



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0006269  
(43) 공개일자 2024년01월15일

- |  |  |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/> <i>H04W 24/02</i> (2009.01) <i>G06N 3/04</i> (2023.01)<br/> <i>G06N 3/08</i> (2023.01) <i>H04W 24/08</i> (2009.01)<br/> <i>H04W 24/10</i> (2009.01) <i>H04W 84/12</i> (2009.01)</p> <p>(52) CPC특허분류<br/> <i>H04W 24/02</i> (2013.01)<br/> <i>G06N 3/04</i> (2023.01)</p> <p>(21) 출원번호 10-2022-0083124<br/> (22) 출원일자 2022년07월06일<br/> 심사청구일자 2022년07월06일</p> | <p>(71) 출원인<br/> <b>포항공과대학교 산학협력단</b><br/> 경상북도 포항시 남구 청암로 77 (지곡동)</p> <p>(72) 발명자<br/> <b>양현중</b><br/> 경상북도 포항시 남구 청암로 77<br/> <b>노현호</b><br/> 경상북도 포항시 남구 청암로 77</p> <p>(74) 대리인<br/> <b>특허법인(유한)아이시스</b></p> |
|--|--|

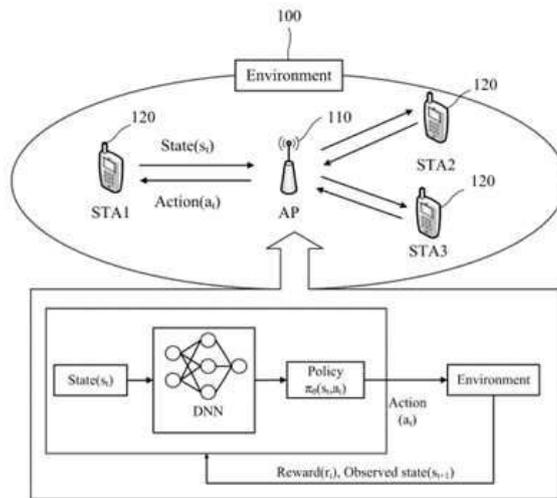
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 무선 통신 시스템에서 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법 및 장치

(57) 요약

본 개시의 일 실시예에 따라 무선 통신 시스템에서 상향링크 통신을 수행하는 AP의 방법이 제공된다. 상기 방법은, 복수의 STA(station)들로부터 상기 AP 및 상기 복수의 STA들 간의 상기 무선 통신 시스템에 대한 상태 정보를 수신하는 과정; 상기 상태 정보를 강화학습 네트워크에 입력하여 상기 복수의 STA들 중 주파수를 할당할 적어도 하나의 STA를 예측하는 과정; 및 상기 예측 결과를 이용하여 통신 성능을 최대로 하는 적어도 하나의 최적의 STA를 선택하는 과정을 포함한다.

대표도 - 도1



(52) CPC특허분류

*G06N 3/08* (2023.01)

*H04W 24/08* (2013.01)

*H04W 24/10* (2013.01)

*H04W 84/12* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

무선 통신 시스템에서 상향링크 통신을 수행하는 AP(access point)의 방법에 있어서, 상기 방법은, 복수의 STA(station)들로부터 상기 AP 및 상기 복수의 STA들 간의 상기 무선 통신 시스템에 대한 상태 정보를 수신하는 과정; 상기 상태 정보를 강화학습 네트워크에 입력하여 상기 복수의 STA들 중 주파수를 할당할 적어도 하나의 STA를 예측하는 과정; 및 상기 예측 결과를 이용하여 통신 성능을 최대로 하는 적어도 하나의 최적의STA를 선택하는 과정을 포함하는 방법.

#### 청구항 2

제1항에 있어서, 상기 상태 정보는 상기 AP 및 상기 복수의 STA들 간의 채널 이득 정보 또는 상기 복수의 STA들에 대한 SNR(signal to noise)정보를 포함함을 특징으로 하는 방법.

#### 청구항 3

제1항에 있어서, 상기 복수의 STA들에게 상기 상태 정보를 측정하기 위한 참조 신호를 전송하는 과정을 더 포함함을 특징으로 하는 방법.

#### 청구항 4

제1항에 있어서, 상기 선택된 적어도 하나의 최적의 STA와 상향링크 통신을 수행하는 과정을 더 포함함을 특징으로 하는 방법.

#### 청구항 5

제1항에 있어서, 상기 강화학습 네트워크는 DNN(deep neural network), CNN(convolution neural network), 또는 RNN(recurrent neural network) 중 적어도 하나를 포함함을 특징으로 하는 방법.

#### 청구항 6

제1항에 있어서, 상기 강화학습 네트워크의 상기 상태 정보는 이하의 수학적식에 따라 표현됨을 특징으로 하는 방법,

$$s_t = [h_{1t}, h_{2t}, \dots, h_{Mt}]$$

여기서  $s_t$ 는 슬롯 t에서 AP의 상태 정보이고,  $h_{it}$ 는 시간 슬롯 t에서 AP와 i번째 STA간 채널 이득 정보를 나타냄.

### 청구항 7

제6항에 있어서,

상기 강화학습 네트워크의 행동 정보는 이하의 수학적식에 따라 표현됨을 특징으로 하는 방법,

$$a_t = A = \{x \mid x \text{는 선택된 STA의 인덱스, } 1 \leq x \leq M\}$$

여기서  $a_t$ 는 상기 시간 슬롯 t에서 상기 AP의 행동 정보이고, A는 상기 복수의 STA들 중 상기 AP에 의해 선택된 상기 적어도 하나의 STA의 인덱스를 원소로 하는 집합이고, A는  $1 \leq n(A) \leq \min(N, M)$ 을 만족하고, N은 상기 AP의 안테나 수이고, M은 상기 복수의 STA들의 수를 나타냄.

### 청구항 8

제7항에 있어서,

상기 강화학습 네트워크는 이하의 수학적식에 따른 보상이 최대값이 되도록 학습됨을 특징으로 하는 방법,

$$r_t = \sum_{i \in A} BW_{it} \log_2 \left( 1 + \frac{P_{si}}{\sigma^2} \text{diag}_{i+1}((H^H H)^{-1}) \right)$$

여기서  $r_t$ 는 상기 시간 슬롯 t에서 상기 강화학습 네트워크의 보상이고,  $BW_{it}$ 는 시간 슬롯 t에서 A에 포함된 i번째 STA에게 할당된 주파수 대역을 나타내고,  $P_{si}$ 는 A에 포함된 i번째 STA가 상향링크를 통해 전송하는 신호의 세기를 나타내고,  $\sigma$ 는 가우시안 노이즈(Gaussian noise) n의 표준편차를 나타냄.

### 청구항 9

무선 통신 시스템에서 상향링크 통신을 수행하는 AP(access point)에 있어서, 상기 AP는,

통신부; 및

복수의 STA(station)들로부터 상기 AP 및 상기 복수의 STA들 간의 상기 무선 통신 시스템에 대한 상태 정보를 수신하도록 상기 통신부를 제어하고, 상기 상태 정보를 강화학습 네트워크에 입력하여 상기 복수의 STA들 중 주파수를 할당할 적어도 하나의 STA를 예측하고, 상기 예측 결과를 이용하여 통신 성능을 최대로 하는 적어도 하나의 최적의 STA를 선택하는 제어부를 포함하는 AP.

### 청구항 10

제9항에 있어서,

상기 상태 정보는 상기 AP 및 상기 복수의 STA들 간의 채널 이득 정보 또는 상기 복수의 STA들에 대한 SNR(signal to noise)정보를 포함함을 특징으로 하는 AP.

### 청구항 11

제9항에 있어서, 상기 제어부는,

상기 복수의 STA들에게 상기 상태 정보를 측정하기 위한 참조 신호를 전송하도록 상기 통신부를 제어함을 특징으로 하는 AP.

#### 청구항 12

제9항에 있어서, 상기 제어부는,

상기 선택된 적어도 하나의 최적의 STA와 상향링크 통신을 수행함을 특징으로 하는 AP.

#### 청구항 13

제9항에 있어서,

상기 AP는 상기 강화학습 네트워크를 저장하는 저장부를 더 포함하고,

상기 강화학습 네트워크는 DNN(deep neural network), CNN(convolution neural network), 또는 RNN(recurrent neural network) 중 적어도 하나를 포함함을 특징으로 하는 AP.

#### 청구항 14

제9항에 있어서,

상기 강화학습 네트워크의 상기 상태 정보는 이하의 수학식에 따라 표현됨을 특징으로 하는 AP,

$$s_t = [h_{1t}, h_{2t}, \dots, h_{Mt}]$$

여기서  $s_t$ 는 슬롯 t에서 AP의 상태 정보이고,  $h_{it}$ 는 시간 슬롯 t에서 AP와 i번째 STA간 채널 이득 정보를 나타냄.

#### 청구항 15

제14항에 있어서,

상기 강화학습 네트워크의 행동 정보는 이하의 수학식에 따라 표현됨을 특징으로 하는 AP,

$$a_t = A = \{x \mid x \text{는 선택된 STA의 인덱스, } 1 \leq x \leq M\}$$

여기서  $a_t$ 는 상기 시간 슬롯 t에서 상기 AP의 행동 정보이고, A는 상기 복수의 STA들 중 상기 AP에 의해 선택된 상기 적어도 하나의 STA의 인덱스를 원소로 하는 집합이고, A는  $1 \leq n(A) \leq \min(N, M)$ 을 만족하고, N은 상기 AP의 안테나 수이고, M은 상기 복수의 STA들의 수를 나타냄.

#### 청구항 16

제15항에 있어서,

상기 강화학습 네트워크는 이하의 수학식에 따른 보상이 최대값이 되도록 학습됨을 특징으로 하는 AP,

$$r_t = \sum_{i \in A} BW_{it} \log_2 \left( 1 + \frac{P_{si}}{\sigma^2} \text{diag}_{i+1}((H^H H)^{-1}) \right)$$

여기서  $r_i$ 는 상기 시간 슬롯 t에서 상기 강화학습 네트워크의 보상이고,  $BW_i$ 는 시간 슬롯 t에서 A에 포함된 i번째 STA에게 할당된 주파수 대역을 나타내고,  $P_i$ 는 A에 포함된 i번째 STA가 상향링크를 통해 전송하는 신호의 세기를 나타내고,  $\sigma$ 는 가우시안 노이즈(Gaussian noise) n의 표준편차를 나타냄.

### 발명의 설명

#### 기술 분야

[0001] 본 개시는 무선 통신 시스템에서 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법 및 장치에 관한 것이다.

#### 배경 기술

[0002] IEEE(Institute of Electrical and Electronic Engineers) 802.11ax는 증가하는 데이터 트래픽에 맞는 서비스를 제공하기 위하여 개발된 고효율 무선 통신 프로토콜이다. 제안된 프로토콜의 특징점으로는 OFDMA(orthogonal frequency division multiplexing access), DL/UL MU-MIMO(downlink/uplink multi-user-multi-input-multi-output) 등이 있다. 특히 OFDMA는 가용한 주파수 자원을 RU(resource unit)들이라 불리는 서브 채널로 나누어 주파수 효율을 높이는 기술이다. AP(access point)는 상향 링크 통신에 참여하는 STA(station)들에 RU를 적절하게 할당한다. 하지만, 상향 링크 통신 환경에서 OFDMA를 사용한 주파수 자원 할당 기술과 사용자 선택 알고리즘에 대하여 고려한 연구가 매우 부족하다.

[0003] 기존에는 OFDMA 및 MU-MIMO의 두 기술을 도입한 상향 링크 무선 통신에서의 사용자 선택 및 주파수 자원 할당 문제를 해결하는 방안은 모든 경우의 수를 계산하는 전역 탐색 기법(optimal)과 가장 높은 신호 대 잡음비(signal to noise ratio, SNR)를 갖는 사용자를 선택하여 서비스하는 최대 SNR(max-SNR) 기법이 대표적이다. 그러나, 전역 탐색 기법의 경우 선택 가능한 사용자 수가 증가할수록 계산 복잡도가 증가하고, 최대 SNR 기법의 경우 계산 복잡도는 전역 탐색 기법에 비해 감소하지만 통신 성능이 저하될 수 있다.

[0004] 따라서, 상향 링크 무선 통신에서 계산 복잡도를 감소시키면서 통신 성능을 개선하는 사용자 선택 기법이 필요하다.

### 발명의 내용

#### 해결하려는 과제

[0005] 본 개시는 무선 통신 시스템에서 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법 및 장치를 제공하는데 있다.

#### 과제의 해결 수단

[0006] 본 개시의 일 실시예에 따르면, 무선 통신 시스템에서 상향링크 통신을 수행하는 AP의 방법이 제공된다. 상기 방법은, 복수의 STA(station)들로부터 상기 AP 및 상기 복수의 STA들 간의 상기 무선 통신 시스템에 대한 상태 정보를 수신하는 과정; 상기 상태 정보를 강화학습 네트워크에 입력하여 상기 복수의 STA들 중 주파수를 할당할 적어도 하나의 STA를 예측하는 과정; 및 상기 예측 결과를 이용하여 통신 성능을 최대로 하는 적어도 하나의 최적의STA를 선택하는 과정을 포함한다.

[0007] 본 개시의 일 실시예에 따르면, 무선 통신 시스템에서 상향링크 통신을 수행하는 AP가 제공된다. 상기 AP는, 통신장치; 및 복수의 STA(station)들로부터 상기 AP 및 상기 복수의 STA들 간의 상기 무선 통신 시스템에 대한 상태 정보를 수신하도록 상기 통신장치를 제어하고, 상기 상태 정보를 강화학습 네트워크에 입력하여 상기 복수의 STA들 중 주파수를 할당할 적어도 하나의 STA를 예측하고, 상기 예측 결과를 이용하여 통신 성능을 최대로 하는 적어도 하나의 최적의STA를 선택하는 제어부를 포함한다.

### 발명의 효과

[0008] 본 개시의 실시예들은 다음의 장점들을 포함하는 효과를 가질 수 있다. 다만, 본 개시의 실시예들이 이를 전부 포함하여야 한다는 의미는 아니므로, 본 개시의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될

것이다.

[0009] 본 개시의 일 실시예에 따르면 강화학습 네트워크를 이용하여 통신 성능을 최대로 하는 최적의 사용자를 선택하는 효과가 있다.

[0010] 또한, 본 개시의 일 실시예에 따르면, 강화학습 네트워크를 이용하여 계산 복잡도를 감소시키는 효과가 있다.

**도면의 간단한 설명**

[0011] 도 1은 본 개시의 일 실시예에 따른 강화학습을 이용한 채널 정보가 제한된 네트워크에서 기지국의 제어 프로토콜을 나타낸 도면이다.

도 2a는 본 개시의 일 실시예에 따른 총 주파수 대역이 20MHz인 경우 RU 위치의 예시를 나타낸 도면이다.

도 2b는 본 개시의 일 실시예에 따른 사용자 선택 및 자원 할당의 예시를 나타낸 도면이다.

도 3은 본 개시의 일 실시예에 따른 AP의 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법에 대한 순서도를 나타낸 도면이다.

도 4는 본 개시의 일 실시예에 따른 상향링크 통신을 위한 강화학습 기반 사용자 선택을 수행하는 AP에 대한 블록도를 나타낸 도면이다.

도 5는 본 개시의 일 실시예에 따른 강화학습 네트워크를 나타낸 도면이다.

도 6은 본 개시의 일 실시예에 따른 데이터에 대한 손실의 변화를 나타낸 도면이다.

도 7은 본 개시의 일 실시예에 따른 통신 시스템에서 상향링크 통신을 위한 강화학습 네트워크의 통신 성능을 기존 방식과 비교하여 나타낸 도면이다.

**발명을 실시하기 위한 구체적인 내용**

[0012] 본 개시의 실시예들의 다음의 상세한 설명은 여기에 설명된 방법, 장치, 및/또는 시스템의 포괄적인 이해를 얻는 것을 돕기 위해 제공된다. 따라서, 본 명세서에 기술된 시스템(들), 장치(들) 및/또는 방법(들)의 다양한 변경들, 수정들 및 균등물들은 본 개시의 기술분야의 당업자에게 제안될 것이다. 또한, 명확성 및 간결함을 위하여 잘 알려진 기능들 및 구성들에 대한 설명은 생략될 수 있다.

[0013] "포함하다(comprise)", "포함하는(comprising)", 또는 그 변형된 용어들은 비배타적인 포함을 커버하도록 의도된 것으로, 예를 들어, 구성들 또는 단계들의 목록을 포함하는 설정, 장치, 또는 방법은 해당 구성들 또는 단계들만을 포함하지 않고, 그러한 설정 또는 장치 또는 방법에 명시적으로 나열되거나 고유하지 않은 다른 구성들 또는 단계들을 포함할 수 있다. 즉, "뿐만 포함하다"에 의해 진행되는 시스템, 장치(들), 또는 방법(들)에 포함된 하나 이상의 요소(들)은 추가적인 제한이 없는 경우 상기 시스템, 장치(들) 또는 방법(들)에서 다른 요소들 또는 추가적인 요소들의 존재를 배제하지 않는다.

[0014] 본 개시의 실시예들의 다음의 상세한 설명에서, 본 명세서의 일부를 형성하고, 본 개시에서 실시될 수 있는 특정 실시예들을 예시로서 도시하는 첨부 도면이 참조된다. 이러한 실시예들은 당업자가 본 개시를 실시할 수 있도록 충분히 상세하게 설명되고, 또한 다른 실시예들이 활용될 수 있고 본 개시의 범위를 벗어나지 않는 선에서 변경이 이루어질 수 있음을 이해해야 한다. 따라서, 다음의 설명은 제한적인 의미로 받아들여서는 안 된다.

[0015] 본 명세서에서 사용된 바와 같이, 단수 형태의 용어들은 문맥이 명백하게 달리 나타내지 않는 한 복수 형태도 나타내는 것으로 의도된다. 용어들 "포함한다(comprise)", "포함하는(comprising)", "포함한다(includes)" 및/또는 "포함하는(including)"은 본 명세서에서 사용될 때, 언급되는 특징들, 정수들, 단계들, 동작들, 요소들 및/또는 구성들의 존재를 특정하고, 하나 이상의 다른 특징들, 정수들, 단계들, 동작들, 요소들, 구성들, 및/또는 그룹들의 존재 또는 추가를 배제하지 않는 것으로 이해될 수 있다.

[0016] 본 개시의 실시예들의 다음의 상세한 설명은 무선 로컬 영역 네트워크(wireless local area network, WLAN)에 적용된다. 다음의 설명은 차세대 WLAN 방법(예: IEEE 802.11ax) 등에 적용될 수 있다. IEEE 802.11ax는 기존의 IEEE 802.11a/b/g/n/ac와의 호환성을 유지한다. 다음의 설명은 IEEE 802.11ax 환경에서 실행될 수 있고, 또한 종래의 11a/b/g/n/ac 환경에서 호환된다.

[0017] WLAN에서, 단일 기본 서비스 집합(basic service set, BSS)은 단일 액세스 포인트(access point, AP) 스테이션

(station, STA) 및 복수의 비-AP STA들인 두 종류의 엔티티들로 구성된다. STA들은 WLAN 동작 대역폭 옵션들(예: 20/40/80/160MHz) 중 하나와 동일한 무선 주파수 채널을 공유한다. 이하에서는 AP STA 및 비-AP STA는 각각 AP 및 STA로 지칭될 수 있다.

- [0018] WLAN은 BSS 내에서 특정 규칙에 따라 스케줄링되는 다중 사용자들의 동시 송수신 동작을 포함한다. 여기서, 다중 사용자(multi-user, MU) 전송은 특정 BSS에 포함된 서로 다른 자원들을 기반으로 복수의 STA들로(예: 하향링크(downlink, DL) MU)/로부터(예: 상향링크(uplink, UL) MU) 동시에 전송되는 것을 의미한다. 예를 들어, 아래 도 1과 같이, 서로 다른 자원들은 직교 주파수 분할 다중 접속(orthogonal frequency division multiplexing access, OFDMA)에서 서로 다른 주파수 자원들이거나, MU 전송의 DL-OFDMA, DL-MU-MIMO, UL-OFDMA 및 UL-MU-MIMO를 사용하는 다중 사용자 다중 입력 다중 출력(MU-multiple input multiple output, MU-MIMO) 전송에서 서로 다른 공간 스트림들일 수 있다.
- [0019] 도 1은 본 개시의 일 실시예에 따른 무선 통신 시스템에서 상향링크 통신을 위한 강화학습 기반 AP의 사용자 선택 프로토콜을 나타낸 도면이다.
- [0020] 도 1을 참조하면, 무선 통신 시스템 환경(100)(예: WLAN BSS)은 AP(110) 및 복수의 STA들(120)을 포함할 수 있다. AP(110)는 복수의 STA들(120)로부터 채널 상태 정보를 수신하고, 채널 상태 정보에 기반하여 복수의 STA들(120) 중 AP(110)와 상향 링크 통신을 수행할 하나 이상의 STA를 선택할 수 있다. 선택 가능한 STA의 최대 수는 AP(110)의 안테나 수에 기반한다. 채널 상태 정보는 AP(110) 및 복수의 STA들(120) 간의 채널 이득 정보 및/또는 SNR 정보를 포함할 수 있다.
- [0021] 한편, AP(110)는 강화학습 네트워크를 설계할 수 있다. 강화학습 네트워크는 AP(110)가 복수의 STA들(120) 중 AP(110)와 최적의 상향링크 통신을 수행할 선택된 하나 이상의 STA를 포함하는 집합을 예측할 수 있다. 예를 들어, AP(110)는 임의의 시간  $t$ 에서 복수의 STA들(120)로부터 상태 정보(state)  $S_t$ (예: 채널 이득 정보, SNR 정보 등)를 수신하고, 상태 정보  $S_t$ 를 기반으로 DNN(deep neural network)을 통해 정책(policy)  $\pi^\theta$ 에 따른 행동(action)  $A_t$ (예: 사용자 선택, 자원 할당 등)를 결정한다. AP(110)는 결정된 행동  $A_t$ 에 따라 복수의 STA들과 상향링크 통신을 수행함으로써 보상(reward)  $R_t$ (예: 데이터 처리율)를 획득하고, 임의의 시간  $t+1$ 에서 복수의 STA들(120)로부터 상태 정보  $S_{t+1}$ 을 수신한다. AP(110)는 획득된 보상  $R_t$  및  $S_{t+1}$ 를 기반으로 DNN을 학습하여 최종적으로 최적의 사용자 선택 기능을 수행하는 정책  $\pi^\theta$ 을 결정할 수 있다.
- [0022] 도 2a는 본 개시의 일 실시예에 따른 총 주파수 대역이 20MHz인 경우 RU 위치의 예시를 나타낸 도면이고, 도 2b는 본 개시의 일 실시예에 따른 사용자 선택 및 자원 할당의 예시를 나타낸 도면이다.
- [0023] 도 2a를 참조하면, 총 주파수 대역이 20MHz일 때 총 주파수 대역은 242개의 서브캐리어(subcarrier)들로 이루어져 있고 양 끝에 각각 6개 서브캐리어들 및 5개의 서브캐리어들을 포함하는 가드(guard)를 가진다. 또한, 20MHz의 주파수 대역은 채널 환경에 따라 자유롭게 나누어 할당 가능하다. 예를 들어, 20MHz의 주파수 대역은 26개의 서브캐리어들로 이루어진 RU(resource unit)(이하, 26-RU) 9개로 구성될 수 있다. 구체적으로, 20MHz의 주파수 대역은 중앙에 7개의 서브캐리어들로 이루어진 DC(direct conversion)를 포함하고, DC를 기준으로 바로 양 옆에 각각 13개의 서브캐리어들이 하나의 26-RU를 이루고, 4개의 26-RU들이 각각 위치하여 총 9개의 26-RU를 포함한다. 또한, 20MHz의 주파수 대역은 9개의 26-RU들 간의 일부 경계에 null 서브캐리어들을 포함한다. 또 다른 예를 들어, 20MHz의 주파수 대역은 52개의 서브캐리어들로 이루어진 RU(이하, 52-RU) 4개와 26-RU 하나로 구성될 수 있다. 구체적으로, 20MHz 주파수 대역은 중앙에 7개의 서브캐리어들로 이루어진 DC를 포함하고, DC를 기준으로 바로 양 옆에 각각 13개의 서브캐리어들이 하나의 26-RU를 이루고, 2개의 52-RU들이 각각 위치하여 총 4개의 52-RU들과 하나의 26-RU를 포함한다. 또한, 20MHz의 주파수 대역은 4개의 52-RU들 간의 일부 경계에 null 서브캐리어들을 포함한다. 또 다른 예를 들어, 20MHz의 주파수 대역은 106개의 서브캐리어들로 이루어진 RU(이하, 106-RU) 2개와 26-RU 하나로 구성될 수 있다. 구체적으로, 20MHz 주파수 대역은 중앙에 7개의 서브캐리어들로 이루어진 DC를 포함하고, DC를 기준으로 바로 양 옆에 각각 13개의 서브캐리어들이 하나의 26-RU를 이루고, 1개의 106-RU들이 각각 위치하여 총 2개의 106-RU들과 하나의 26-RU를 포함한다.
- [0024] 도 2b를 참조하면, 총 주파수 대역이 20MHz일 때, RU의 위치를 기반으로 5개의 STA들 중 주파수 자원을 할당할 STA들의 선택 및 선택된 STA들에 할당된 주파수 자원을 나타낸다. 예를 들어, 5개의 STA들 중 주파수 자원을 할

당할 4개의 STA들로 STA1, STA2, STA3 및 STA4를 선택한다. 또한, 20MHz 주파수 대역이 도 2a에서와 같이 2개의 106-RU 및 하나의 26-RU를 포함할 때, STA1에게 106-RU인 RU1을 할당하고, STA2에게 26-RU인 RU2를 할당하고, STA3 및 STA4는 MU-MIMO 방식으로 106-RU인 RU3을 동시에 할당한다.

[0025] 도 2a 및 도 2b에서는 총 주파수 대역이 20MHz인 경우를 가정하였으나 본 개시의 실시예들은 총 주파수 대역이 40MHz, 80MHz 및 160MHz인 경우에도 적용 가능하며 각각의 주파수 대역에서의 RU의 위치는 본 개시의 적용을 받는 802.11ax의 규격에 개시된 내용을 참조할 수 있다.

[0027] 한편, 설명의 편의를 위해 N개의 안테나를 가지는 하나의 AP와 M개의 STA들이 통신을 수행하는 무선 통신 환경을 가정한다. 이 때, STA의 인덱스를  $i \in \{1, 2, \dots, M\}$  으로 나타내고, 사용자 선택의 최소 단위인 타임 슬롯은  $t \in \{1, 2, \dots, T\}$  로 나타낸다.

[0028] 본 개시의 실시예들에 따른 AP는 아래의 과정에 따라 보상이 최대가 되도록 강화학습 네트워크를 학습하는 과정을 거친다. 먼저 강화학습 네트워크의 상태 정보  $\mathbf{s}_t$  는 다음 수학적 식 1로 정의된다.

### 수학적 식 1

[0029] 
$$\mathbf{s}_t = [h_{1t}, h_{2t}, \dots, h_{Mt}]$$

[0030] 여기서,  $h_{it}$  는 시간 슬롯 t에서 AP와 i번째 STA간 채널 이득을 나타낸다.

[0031] 강화학습 네트워크에서 AP는 M개의 STA들 중 자원을 할당할 STA(들)를 선택하므로 AP의 행동  $\mathbf{a}_t$  는 다음 수학적 식 2로 정의된다.

### 수학적 식 2

[0032] 
$$\mathbf{a}_t = A = \{x \mid x \text{는 선택된 STA의 인덱스, } 1 \leq x \leq M\}$$

[0033] 이 때, A에 대한 제약조건은 다음 수학적 식 3으로 정의된다.

### 수학적 식 3

[0034] 
$$1 \leq n(A) \leq \min(N, M)$$

[0035] 여기서,  $n(\cdot)$  는 집합에 포함된 원소의 수를 나타내는 함수를 의미한다.

[0036] 한편, AP가 선택된 STA(들)로부터 상향링크를 통해 시간 슬롯 t에서 수신하는 신호  $\mathbf{y}_t$  는 다음 수학적 식 4로 정의된다.

### 수학적 식 4

[0037] 
$$\mathbf{y}_t = \sum_{i \in A} s_{it} \mathbf{x}_{it} + \mathbf{n} = \sum_{i \in A} h_{it} \mathbf{x}_{it} + \mathbf{n} = H\mathbf{x} + \mathbf{n}$$

[0038] 여기서,  $\mathbf{x}_{it}$  는 시간 슬롯 t에서 A에 포함된 i번째 STA가 상향링크를 통해 송신하는 심볼을 나타내고,  $\mathbf{x}$  는  $\mathbf{x}_{it}$  로 표현되는 심볼들의 행렬을 나타내고, 결국 AP로부터 선택된 STA(들)의 심볼(들)만이 AP에게 수신된다. 또한,

H는 시간 슬롯 t에서 A에 포함된 i번째 STA와 AP 간의 채널 이득  $h_{it}$ 의 행렬(또는 시간 슬롯 t에서의 상태 정보  $s_{it}$ 의 행렬)로 AP와 AP로부터 선택된 STA(들) 간의 채널 이득만을 포함한다. 또한, n은 노이즈(noise)를 나타내며 가우시안 노이즈(Gaussian noise)에 해당된다.

[0039] 한편, 본 개시의 실시예들에 따른 AP는 간섭 제거(zero-forcing) 수신기를 사용한다고 가정한다. 간섭 제거 수신기 W는 다음 수학식 5와 같이 정의된다.

**수학식 5**

[0040] 
$$W = (H^H H)^{-1} H^H$$

[0041] 여기서, 행렬 B에 대한  $B^H$ 는 행렬 B의 헤르미티안 행렬을 나타낸다.

[0042] AP가 간섭 제거 수신기 W를 통해 시간슬롯 t에서 수신한 상량링크 신호  $\hat{y}_t$ 는 다음 수학식 6과 같이 정의된다.

**수학식 6**

[0043] 
$$\hat{y}_t = W y_t = W H x + W n = (H^H H)^{-1} H^H H x + (H^H H)^{-1} H^H n = x + (H^H H)^{-1} H^H n$$

[0044] 따라서, 간섭 제거 수신기를 통해 선택된 STA(들) 간의 간섭을 제거할 수 있다. 이 때 노이즈 파워  $P_w$ 는 다음 수학식 7과 같이 정의된다.

**수학식 7**

[0045] 
$$P_w = E[(H^H H)^{-1} H^H m^H H (H^H H)^{-1}] = (H^H H)^{-1} H^H E[m^H] H (H^H H)^{-1} = \sigma^2 (H^H H)^{-1} H^H H (H^H H)^{-1} = \sigma^2 (H^H H)^{-1}$$

[0046] 여기서,  $\sigma$ 는 가우시안 노이즈 n의 표준편차를 나타낸다.

[0047] A에 포함된 i번째 STA의 SINR(signal-to-interference-plus-noise-ratio)는 다음 수학식 8과 같이 정의된다.

**수학식 8**

[0048] 
$$SINR_{it} = \frac{P_{si}}{\sigma^2} \text{diag}_{i+1}((H^H H)^{-1}), i \in A$$

[0049] 여기서,  $P_{si}$ 는 A에 포함된 i번째 STA가 상량링크를 통해 전송하는 신호의 세기를 나타내고,  $\text{diag}_{i+1}(B)$ 는 행렬 B의 i+1번째 대각 성분을 나타낸다. 예를 들어,  $P_{si}$ 는 모두 동일한 신호 세기를 가질 수 있다.

[0050] 다음으로 본 개시에 따른 강화학습 네트워크의 행동을 주파수를 할당 받을 사용자(또는 해당 STA) 선택으로 설정하였기 때문에, 본 개시에 따른 강화학습 네트워크의 보상은 선택된 사용자(또는 해당 STA)의 SINR을 기반으로 한 합계 전송률(sum-rate)로서 다음 수학식 9와 같이 정의된다.

수학식 9

$$r_t = \sum_{i \in A} BW_{it} \log_2(1 + SINR_{it}) = \sum_{i \in A} BW_{it} \log_2\left(1 + \frac{P_{si}}{\sigma^2} \text{diag}_{i+1}((H^H H)^{-1})\right)$$

[0051]

[0052]

여기서,  $BW_{it}$ 는 시간 슬롯  $t$ 에서  $A$ 에 포함된  $i$ 번째 STA에게 할당된 주파수 대역으로 AP에 의해 선택된 STA(들)의 주파수 대역만을 나타낸다.  $BW_{it}$ 는  $A$ 에 포함된  $i$ 번째 STA가 전송한 상태 정보 및 총 주파수 대역에 포함된 RU의 위치를 고려하여 할당될 수 있다.

[0053]

AP는 수학식 1, 수학식 2, 및 수학식 9에서 정의된 강화학습 네트워크의 상태, 행동, 보상 정보를 학습하여 합계 전송률을 최대로 하는 최적의 사용자 선택(또는 STA 선택)을 수행할 수 있다.

[0055]

도 3은 본 개시의 일 실시예에 따른 AP의 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법에 대한 순서도를 나타낸 도면이다.

[0056]

도 3을 참조하면, AP의 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법(300)은 310 내지 330 단계를 포함한다. AP의 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법(300)은 AP의 제어부(또는 프로세서)를 통해 수행될 수 있다.

[0057]

310 단계에서 AP는 복수의 STA들로부터 상태 정보를 수신한다. 상태 정보는 AP 및 복수의 STA들 간 채널 이득 정보 및/또는 복수의 STA들의 SNR 정보를 포함한다. AP는 복수의 STA들로부터 상태 정보를 수신하기 위해 복수의 STA들에게 참조 신호를 전송할 수 있고, 복수의 STA들은 수신된 참조 신호를 기반으로 상태 정보를 AP에 전송할 수 있다.

[0058]

320 단계에서 AP는 수신한 상태 정보를 이용하여 강화학습 네트워크를 통해 사용자(또는 STA) 선택을 예측한다. 예컨대, 강화학습 네트워크에 수신한 상태 정보를 입력하여 사용자(또는 STA) 선택을 예측할 수 있다. 선택 가능한 사용자(또는 STA)의 최대 수는 AP의 안테나 수에 기반한다. 강화학습 네트워크의 예측값은 합계 전송률을 고려하여 선택할 사용자(들)의 수 및/또는 선택할 사용자(들)을 결정하기 위한 값이다. 강화학습 네트워크는 DNN, CNN(convolution neural network), 또는 RNN(recurrent neural network)를 포함한다.

[0059]

330 단계에서 AP는 강화학습 네트워크의 예측값을 통해 최적의 사용자(또는 STA)(들)을 선택한다. 예컨대, 강화학습 네트워크의 예측값을 기준값으로 하여 AP는 통신 성능을 최대로 하는 주파수를 할당할 최적의 사용자(또는 STA)(들)을 선택할 수 있다. 또한, AP는 선택된 최적의 사용자(또는 STA)(들)과 상향링크 통신을 수행할 수 있다.

[0061]

도 4는 본 개시의 일 실시예에 따른 상향링크 통신을 위한 강화학습 기반 사용자 선택을 수행하는 AP에 대한 블록도를 나타낸 도면이다.

[0062]

도 4를 참조하면, 상향링크 통신을 위한 강화학습 기반 사용자 선택을 수행하는 AP(400)는 통신부(410), 저장부(420) 및 제어부(430)를 포함한다.

[0063]

통신부(410)는 AP(400)와 복수의 STA들 간의 상향링크에 대한 상태 정보를 수신한다. 예컨대, 통신부(410)는 상태 정보를 수신하기 전 복수의 STA들에게 참조 신호를 전송하고, 복수의 STA들로부터 참조 신호에 기반하여 측정된 상태 정보를 수신할 수 있다.

[0064]

저장부(420)는 강화학습 네트워크를 저장한다. 저장부(420)는 AP(400)의 메모리로 구현될 수 있다. 저장부(420)는 학습이 완료된 강화학습 네트워크를 저장할 수도 있고, 초기 상태의 강화학습 네트워크를 저장하고 이후 학습 과정에 따라 저장된 강화학습 네트워크를 불러올 수도 있다.

[0065]

제어부(430)는 상태 정보를 강화학습 네트워크에 입력하여 사용자(또는 STA) 선택을 예측한다. 제어부(430)는 AP(400)의 CPU(central processing unit) 또는 AP(application processor)로 구현될 수 있다. 또한, 제어부

(430)는 강화학습 네트워크의 예측 결과를 기반으로 AP(400)가 주파수를 할당할 사용자(또는 STA) 선택을 제어할 수 있도록 한다.

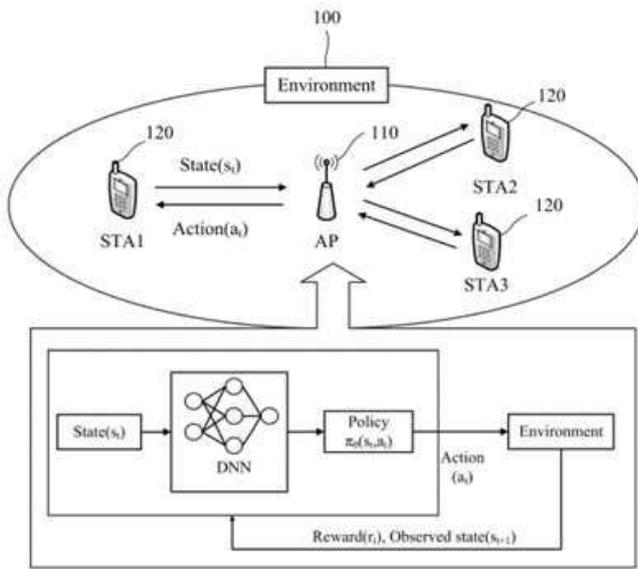
- [0066] 한편, 상술한 바와 같은 상향링크 통신을 위한 강화학습 기반 사용자 선택을 수행하는 AP(400)는 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.
- [0067] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM (read-only memory), PROM (programmable read only memory), EPROM(erasable PROM) 또는 EEPROM(electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.
- [0068] 일시적 판독 가능 매체는 스태틱 램(static RAM, SRAM), 다이내믹 램(dynamic RAM, DRAM), 싱크로너스 디램(synchronous DRAM, SDRAM), 2배속 SDRAM(double data rate SDRAM, DDR SDRAM), 증강형 SDRAM(enhanced SDRAM, ESDRAM), 동기화 DRAM(synclink DRAM, SLDRAM) 및 직접 램버스 램(direct rambus RAM, DRRAM) 과 같은 다양한 RAM(random access memory)을 의미한다.
- [0070] 도 5는 본 개시의 일 실시예에 따른 강화학습 네트워크를 나타낸 도면이다. 도 5에서는 강화학습 네트워크로 DNN을 사용하여 설명한다. 이 때, AP의 안테나 수를 4로 하고, 복수의 STA들의 수를 10으로 하고, 각각의 STA들은 하나의 안테나를 가지는 것으로 가정한다. 또한, 총 주파수 대역폭은 20MHz이고, 이를 최대 4명에게 할당하는 방식을 가정한다.
- [0071] 도 5를 참조하면, DNN(500)은 하나의 입력 층(input layer), 복수의 은닉 층(hidden layer)들 및 하나의 출력 층(output layer)을 포함한다. 예를 들어, DNN(500)은 20MHz 주파수 대역이 9개의 26-RU들을 포함하는 경우 입력 층(501)은 각각 9개의 26-RU들에 대해 4개의 안테나를 포함하는 AP 및 10개의 STA들에 대한 채널 정보로 주어지고, 각 채널 정보는 복소수로 이루어져 있고 이를 실수와 허수부로 나눌 수 있다(9x10x4x2). 이를 통해 입력 층(501)의 채널 정보들을 평평하게 하여 720x1 크기의 채널 정보를 포함하는 플래튼(flatten) 층(502)으로 구성한다. 이후 플래튼 층(502)의 채널 정보를 5개의 선형(linear)과 구멍난(leaky) ReLU(rectified linear unit)으로 구성된 은닉 층들(503,504,505,506,507)을 통과한다. 구체적으로 720x1 크기의 채널 정보를 포함하는 플래튼(flatten) 층(502)은 은닉 층 1(503), 은닉 층 2(504), 은닉 층 3(505) 및 은닉 층 4(506)를 통과하면서 12320x1 크기로 변환되고, 은닉 층 5(507)을 통과하면서 1540x1 크기로 변환된다. 결과적으로 1540x1 크기의 데이터를 재구축(reshape)하여 385x4 크기의 재구축 층(508)을 구성한다. 여기서, 385는 행동 공간(action space)의 크기를 나타내고, 4는 최대 선택 가능한 사용자(또는 STA)(들)의 수를 나타낸다. 재구축된 데이터에 행의 합을 구하여 385x1 크기의 출력 층(509)을 구성한다.
- [0073] 도 6은 본 개시의 일 실시예에 따른 데이터에 대한 손실의 변화를 나타낸 도면이다. 도 6의 실험결과는 총 주파수 대역이 20MHz인 환경에서 AP의 안테나 수는 4개이고, 복수의 STA들의 수가 10개 일 때 최대 4개의 STA들에 대해 주파수 자원을 할당하는 방식을 가정하여 강화학습 네트워크를 서로 다른 30만개의 채널에 대해 학습한 결과들이다.
- [0074] 도 6을 참조하면, 데이터에 대한 손실(training loss)이 학습이 진행됨에 따라점점 더 낮아지는 것을 확인할 수 있다. 즉, 학습이 진행됨에 따라 강화학습 네트워크가 환경을 이해하고 학습 파라미터를 조절함으로써 데이터에 대한 손실이 낮아짐을 알 수 있다.
- [0075] 도 7은 본 개시의 일 실시예에 따른 통신 시스템에서 상향링크 통신을 위한 강화학습 네트워크의 통신 성능을 기존 방식과 비교하여 나타낸 도면이다. 도 7에서 기존 방식은 전역 탐색 기법 및 최대 SNR 기법을 포함한다. 또한, 도 7의 실험결과는 도 6과 마찬가지로 총 주파수 대역이 20MHz인 환경에서 AP의 안테나 수는 4개이고, 복수의 STA들의 수가 10개 일 때 최대 4개의 STA들에 대해 주파수 자원을 할당하는 방식을 가정하여 강화학습 네트워크를 서로 다른 30만개의 채널에 대해 학습한 결과들이다.

[0076] 도 7을 참조하면, 학습이 진행될수록 본 개시의 강화학습을 이용한 사용자 선택 방식이 기존의 최대 SNR 기법과 비교하여 더 나은 통신 성능을 나타내는 것을 확인할 수 있다.

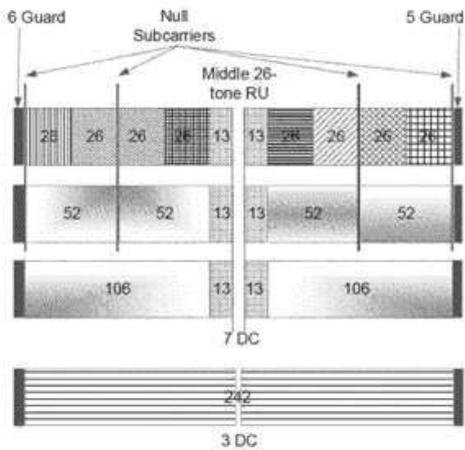
[0078] 본 개시의 일 실시예에 따른 상향링크 통신을 위한 강화학습 기반 사용자 선택 방법 및 장치는 이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 개시의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

도면

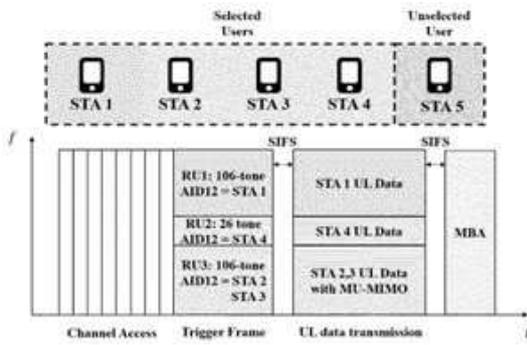
도면1



도면2a

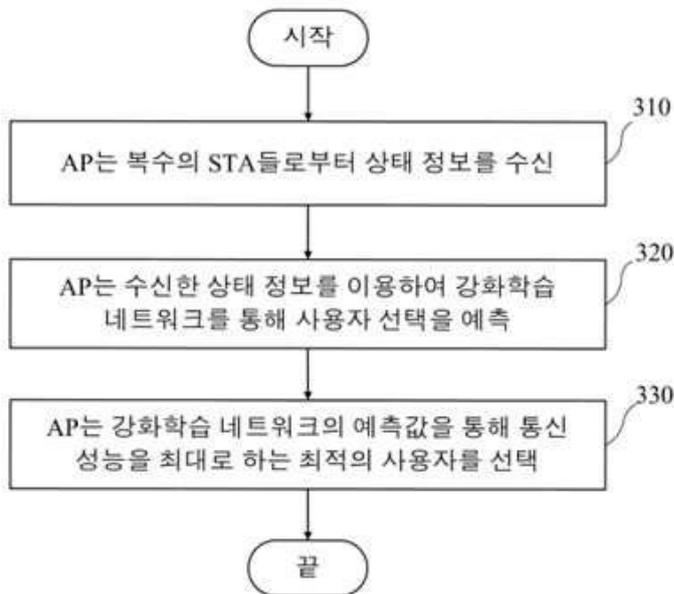


도면2b

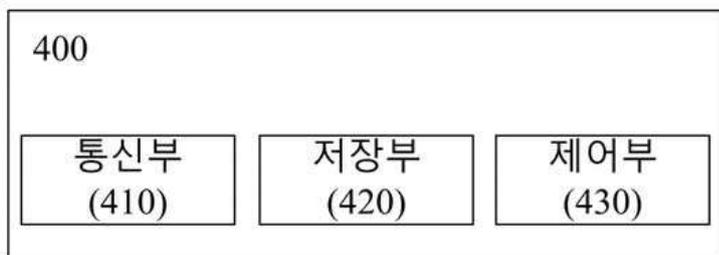


도면3

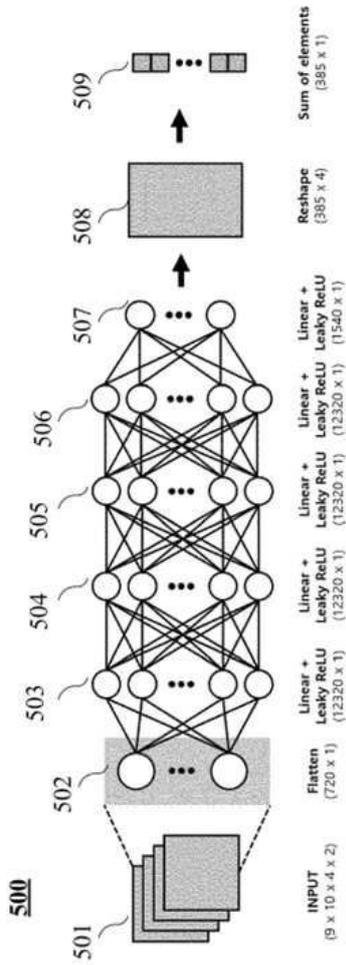
300



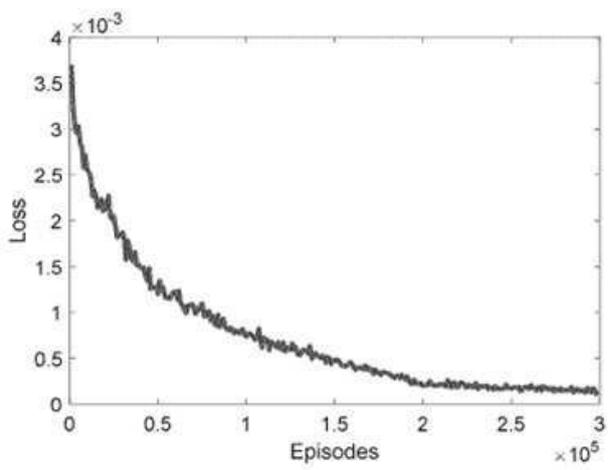
도면4



도면5



도면6



도면7

