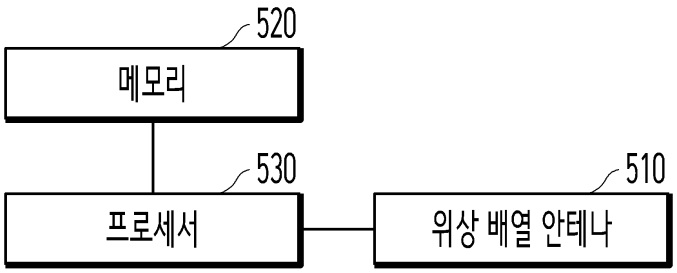
도면3


RL Agent
State S_t
Actor
Action A_t
Critic
TD
error
Reward R_t
Environment
observation

도면4


Input layer(410)
Hidden layers(420)
Output layer(430)
431
432
433
434
$s^{(t)}$
N_a configuration groups for $\pi_\phi(\cdot|s^{(t)})$
Indication group for $\pi_\phi(\cdot|s^{(t)})$
$V_\phi(s^{(t)})$
ReLU(440)
Softmax(450)
Fully Connected

도면5


520
메모리
530
프로세서
510
위상 배열 안테나