(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0092294
(43) 공개일자 2023년06월26일

(51) 국제특허분류(Int. Cl.)
H04W 72/56 (2023.01) *G06N 3/08* (2023.01)
H04J 99/00 (2009.01) *H04W 72/04* (2009.01)
(52) CPC특허분류
H04W 72/56 (2023.01)
G06N 3/084 (2023.01)
(21) 출원번호 10-2021-0181564
(22) 출원일자 2021년12월17일
심사청구일자 2021년12월17일

(71) 출원인
조선대학교산학협력단
광주광역시 동구 필문대로 309 (서석동)
(72) 발명자
최우열
광주광역시 동구 필문대로 309 조선대학교 컴퓨터공학과
레즈완 시팟
광주광역시 동구 필문대로 309 조선대학교 컴퓨터공학과
(74) 대리인
특허법인지원

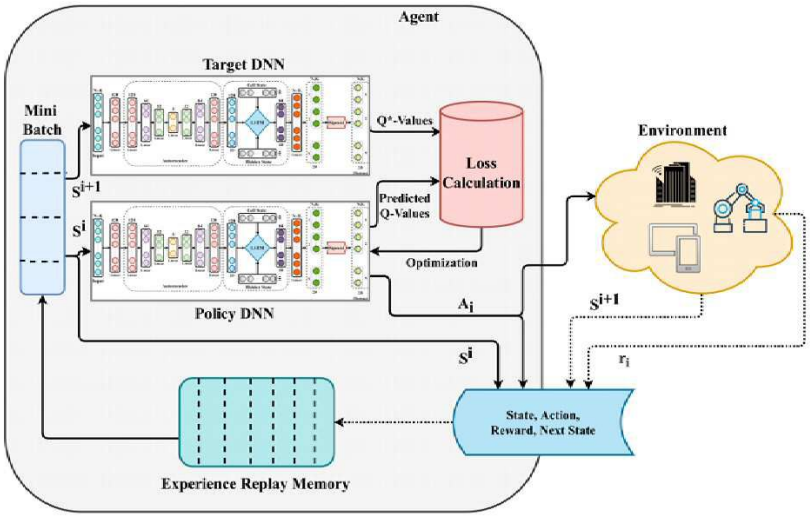
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **심층 Q-학습을 이용한 우선순위 기반 자원 할당 방법 및 장치**

(57) 요 약

자원 할당 방법 및 장치가 제공된다. 일실시예에 따른 자원 할당 방법은 적어도 하나의 디바이스에 전력을 할당하는 단계; 상기 적어도 하나의 디바이스의 우선순위를 결정하는 단계; 및 Q-학습(Q-Learning)을 이용하여 전력 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 상기 적어도 하나의 디바이스에 채널을 할당하는 단계를 포함할 수 있다.

대 표 도 - 도8



(52) CPC특허분류
H04J 99/00 (2022.08)
H04W 72/0473 (2013.01)
H04W 72/53 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711143923
과제번호	2019R1F1A1046687
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)(1311101501551200123430 1)
연구과제명	전이중 클라우드 무선 접속 네트워크를 위한 딥러닝 기반 빔형성 프로토콜 개발
기 여 율	1/1
과제수행기관명	조선대학교 산학협력단
연구기간	2021.03.01 ~ 2022.02.28

명 세 서

청구범위

청구항 1

비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 방법에 있어서,

(a) 적어도 하나의 디바이스에 전력을 할당하는 단계;

(b) 상기 적어도 하나의 디바이스의 우선순위를 결정하는 단계; 및

(c) Q-학습(Q-Learning)을 이용하여 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 상기 적어도 하나의 디바이스에 채널을 할당하는 단계

를 포함하는,

자원 할당 방법.

청구항 2

제1항에 있어서, 상기 단계 (c)는,

(d) 상기 Q-학습의 상태, 행동 및 보상에 대해, 상기 디바이스의 채널 대 잡음비를 상태로, 상기 채널 할당을 행동으로, 그리고 채널에 대한 합속도를 보상으로 각각 설정하는 단계;

(e) 현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하는 단계;

(f) 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하는 단계; 및

(g) 상기 단계 (e) 및 상기 단계 (f)를 반복 수행하면서 채널 할당 정책을 결정하는 단계

를 포함하는,

자원 할당 방법.

청구항 3

제1항에 있어서, 상기 단계 (a)는,

합속도(sum-rate)에 기초하여 적어도 하나 이상의 디바이스에 전력을 할당하는

자원 할당 방법.

청구항 4

제3항에 있어서, 상기 합속도는 채널에 대한 각 장치의 데이터전송률을 합산하여 계산되는 속도인

자원 할당 방법.

청구항 5

제3항에 있어서, 상기 전력을 할당하는 단계는, 미리 지정되는 임계값 이상으로 전력을 할당하는

자원 할당 방법.

청구항 6

제1항에 있어서, 상기 단계 (b)는,

상기 적어도 하나의 디바이스에 요구되는 통신 품질 요구사항을 기준으로 우선순위를 결정하는

자원 할당 방법.

청구항 7

제1항에 있어서, 상기 단계 (b)는,

상기 적어도 하나의 디바이스와 기지국과의 거리를 기준으로 우선순위를 결정하는

자원 할당 방법.

청구항 8

제2항에 있어서, 상기 적어도 하나의 디바이스는,

향상된 모바일 광대역(eMBB) 장치, 대규모 사물통신(mMTC) 장치 및 초고신뢰·저지연 통신(URLLC) 장치 중 적어도 하나 이상을 포함하는

자원 할당 방법.

청구항 9

비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 장치에 있어서,

적어도 하나의 디바이스의 우선순위를 결정하고, 전력 및 채널을 할당하는 할당부; 및

Q-학습(Q-Learning)을 이용하여 상기 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 합속도가 미리 지정되는 값 이상이 되도록하는 채널 할당 정책을 결정하는 Q-학습부

를 포함하는,

자원 할당 장치.

청구항 10

제9항에 있어서, 상기 Q-학습부는,

타겟 DNN에 의해 계산되는 Q*-값과 정책 DNN에 의해 계산되는 Q-값 사이의 차이를 범주형 교차 엔트로피 손실 함수를 이용하여 계산하고 아담 최적화기(Adam optimizer)를 사용하여 정책 DNN을 업데이트하는

자원 할당 장치.

청구항 11

제9항에 있어서, 상기 Q-학습부는,

상기 Q-학습의 상태, 행동 및 보상에 대해, 상기 디바이스의 채널 대 잡음비를 상태로, 상기 채널 할당을 행동으로, 그리고 채널에 대한 합속도를 보상으로 각각 설정하고,

현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하며, 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하여 채널 할당 정책을 결정하는

자원 할당 장치.

청구항 12

컴퓨터 장치로 하여금, 제1항 내지 제8항 중 어느 한 항의 방법을 수행하도록 구현된 명령어(instructions)를 포함하는, 기계 판독 가능한 비일시적 기록 매체에 저장된, 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 우선순위 기반 자원 할당 방법 및 장치에 관한 것으로, 더욱 상세하게는 심층 Q-학습을 이용한 우선순위 기반 자원 할당 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 사물인터넷(Internet of Things, IoT)과 클라우드 컴퓨팅의 대중화가 급격히 증가함에 따라 무선 통신 네트워크에 대한 높은 신뢰성의 데이터 속도와 대규모 연결에 대한 요구가 날로 증가하고 있다.

[0003] 이러한 요구를 충족시키기 위해 3GPP(3rd Generation Partnership Project)는 3가지 주요 서비스를 제공하는 5세대(5G) 무선 네트워크를 도입했다. 5세대 무선 네트워크의 주요 서비스에는 IoT 장치를 위한 대규모 연결을 지원하는 대규모 기계 유형 통신(massive Machine Type Communication, mMTC), 모바일 플랫폼을 위한 높은 데이터 속도를 제공하는 향상된 모바일 광대역(enhanced Mobile Broadband, eMBB), 그리고 매우 민감하고 중요한 애플리케이션을 위해 낮은 대기 시간과 안정성을 보장하는 초신뢰성 및 저지연 통신(Ultra-Reliable and Low-Latency Communication, URLLC) 등이 있다.

[0004] 이러한 서비스는 서비스 품질(quality-of-service, QoS) 측면에서 분류될 수 있으며, URLLC는 높은 신뢰성과 낮은 대기 시간에 대한 엄격한 QoS 정책을 갖고 eMBB 서비스는 중간 QoS 정책을 가지나, mMTC는 대규모 연결을 제외하고 특정한 QoS 정책을 갖지 않는다.

[0005] QoS 정책은 제한된 스펙트럼 리소스, 큰 전송 손실 및 긴 대기 지연으로 인해 기존의 직교 다중 액세스(Orthogonal Multiple Access, OMA)로 수행하기 어렵다. 따라서 다양한 QoS 요구 사항을 유지하기 위해 많은 기술이 5G 통신 네트워크에 도입되었으며, 그 중 비직교 다중 액세스(Non-orthogonal Multiple Access, NOMA)는 제한된 리소스, 매우 안정적인 전송, 낮은 전송 지연 및 높은 스펙트럼 효율성으로 대규모 연결을 지원할 수 있기 때문에 인기가 증가하고 있다.

[0006] 그러나 NOMA 시스템은 전력 할당 및 채널 할당을 포함하는 자원 할당과 관련한 문제를 갖는다. 예를 들어, 채널 할당과 전력 할당의 가능한 모든 조합은 최적의 솔루션에 도달할 것을 요구하게 되고, 그것은 시스템을 복잡하게 만들고 극도로 높은 계산 능력을 필요로 한다. 특히 다중 반송파(Multi-Carrier) NOMA 시스템의 경우 시스템이 더 복잡해질 수 있다.

[0007] 시스템 합 속도(system sum-rate)의 증가가 각 채널의 합 속도(channel sum-rate)를 반드시 증가시키는 것은 아니기 때문에, 다중 반송파 NOMA의 또 다른 문제는 채널 합 속도의 공정성이다. 모든 채널의 열악한 합 속도(sum-rate)는 해당 채널에 할당된 장치의 성능을 저하시킬 수 있다.

[0008] 또한 순차적 간섭 제거(Successive Interference Cancellation, SIC)를 사용한 완벽한 신호 디코딩과 5G 서비스의 QoS 요구 사항 충족도 전력 할당 및 채널 할당에 달려 있다. 불완전한 SIC와 부적절한 채널 할당은 시스템의 전체 성능을 쉽게 저하시킬 수 있다.

선행기술문헌

특허문헌

[0009] (특허문헌 0001) 한국공개특허 제10-2020-0081630호

발명의 내용

해결하려는 과제

[0010] 발명은 전술한 문제점을 해결하기 위하여 창출된 것으로, 우선순위 기반 딥러닝 모델을 이용한 자원 할당 방법 및 장치를 제공하는 것을 그 목적으로 한다.

[0011] 본 발명의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0012] 상기한 목적들을 달성하기 위하여, 본 발명의 일 실시예에 따른 비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 방법은, (a) 적어도 하나의 디바이스에 전력을 할당하는 단계; (b) 상기 적어도 하나의 디바이스의 우선순위를 결정하는 단계; 및 (c) Q-학습(Q-Learning)을 이용하여 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 상기 적어도 하나의 디바이스에 채널을 할당하는 단계를 포함할 수 있다.

[0013] 다른 일실시예에 따른 자원 할당 방법은 (d) 상기 Q-학습의 상태, 행동 및 보상에 대해, 상기 디바이스의 채널 대 잡음비를 상태로, 상기 채널 할당을 행동으로, 그리고 채널에 대한 합속도를 보상으로 각각 설정하는 단계; (e) 현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하는 단계; (f) 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하는 단계; 및 (g) 상기 단계 (e) 및 상기 단계 (f)를 반복 수행하면서 채널 할당 정책을 결정하는 단계를 포함할 수 있다.

[0014] 다른 일실시예에 따른 자원 할당 방법은 합속도(sum-rate)에 기초하여 적어도 하나 이상의 디바이스에 전력을 할당할 수 있다.

[0015] 또 다른 일실시예에 따른 자원 할당 방법에서 상기 합속도는 채널에 대한 각 장치의 데이터전송률을 합산하여 계산되는 속도일 수 있다.

[0016] 다른 일실시예에 따른 자원 할당 방법에서 상기 전력을 할당하는 단계는, 미리 지정되는 임계값 이상으로 전력을 할당하는 단계일 수 있다.

[0017] 다른 일실시예에 따른 자원 할당 방법은 상기 적어도 하나의 디바이스에 요구되는 통신 품질 요구사항을 기준으로 우선순위를 결정할 수 있다.

[0018] 또 다른 일실시예에 따른 자원 할당 방법은 상기 적어도 하나의 디바이스와 기지국과의 거리를 기준으로 우선순위를 결정할 수도 있다.

[0019] 일실시예에 따른 자원 할당 방법에서 상기 적어도 하나의 디바이스는, 향상된 모바일 광대역(eMBB) 장치, 대규모 사물통신(mMTC) 장치 및 초고신뢰·저지연 통신(URLLC) 장치 중 적어도 하나 이상을 포함할 수 있다.

[0020] 일실시예에 따르면 비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 장치에 있어서, 적어도 하나의 디바이스의 우선순위를 결정하고, 전력 및 채널을 할당하는 할당부; 및 Q-학습(Q-Learning)을 이용하여 상기 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 합속도가 미리 지정되는 값 이상이 되도록하는 채널 할당 정책을 결정하는 Q-학습부를 포함하는, 자원 할당 장치가 개시된다.

[0021] 다른 일실시예에 따르면 상기 Q-학습부는, 타겟 DNN에 의해 계산되는 Q*-값과 정책 DNN에 의해 계산되는 Q-값 사이의 차이를 범주형 교차 엔트로피 손실 함수를 이용하여 계산하고 아담 최적화기(Adam optimizer)를 사용하여 정책 DNN을 업데이트하는 자원 할당 장치도 개시된다.

[0022] 또 다른 일실시예에 따르면 상기 Q-학습부는, 상기 Q-학습의 상태, 행동 및 보상에 대해, 상기 디바이스의 채널 대 잡음비를 상태로, 상기 채널 할당을 행동으로, 그리고 채널에 대한 합속도를 보상으로 각각 설정하고, 현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하며, 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하여 채널 할당 정책을 결정하는 자원 할당 장치가 제시된다.

[0023] 일실시예에 따르면 컴퓨터 장치로 하여금, 청구항 제1항 내지 제8항 중 어느 한 항의 방법을 수행하도록 구현된 명령어(instructions)를 포함하는, 기계 판독 가능한 비일시적 기록 매체에 저장된, 컴퓨터 프로그램도 개시된다.

[0024] 상기한 목적들을 달성하기 위한 구체적인 사항들은 첨부된 도면과 함께 상세하게 후술될 실시예들을 참조하면 명확해질 것이다.

[0025] 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라, 서로 다른 다양한 형태로 구성될 수 있으며, 본 발명의 개시가 완전하도록 하고 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자(이하, "통상의 기술자")에게 발명의 범주를 완전하게 알려주기 위해서 제공되는 것이다.

발명의 효과

[0026] 본 발명의 일 실시예에 의하면, Q-학습을 사용하여 채널을 할당함으로써, 합속도가 큰 최적의 자원 할당 방법을 구할 수 있다.

[0027] 또한, 본 발명의 일 실시예에 의하면, Q-학습 모델은 DNN을 사용함으로써, 비지도 반복 학습시키면서 자원 할당 효율을 증가시킨다.

[0028] 본 발명의 효과들은 상술된 효과들로 제한되지 않으며, 본 발명의 기술적 특징들에 의하여 기대되는 잠정적인

효과들은 아래의 기재로부터 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0029] 도 1은 일 실시 예에 따른 NOMA 시스템을 도시한다.

도 2는 일 실시 예에 따른 SISO(Single Input Single Output)-NOMA 시스템을 도시한다.

도 3은 3종 디바이스 NOMA 시스템에 채널을 할당하는 예시를 도시한다.

도 4는 일 실시 예에 따른 Q-학습 개념도이다.

도 5는 일 실시 예에 따른 DNN 구조를 도시한다.

도 6은 일 실시 예에 따른 오토인코더를 도시한다.

도 7은 일 실시 예에 따른 LSTM 셀을 도시한다.

도 8은 일 실시 예에 따른 DQL 프레임워크를 도시한다.

도 9는 일 실시 예에 따른 DQL 프레임워크의 알고리즘을 도시한다.

도 10은 일 실시예에 따른 심층 Q-학습을 이용한 우선순위 기반 자원 할당 방법을 도시한다.

도 11은 일 실시예에 따른 심층 Q-학습을 이용한 우선순위 기반 자원 할당 장치를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 발명은 다양한 변경을 가할 수 있고, 여러 가지 실시예들을 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 이를 상세히 설명하고자 한다.

[0031] 청구범위에 개시된 발명의 다양한 특징들은 도면 및 상세한 설명을 고려하여 더 잘 이해될 수 있을 것이다. 명세서에 개시된 장치, 방법, 제법 및 다양한 실시예들은 예시를 위해서 제공되는 것이다. 개시된 구조 및 기능상의 특징들은 통상의 기술자로 하여금 다양한 실시예들을 구체적으로 실시할 수 있도록 하기 위한 것이고, 발명의 범위를 제한하기 위한 것이 아니다. 개시된 용어 및 문장들은 개시된 발명의 다양한 특징들을 이해하기 쉽게 설명하기 위한 것이고, 발명의 범위를 제한하기 위한 것이 아니다.

[0032] 본 발명을 설명함에 있어서, 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우, 그 상세한 설명을 생략한다.

[0033] 이하, 본 발명의 일 실시예에 따른 우선순위 기반 심층 Q-학습을 이용한 자원 할당 방법 및 장치를 설명한다.

[0034] 도 1은 일 실시 예에 따른 다중 반송파(Multi-carrier) NOMA 시스템을 도시한다.

[0035] NOMA 시스템은 업링크(uplink) 및 다운링크(downlink) 전송 모두에 대해 전력 도메인을 활용하는 동일한 무선 자원 블록(Radio Resource Block, RRB)을 사용하여 여러 장치에 서비스를 제공할 수 있다. 간단한 다운링크 다중 반송파 NOMA 시스템에서 기지국(Base Station, BS)은 무선 채널을 통해 동시에 다른 유형의 장치를 서비스한다.

[0036] 도 1은 3개의 서로 다른 장치로 구성되는 5G 네트워크의 시나리오를 도시한다. 기지국은 3개의 장치마다 하나의 채널을 할당하며, 여기서 3개의 장치의 신호들은 서로 다른 전력 레벨에서 다중화(multiplexed)된다. 따라서 이 장치들은 원하는 신호와 함께 해당 채널의 나머지 두 장치의 신호를 잡음 또는 간섭으로 수신한다. 원하는 신호의 전력 레벨이 높으면 원하지 않는 신호는 노이즈로 작용하고, 그렇지 않으면 원하지 않는 신호는 간섭으로 작용한다.

[0037] 원하는 신호를 디코딩하기 위해 각 장치는 순차적 간섭 제거(Successive Interference Cancellation, SIC) 기술을 사용한다. 순차적 간섭 제거는 가장 높은 전력으로 신호를 디코딩하고 원하는 신호를 디코딩할 때까지 메인 신호에서 해당 신호를 제거한다. 완벽한 순차적 간섭 제거는 신호 대 간섭 잡음비(Signal to Interference plus Noise Ratio, SINR)과 같은 채널 상태 정보(CSI)에 의존하고, 신호대 간섭 잡음비는 채널 할당 및 전력 할당에 의존한다. 이 경우 채널에 대한 각 장치의 데이터 전송률(data rate)은 아래의 식(1)을 사용하여 계산할 수 있다.

수학식 1

[0038] $$R_i^k = \log_2(1 + \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}), \qquad k, i = 1, 2, 3,$$

[0039] 여기서 Γ은 할당된 채널 k에 대한 채널 대 잡음비(Channel to Noise Ratio, CNR)이고 P는 할당된 전력이다.

[0040] 도 2는 일실시예에 따른 기지국과 5G가 지원 가능한 장치로 구성된 5G 네트워크의 마이크로 셀(micro-cell)을 설명한다. 보다 상세하게는 도 2에서 총 장치의 개수는 N이고 채널 수는 K인, 단일 입력 및 단일 출력(Single-Input and Single-Output, SISO) NOMA 시스템의 다운링크를 도시한다.

[0041] 도 2는 서로 다른 5G 네트워크의 서비스를 요구하는 3 종의 장치들이 도시한다. 구체적으로 eMBB 장치(UE_1, UE_2, …, UE_e), URLLC 장치(UL_1, UL_2, …, UL_l) 및 mMTC 장치(MC_1, MC_2, …, MC_m)가 존재한다.

[0042] 예시적으로 그러나 한정되지 않게, 총 가용 대역폭(BW_t)이 180kHz의 채널 대역폭(BW_{ch})을 갖는 모든 채널로 분할되는 것을 가정한다. 채널당 최대 장치 개수는 n이며, n의 범위는 $2 \leq n \leq N$이고, 총 채널 수는 K이며, K = ceil(N/n) 이다.

[0043] 일실시예에 따르면 완전한 채널 상태 정보(CSI)를 가정하지만, 실제 무선 환경을 고려하여 불완전한 CSI도 가정한다.

[0044] k번째 채널이 n개의 디바이스에 할당되었다고 가정하고, 여기서 n번째 디바이스에 할당된 전력은 P_n이고 n번째 디바이스의 원하는 신호는 x_n 이다. n개의 장치의 신호를 결합한 후, 기지국은 k 번째 채널을 통해 다음의 수학식 2와 같은 신호를 전송한다.

수학식 2

[0045] $$X^k = \sum_{i=1}^{n} \sqrt{P_i^k} x_i, \qquad i = 1, 2, \ldots, n$$

[0046] 디바이스 단에서 전송된 신호는 경로 손실 성분과 추가 백색 가우스 잡음(Additive White Gaussian Noise, AWGN)으로 도달한다. 도달된 신호는 다음의 수학식 3과 같이 표현될 수 있다.

수학식 3

[0047] $$y^k = \sum_{i=1}^{n} \sqrt{P_i^k} h_i^k x_i + w^k, \qquad i = 1, 2, \ldots, n$$

[0048] 여기서 h^k_i는 i번째 디바이스의 채널 이득(gain)이고 w_k는 추가 백색 가우스 잡음(AWGN)을 의미하며, 상기 추가 백색 가우스 잡음은 온도 잡음 전력 분산 σ_k를 포함할 수 있다.

[0049] 신호를 수신한 후 수신기는 순차적 간섭 제거(SIC) 기술을 사용하여 신호를 디코딩(decode)한다. 완벽한 순차적 간섭 제거는 통신에 사용되는 해당 채널의 장치 신호 대 간섭 잡음비(SINR)에 의존한다. k번째 채널에 대한 n번째 장치의 채널 대 잡음비(CNR)가 다음과 같다고 가정한다.

수학식 4

[0050] $$\Gamma_n^k = \frac{|h_i|^2}{\sigma_k}$$

[0051] 상술한 것과 같이 채널의 장치에 서로 다른 전력 레벨이 할당될 수 있다. NOMA에 따라, 채널 대 잡음비가 가장 낮은 장치에 가장 높은 전력이 할당되고 그 역도 성립한다. 예를 들어, 채널 대 잡음비가 $\Gamma^k_1 > \Gamma^k_2 > \cdots > \Gamma^k_n$ 인 장치의 경우전력이 $P^k_1 < P^k_2 < \cdots < P^k_n$ 으로 각각 할당된다. 따라서 특정 채널의 각 장치에 대한 신호 대 간섭 잡음비(SINR) 및 데이터 전송률(data rate)은 각각 수학식 5와 앞서 설명한 수학식 1로 표현될 수 있다.

수학식 5

[0052] $$\gamma_i^k = \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}, \qquad k, i = 1, 2, \ldots, n$$

[0053] 완벽한 순차적 간섭 제거를 수행하기 위해, 기지국은 아래의 수학식 6과 같이 특정 임계값(P_{th}) 이상으로 각 디바이스에 전력을 할당한다. 예를 들어, 낮은 채널 대 잡음비를 갖는 장치는, 순차적 간섭 제거 기술을 완벽하게 완성하기 위해, 높은 채널 대 잡음비를 갖는 다른 장치들의 전력의 합보다 더 높은 전력을 가져야 한다.

수학식 6

[0054] $$\left(P_i^k - \left(\sum_{j=1}^{i-1} \sqrt{P_j^k}\right)\right) \Gamma_d^k \geq P_{th}, \quad i = 1, 2, \ldots, n, \quad d = n, \ldots, 2, 1, \quad k = 1, 2, \ldots, K.$$

[0055] 각 디바이스는 채널 할당을 위한 채널의 집합 $\Gamma_N = \{\Gamma^1_N, \Gamma^2_N, \cdots, \Gamma^k_N\}$ 및 전력 범위 $P_N \in [0.01, 0.99] \times P_T$ 를 갖고, 여기서 P_T는 전력 할당을 위한 채널당 총 전력 예산이다. 일실시예에서 NOMA 시스템의 채널 할당 및 전력 할당의 최적화를 위한 핵심 성능 지표인 합속도(sum-rate)에 중점을 둔다.

수학식 7

[0056] $$R_{sum} = \sum_{1}^{K} \sum_{i=1}^{n} \log_2\left(1 + \frac{P_i \Gamma_i^k}{\sum_{j=1}^{i-1} P_j \Gamma_i^k + 1}\right), \qquad i = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, K.$$

[0057] 모든 디바이스의 최소 데이터 전송률(minimum data rate) 요구조건은 다음의 수학식 8과 같이 표현될 수 있다.

수학식 8

[0058] $$\log_2\left(1 + \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}\right) \geq R_i^k, \qquad i = 1, 2, \ldots, n, \quad k = 1, 2, \ldots, K.$$

[0059] 채널의 디바이스당 전력 합계는 P_T보다 작거나 같아야 하며 다음의 수학식 9와 같이 표현될 수 있다.

수학식 9

[0060] $$\sum_{i=1}^{n} P_i^k \le P_T, \qquad k = 1, 2, \ldots, K.$$

[0061] 이하에서는 최적의 전력 할당 방식을 도출하고, 디바이스들 사이의의 공정성과 시스템 성능 향상을 보장하기 위하여 5G 서비스의 QoS, 합속도 최대화(Maximizing Sum-Rate, MSR), 그리고 채널 합속도 최대화(Maximizing Channel Sum-Rate, MCSR)를 유지하기 위한 심층 Q-학습(Deep Q-learning, DQL) 알고리즘을 이용한 우선 순위 기반 채널 할당 방법을 설명한다. 구체적으로, DQL은 채널 할당을 평가하고 DNN을 훈련하기 위해 전력 할당이 필요하므로 먼저 주어진 채널에 대한 전력 할당 솔루션을 설명하고, 다음으로 NOMA 시스템에 대한 최적의 솔루션을 얻기 위해 우선순위 기반 채널 할당을 위한 DQL 프레임워크를 설명한다.

[0062] 일실시예에 따른 NOMA의 다양한 제약 조건을 고려하면서 최대 합속도(Maximum Sum-Rate)와 시스템 효율성을 높이기 위해 주어진 채널에 대한 최적의 전력 할당을 설명한다. 예시적으로 기지국로부터의 거리에 따라 내림차순으로 디바이스를 정렬할 수 있다. 주된 목적은 합속도(sum-rate)를 최대화하는 것이므로 수학식 6, 8, 9를 고려하여 주어진 채널 k에 대한 볼록 함수(convex function)를 최대화하는 것으로 수학식 7로 나타낼 수 있으며, 이는 다음의 수학식 10과 같이 공식화될 수 있다.

수학식 10

[0063] $$maximize\ P_i^k, \sum_{i=1}^{n} \log_2\left(1 + \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}\right),$$

[0064] $$\text{subject to } \log_2\left(1 + \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}\right) \ge R_i^k,$$

[0065] $$\sum_{i=1}^{n} P_i^k \le P_T,$$

[0066] $$\left(P_i^k - \left(\sum_{j=1}^{i-1} P_j^k\right)\right)\Gamma_d^k \ge P_{th}, \quad \forall i = 1, 2, \ldots, n; \quad d = n, \qquad \ldots, 2, 1.$$

[0067] 상기 수학식 10의 볼록 문제(convex problem)는 다음의 수학식 11과 같이 라그랑주 형식(Lagrangian form)으로도 표현될 수 있다.

수학식 11

[0068] $$\mathcal{L}(P, \tau, \nu, \psi) = \sum_{i=1}^{n} \log_2\left(1 + \frac{P_i^k \Gamma_i^k}{\sum_{j=1}^{i-1} P_j^k \Gamma_i^k + 1}\right)$$

[0069] $$= \tau^k\left(P_T - \sum_{i=1}^{n} P_i^k\right) + \sum_{i=1}^{n} \nu_i^k \left\{P_i^k \Gamma_i^k - \left(\sum_{j=1}^{i-1} P_j^k \Gamma_i^k - 1\right) \times \left(\varphi_i^k - 1\right)\right\} + \sum_{i=2}^{n} \psi_i^k \left(P_i^k \Gamma_d^k - \sum_{l=1}^{i} P_l^k \Gamma_d^k - P_{Th}\right)$$

[0070] 여기서 τ, v 및 Ψ는 라그랑주 승수(Lagrange multiplier)이고, ∀i = 1, 2, …, n 이며, $\phi^{k}_{i} = 2^{Rki/KBWch}$ 이다.

[0071] P_i, τ, v 및 Ψ에 대해 수학식 11을 미분하면, 다중 KKT(Karush-Kuhn-Tucker) 조건을 얻을 수 있다. n개의 디바이스를 갖는 NOMA의 경우 2n개의 라그랑주 승수가 있으므로 2^{2n}개의 조합이 생성된다. 예를 들어, n = 2, 3, 4, …, 8 인 경우에 조합의 수는 각각 16, 64, 256, …, 65536 이다. 그러나 모든 유형의 조합을 확인하는 것은 계산적으로 가능하지 않다. 따라서 2, 3, 4개의 디바이스를 갖는 NOMA에 대해 n개의 방정식만 풀면, KKT 조건을 각각 만족하는 2, 4, 8 조합을 찾을 수 있다. 따라서 주어진 채널 k에 대한 n개의 디바이스를 갖는 NOMA에 대한 전력 할당의 폐쇄형 솔루션은 거의 최적이며, 다음의 수학식 12와 같이 나타낼 수 있다.

수학식 12

[0072]
$$P_x = \frac{P_T}{2^{(n-1)}} + \frac{(x-1)P_{th}}{2^{(x-1)}\Gamma_{(x-1)}} - \left(\sum_{i=x}^{n-1}\frac{P_{th}}{2^{i}\Gamma_i}\right),$$

[0073]
$$P_j = \frac{P_T}{2^{(n-q-2)}} + \frac{P_{th}}{2\Gamma_{(j-1)}} - \left(\sum_{i=j}^{n-1}\frac{P_{th}}{2\Gamma_i}\right)$$

[0074] 여기서 x = 1, 2 이고, j = 3, 4, …, n 이고, q = 0, 1, …, (n - 3) 이다. 그리고 디바이스들은 각각 $P^{k}_{1} < P^{k}_{2} < \cdots < P^{k}_{n}$ 인 전력과 함께 $\Gamma^{k}_{1} > \Gamma^{k}_{2} > \cdots > \Gamma^{k}_{n}$ 인 채널 대 잡음비(CNR)를 갖는다.

[0075] 이하에서는 일실시예에 따른 DQL을 이용한 우선순위 기반 채널 할당 기법을 보다 상세하게 설명한다.

[0076] 합속도 최대화(Maximizing Sum-Rate, MSR), 채널 합속도 최대화(Maximizing Channel Sum-Rate, MCSR) 및 우선순위를 기반으로 채널 할당 문제를 공식화하고, 채널 할당 문제를 강화 작업(Reinforcement task)으로 모델링하고, DQL 프레임워크를 생성하기 위한 LSTM 네트워크가 따르는 오토인코더를 설명한다. 마지막으로 최적에 가까운 전력 할당 솔루션을 사용하고 검증을 위한 DNN 학습을 설명한다.

[0077] 5G 무선 네트워크는 QoS 요구 사항이 다른 세 가지 서비스를 제공한다. URLLC 서비스는 QoS 요구 사항이 가장 높고, eMBB 서비스는 평균 QoS 요구 사항을 가지며, mMTC 서비스는 가장 낮은 QoS 요구 사항을 갖는다. 따라서 사용 중인 서비스와 QoS 요구 사항을 기반으로 네트워크 디바이스의 우선 순위를 지정할 수 있으며, URLLC 디바이스에 가장 높은 우선 순위를, eMBB 디바이스에 두 번째로 높은 우선 순위를, 그리고 mMTC 디바이스에 가장 낮은 우선 순위를 갖도록 할 수 있다.

[0078] 기지국은 URLLC, eMBB 및 mMTC 디바이스를 상기 기지국으로부터의 거리에 따라 내림차순으로 정렬한다. 다음으로 기지국은 게인이 가장 높은 채널에 URLLC 디바이스를 할당하고 도 3과 같이 사용 가능한 채널에 따라 eMBB 디바이스와 mMTC 디바이스를 할당할 수 있다.

[0079] 구체적으로 도 3에서는 4 개의 URLLC, 5개의 eMBB 및 3개의 mMTC 디바이스가 있는 3 종 NOMA 시스템의 우선순위 기반 채널 할당을 도시한다. 도 3에 도시되는 채널 할당 방식은 예시적이며, 기지국과의 각 장치의 CNR에 따라 달라질 수 있다.

[0080] 채널 할당 최적화의 또 다른 주요 요구 사항은 채널 및 전체 합속도를 최대화하는 것이다. 기지국은 각 채널 k에 대해 합속도를 최대화하는지 확인하기 위한 $\binom{N}{n}$의 조합을 가지고 있다. 따라서 전체 조합은 일반적으로 MCSR에 대해 $\sum_{i=1}^{K}\binom{N-(n\times i)}{n}$ 이다. 우선 순위에 관해서는 채널에서 낮은 우선 순위 디바이스가 높은 우선 순위 디바이스를 대체할 수 없다. 그러나 높거나 같은 우선 순위 디바이스는 주어진 채널에서 같거나 낮은 우선 순위 디바이스를 대체할 수 있다. 우선순위 체계와 통합된 최대화 프로세스는 기지국이 디바이스의 가능한 모든 조합을 확인해야 하기 때문에 계산적으로 복잡하다. 따라서 다음으로는 계산 복잡도를 줄이기 위해 우선 순위를

유지하고 합속도를 최대화하면서 디바이스에 채널을 할당하는 DQL을 설명한다.

[0081] 일실시예에 따른 DQL에서는 우선 순위 기반 채널 할당 문제를 최적화할 수 있다. 구체적으로, DQL 알고리즘은 일반적으로 심층신경망(Deep Neural Network, DNN) 에이전트와 환경으로 구성된다. 에이전트는 환경과 상호 작용하고 취할 행동(Action)을 결정한다. 예를 들어 기지국은 에이전트 역할을 하며 URLLC, eMBB 및 mMTC 디바이스 정보로 구성된 환경과 상호 작용한다.

[0082] 최초에, 에이전트는 모든 디바이스의 채널 정보를 수집하기 위해 환경 탐색을 시작한다. 각 시간 단계 t에서, 환경 내 에이전트의 현재 상태 s_t를 기반으로, 에이전트는 DNN을 사용하여 채널을 할당할 때의 행동 a_t를 예측한다. 리턴 값으로, 도 4에서 보이는 것과 같이 에이전트는 환경으로부터 즉각적인 보상 r_t와 다음 상태 s_{t+1}을 받는다. 에이전트는 채널 할당을 잘 수행하면 좋은 보상 r_t를 받는다. 에이전트는 행동을 예측함으로써 환경에 대해 학습하고 최적의 채널 할당 정책 π_c를 달성한다. 최적의 정책은 DNN에 의해 각 시간 단계 t에서 학습된다. 에이전트는 여러 에피소드에 대해 채널 할당 프로세스를 반복하여 정책 π_c를 업데이트하고 개선한다. 할당할 채널이 남아 있지 않으면 한 에피소드가 종료된다.

[0083] 일실시예에 따른 DNN에서 사용하기 위한 상태(state), 행동(action) 및 보상(reward)을 다음과 같이 정의할 수 있다.

[0084] 1) 상태(State): 각 디바이스에 대한 채널 정보를 환경의 상태로 정의한다. K개의 채널 선호도를 갖는 N개의 디바이스에 대하여, 상태 공간은 N×K 개의 요소(element)를 가지며 $S = \{ \Gamma^1_{1,} \Gamma^2_{1,} \Gamma^3_1, \cdots, \Gamma^K_{1,} \Gamma^K_{2,} \Gamma^K_3, \cdots, \Gamma^k_N\}$으로 나타낼 수 있다.

[0085] 2) 행동(Action): 에이전트의 주요 행동은 행동 공간 A에 속하는 디바이스에 채널을 할당하는 것이다. 집합 S의 각 에피소드에서, 에이전트는 집합 S로부터의 K 요소당 하나의 행동을 유지하면서 $N \in A$ 개의 행동을 취해야 한다. 예를 들어 2, 3, …, n 개의 디바이스를 갖는 NOMA에 대해, 에이전트는 하나의 행동을 각각 2, 3, …, n회 수행할 수 있다.

[0086] 3) 보상(Reward): 에이전트가 N개의 행동을 완료할 때마다, 에이전트는 각 행동에 대해 보상 r^l_t을 받는다. 각각의 올바른 행동에 대해 에이전트는 긍정적인 보상 r_i를 받고 에이전트가 올바른 n개의 행동을 취할 때 에이전트는 취한 행동에 대한 보상으로 해당 채널의 합속도를 얻는다. 예를 들어, 3개의 디바이스를 갖는 NOMA를 가정한다. 에이전트는 채널당 3개의 장치를 할당해야 한다. 이 경우 에이전트가 디바이스의 우선 순위에 따라 적절한 채널을 성공적으로 선택하면 에이전트는 긍정적인 보상 r_i (즉, 10)를 얻는다. 에이전트가 3개의 디바이스에 대해 동일한 적절한 채널을 선택할 수 있는 경우, 에이전트는 3개의 작업에 대한 보상으로 수학식 1에 의해 계산된 합속도를 얻는다. 이 때 보상 함수는 다음과 같이 정의할 수 있다.

수학식 13

[0087]
$$r_t^l = \begin{cases} \sum_{i=1}^{n} R_i^k & if\ a_p^k = n \\ 0 < r_i < \sum_{i=1}^{n} R_i^k\ for\ each\ a_i^l & if\ a_p^k < n \\ 0 & if\ a_p^k = 0 \end{cases}$$

[0088] 여기서 a^k_p는 채널 k당 취해진 적절한 행동 a^l_t의 수이고, $\forall l = 1, 2, \cdots, N \in A$ 이다. 전체 시스템의 성능과 공정성을 증가시키는 각 채널에 대한 합속도를 최대화하는 것을 설명한다.

[0089] 도 5에서는 상태, 행동 및 보상과 함께 채널 할당을 위한 정책 컨트롤러로서 DNN 구조를 도시한다. DNN은 Q-테이블(Q-Table)을 대체하고, 환경의 각 상태-행동 쌍에 대한 Q-값을 추정한다. 결국 DNN은 채널 할당을 위한 최적의 정책에 대한 근사치를 계산한다. 일실시예에 따른 DNN은 오토인코더(Autoencoder) 모델과 LSTM(Long Short-Term Memory) 모델의 두 부분으로 구성된다. DNN의 메인 목표는 Q(S, A)로 표현될 수 있는 각 상태 공간

에 대한 각 장치-채널 쌍에 대한 확률을 도출하는 것이다. 이러한 확률은 DQL에 대한 Q-값이다.

[0090] 일실시예에 따른 오토인코더는 입력의 수가 출력 뉴런의 수와 동일한 피드포워드(Feed-forward) 신경망이다. 구체적으로 설명하면, 입력을 저차원 코드로 압축한 다음, 출력 단에서 상기 코드로부터 입력 데이터를 재구성한다. 오토인코더는 화려한 처리나 라벨링 없이 원시 입력 데이터를 쉽게 처리할 수 있다. 따라서 오토인코더는 비지도 학습(Unsupervised learning) 기법의 일부로 간주되며 훈련 데이터로부터 라벨을 생성할 수 있다.

[0091] 도 6은 오토인코더를 도시한다. 예시적으로 그러나 한정되지 않게 오토인코더는 인코더(Encoder), 코드(Code) 및 디코더(Decoder)라는 세 가지 주요 부분으로 구성될 수 있다. 인코더와 디코더는 모두 완전히 연결된 신경망이다. 인코더는 2^n개의 뉴런을 갖는 입력 레이어(Input layer)로 시작하고 2^{n-h}개의 뉴런을 갖는 다중 히든 레이어(Hidden layer)가 이어진다. 여기서 h는 레이어의 위치이다. 히든 레이어당 뉴런의 수는 오토인코더의 코드 부분까지 계속 감소한다. 예시적으로 코드 레이어에 대해 2^3개의 뉴런을 사용한다. 디코더는 출력 레이어로 끝나는 인코더의 거울상 대칭 구조를 갖는다. 상술한 구조는 샌드위치처럼 레이어가 차례로 쌓이기 때문에 스택형 오토인코더(Stacked Autoencoder)이다. 일실시예에 따르면 오토인코더의 각 레이어에 대한 활성화 함수(Activation Function)로 ReLU(Rectified Linear Unit)를 사용할 수 있다.

[0092] 장단기 기억(LSTM)은 순환 신경망(Recurrent Neural Network, RNN)의 진화된 형태이다. LSTM은 장기 종속성을 학습하고 향후 사용을 위해 이전 정보를 기억할 수 있는 특수한 유형의 RNN이다.

[0093] LSTM 네트워크는 여러 개의 LSTM 셀로 구성된 체인 구조를 가지고 있다. 3개의 LSTM 셀을 사용하여 구축한 LSTM 네트워크를 가정한다.

[0094] 도 7에서는 단일 LSTM 셀의 구조가 도시된다. 일실시예에 따른 LSTM 셀에는 3개의 입력 파라미터(parameter)와 2개의 출력 파라미터가 존재한다. 상기 셀과 히든 상태(Hidden State)는 입력과 출력 사이의 공통 파라미터이다. 다른 파라미터는 현재 입력이다. 상기 LSTM 셀은 또한 도 7에서 도시되는 것과 같이 선형 변환을 포함하는 3개의 시그모이드(Sigmoid) 레이어와 2개의 하이퍼탄젠트(tanh) 레이어를 포함할 수 있다. 초기에, 첫 번째 LSTM 셀에 대한 입력과 함께, 무작위 셀(random cell)과 히든 상태(hidden state)가 주어진다. 그러면 두 개의 출력(히든 상태, 셀 상태)이 다음 셀의 세 개의 입력이 된다.

[0095] 입력 및 출력 크기가 128이고 코드 크기가 8인 오토인코더에 이어서 입력 크기가 128, 히든 상태 크기가 64, 순환 레이어가 3인 LSTM 네트워크를 예시적으로 설명한다. 마지막으로 LSTM의 출력은 선형 레이어와 시그모이드 레이어를 통과하여 각 장치의 선호 채널에 대한 확률을 얻는다. 상태 공간(S)은 정책 네트워크의 입력으로 제공된다. 처음에 입력은 먼저 128의 차원에 포함된다. 그 후 도 5와 같이 채널 할당 확률을 생성하기 위해 정책 네트워크를 통과한다.

[0096] 도 8에서는 일실시예에 따른 DQL 프레임워크를 도시한다. 일실시예에 따른 DQL 프레임워크는 에이전트와 환경을 포함할 수 있다. 상기 에이전트에는 타겟 DNN, 정책 DNN, 경험 리플레이 메모리 및 배치를 포함할 수 있다. 이하에서는 각 구성을 상세히 설명한다.

[0097] 일실시예에 따른 DNN은 에피소드 마다 훈련 데이터 집합 $T_{data} = \{S^1, S^2, \cdots, S^{ins}\}$를 이용하여 점차적으로 훈련된다. 각 상태 공간(S)에 대해 디바이스-채널 쌍은 DNN으로부터의 출력 확률에 따라 입실론 탐욕(ε-greedy) 정책을 사용하여 선택된다. 하나의 에피소드는 모든 상태 공간이 DNN을 통과할 때 종료된다. 상태 공간별로 각 디바이스에 대해 행동을 취하는 정책은 다음과 같이 표현될 수 있다.

수학식 14

[0098]
$$a_i^l = \begin{cases} argmax\ Q(S^i, A_i^l) & if\ \varepsilon < \varepsilon_{th}\ ; where\ \varepsilon_{th} \in (0,1] \\ random\ action\ [1,K] & otherwise \end{cases}$$

[0099] $\forall l = 1, 2, \cdots, N\ \ A,$

[0100] $\forall i = 1, 2, \cdots, ins.$

[0101] 상기 수학식 14를 사용하여 행동을 수행한 후 에이전트는 수학식 13과 다음 상태 공간(S^{i+1}) 따라 보상을

받는다.

[0102] DNN을 훈련하기 위해, 손실을 계산하고 역전파를 수행하는 DNN의 파라미터를 최적화한다. 손실을 계산하기 위해 타겟 DNN이라고 부르는 다른 DNN으로부터 S^{i+1}의 각 디바이스-채널 쌍에 대한 최적의 Q*값(Q*-value)들의 근사한다. 타겟 DNN은 정책 DNN과 동일하며 정책 DNN의 파라미터에 의해 초기화된다. 다음 상태 공간(S^{i+1})은 타겟 DNN에 대한 입력으로 주어지고 출력으로부터 최적의 Q*-값이 에이전트에 의해 탐욕스럽게(greedily) 선택된다. 채널 할당은 분류 문제이기 때문에, 범주형 교차 엔트로피 손실 함수(categorical cross-entropy loss function)를 사용하여 최적의 Q* 값과 일반 Q 값 사이의 손실을 계산한다. 상기 손실을 계산한 후 아담 최적화기(Adam optimizer)를 사용하여 정책 DNN을 최적화한다. 최적의 Q*-값을 올바르게 추정하기 위해, 주기적으로 특정 에피소드 후에 정책 DNN의 파라미터로 타겟 DNN을 업데이트한다.

[0103] 최적 정책의 보다 안정적인 수렴을 위해, DQL에 경험 리플레이 메모리(Experience Replay Memory, ERM)를 설명한다. 초기에 에이전트는 환경을 탐색하고 현재 상태, 행동, 보상 및 다음 상태(S^{i}, A_i, r_i, S^{i+1})를 상기 경험 리플레이 메모리의 튜플(Tuple)로 저장한다. 다음으로 상기 에이전트는 상기 경험 리플레이 메모리에서 튜플의 미니 배치를 가져와 정책 DNN을 훈련시킨다. 상기 경험 리플레이 메모리는 각 훈련 데이터에 대해 계속 업데이트된다.

[0104] 도 9는 DQL 프레임워크의 알고리즘을 도시한다. 상기 DQL 프레임워크의 알고리즘을 보다 상세히 설명한다.

[0105] 일실시예에 따른 DQL 프레임워크는 입실론(ε)에 기초해서 채널을 선택하고 현재 상태공간 S^{i}를 위해 액션 공간 A^{i}에 추가하는 단계; 보상 r_i와 다음 상태공간 S^{i+1}을 관찰하는 단계; (S^{i}, A^{i}, r_t, S^{i+1})을 경험 리플레이 메모리(ERM)에 넣는 단계; 및 상기 경험 리플레이 메모리로부터 추출된 무작위 샘플로 미니 배치를 생성하는 단계를 포함할 수 있다.

[0106] 상기 미니 배치를 생성하는 단계는, 미니배치의 튜플마다 정책 DNN을 사용하여 Q값을 얻는 단계; 타겟 DNN을 사용하여 Q*값을 근사하는 단계; Q와 Q*값을 사용하여 손실을 계산하는 단계; 아담 최적화기(adam optimizer)를 사용하여 정책 DNN의 파마미터p를 최적화하는 단계를 포함할 수 있다.

[0108] 도 10은 일 실시에에 따른 심층 Q-학습을 이용한 우선순위 기반 자원 할당 방법을 도시한다. 일 실시예에서, 비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 방법에 있어서, 도 10의 각 단계는 기지국에 의해 수행될 수 있다.

[0109] 도 10을 참고하면, S1001 단계는, 적어도 하나의 디바이스에 전력을 할당하는 단계이다.

[0110] 일 실시예에서, 합속도(sum-rate)에 기초하여 적어도 하나 이상의 디바이스에 전력을 할당할 수 있다.

[0111] 일 실시예에서, 상기 합속도는 채널에 대한 각 장치의 데이터전송률을 합산하여 계산되는 속도일 수 있다.

[0112] 일 실시예에서, 상기 전력을 할당하는 단계는, 미리 지정되는 임계값 이상으로 전력을 할당할 수 있다.

[0113] S1003 단계는, 상기 적어도 하나의 디바이스의 우선순위를 결정하는 단계이다.

[0114] 일 실시예에서, 적어도 하나의 디바이스에 요구되는 통신 품질 요구사항을 기준으로 우선순위를 결정할 수 있다.

[0115] 일 실시예에서, 상기 적어도 하나의 디바이스와 기지국과의 거리를 기준으로 우선순위를 결정할 수 있다.

[0116] 일 실시예에서, 상기 적어도 하나의 디바이스는, 향상된 모바일 광대역(eMBB) 장치, 대규모 사물통신(mMTC) 장치 및 초고신뢰·저지연 통신(URLLC) 장치 중 적어도 하나 이상을 포함할 수 있다.

[0117] S1005 단계는, Q-학습(Q-Learning)을 이용하여 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 상기 적어도 하나의 디바이스에 채널을 할당하는 단계이다.

[0118] 일 실시예에서, 상기 Q-학습의 상태, 행동 및 보상에 대해, 상기 디바이스의 채널 대 잡음비를 상태로, 상기 채널 할당을 행동으로, 그리고 채널에 대한 합속도를 보상으로 각각 설정하고, 현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하고, 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하며, 상기

현재 상태를 기반으로 DNN(Deep Neural Network)을 사용하여 채널을 할당하는 단계와 상기 채널에 대한 합속도 및 다음 상태 정보를 획득하는 단계를 반복 수행하면서 채널 할당 정책을 결정할 수 있다.

[0119] 도 11은 일 실시예에 따른 심층 Q-학습을 이용한 우선순위 기반 자원 할당 장치를 도시한다. 일 실시예에서, 비직교 다중 접속 시스템(Non-Orthogonal Multiple Access)에서 자원을 할당하는 장치에 있어서, 도 11의 자원 할당 장치(1100)는 기지국을 포함할 수 있다.

[0120] 도 11을 참고하면, 자원 할당 장치(1100)는 제어부(1110), 통신부(1120) 및 저장부(1130)를 포함할 수 있다.

[0121] 제어부(1110)는 적어도 하나의 디바이스의 우선순위를 결정하고, 전력 및 채널을 할당할 수 있다.

[0122] 일 실시예에서, 제어부(1110)는 Q-학습(Q-Learning)을 이용하여 상기 채널 할당에 따른 합속도(Data rate)를 학습하고, 학습한 내용에 기초하여 합속도가 미리 지정되는 값 이상이 되도록하는 채널 할당 정책을 결정할 수 있다.

[0123] 일 실시예에서, 제어부(1110)는 적어도 하나의 프로세서 또는 마이크로(micro) 프로세서를 포함하거나, 또는, 프로세서의 일부일 수 있다. 또한, 제어부(1110)는 CP(communication processor)라 지칭될 수 있다. 제어부(1110)는 본 발명의 다양한 실시예에 따른 자원 할당 장치(1100)의 동작을 제어할 수 있다.

[0124] 통신부(1120)는 적어도 하나의 디바이스에게 할당된 채널에 대한 정보를 송신할 수 있다.

[0125] 일 실시예에서, 통신부(1120)는 유선 통신 모듈 및 무선 통신 모듈 중 적어도 하나를 포함할 수 있다. 통신부(1120)의 전부 또는 일부는 '송신부', '수신부' 또는 '송수신부(transceiver)'로 지칭될 수 있다.

[0126] 저장부(1130)는 채널 할당 정책을 저장할 수 있다.

[0127] 일 실시예에서, 저장부(1130)는 휘발성 메모리, 비휘발성 메모리 또는 휘발성 메모리와 비휘발성 메모리의 조합으로 구성될 수 있다. 그리고, 저장부(1130)는 제어부(1110)의 요청에 따라 저장된 데이터를 제공할 수 있다.

[0128] 도 11을 참고하면, 자원 할당 장치(1100)는 제어부(1110), 통신부(1120) 및 저장부(1130)를 포함할 수 있다. 본 발명의 다양한 실시 예들에서 자원 할당 장치(1100)는 도 11에 설명된 구성들이 필수적인 것은 아니어서, 도 11에 설명된 구성들보다 많은 구성들을 가지거나, 또는 그보다 적은 구성들을 가지는 것으로 구현될 수 있다.

[0129] 이상의 설명은 본 발명의 기술적 사상을 예시적으로 설명한 것에 불과한 것으로, 통상의 기술자라면 본 발명의 본질적인 특성이 벗어나지 않는 범위에서 다양한 변경 및 수정이 가능할 것이다.

[0130] 따라서, 본 명세서에 개시된 실시예들은 본 발명의 기술적 사상을 한정하기 위한 것이 아니라, 설명하기 위한 것이고, 이러한 실시예들에 의하여 본 발명의 범위가 한정되는 것은 아니다.

[0131] 본 발명의 보호범위는 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본발명의 권리범위에 포함되는 것으로 이해되어야 한다.

부호의 설명

[0132] 1100: 자원 할당 장치

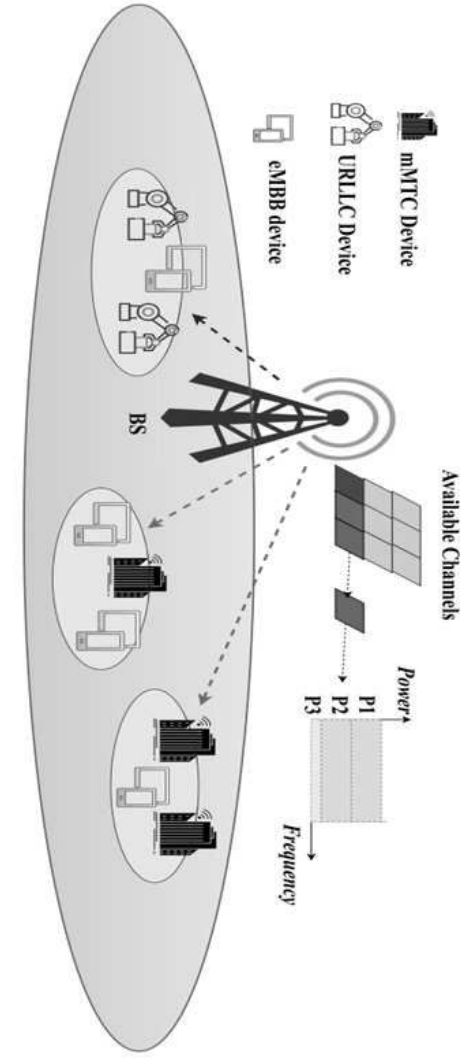
1110: 제어부

1120: 통신부

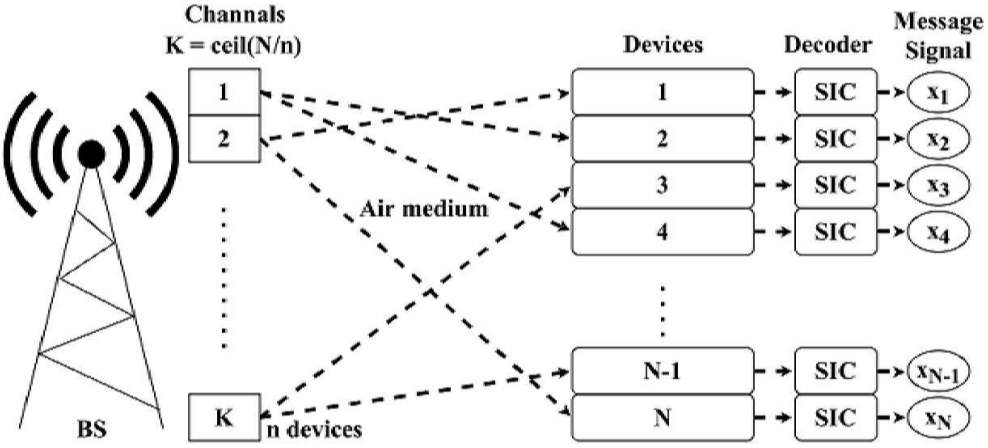
1130: 저장부

도면

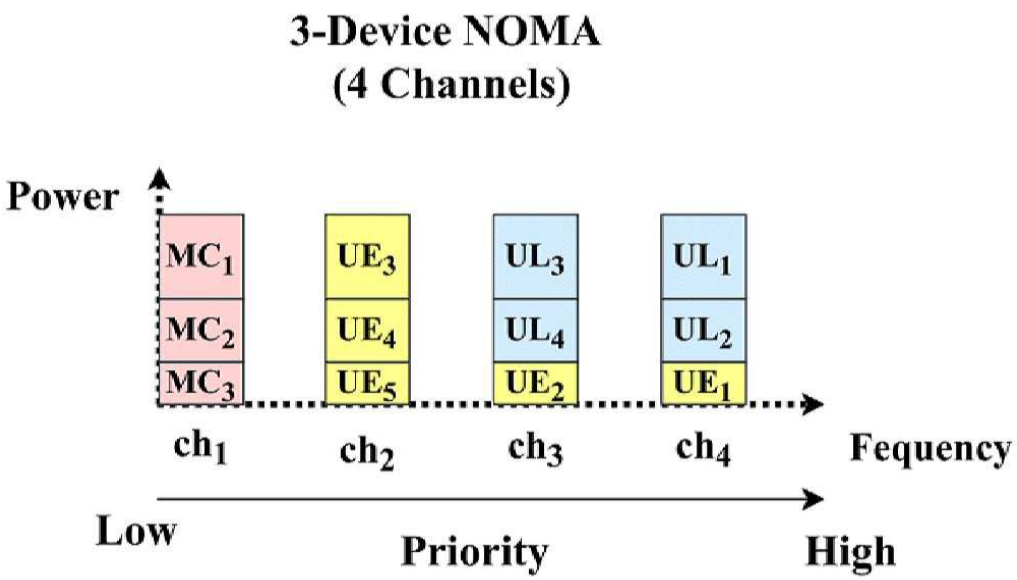
도면1


eMBB device
URLLC Device
mMTC Device
BS
Available Channels
Power
P1
P2
P3
Frequency

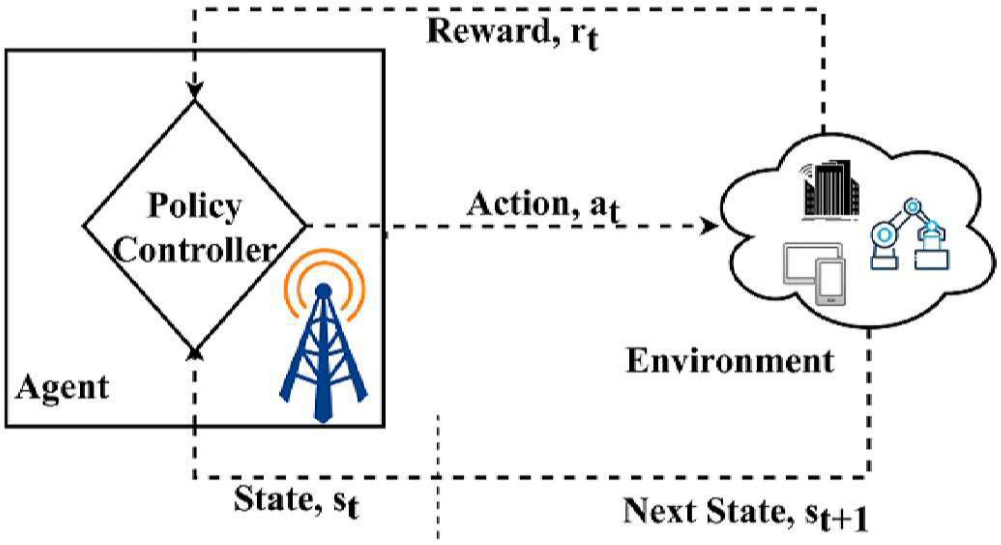
도면2


Channals
K = ceil(N/n)
Devices
Decoder
Message Signal
1
2
K
n devices
Air medium
BS
1
2
3
4
N-1
N
SIC
x_1
x_2
x_3
x_4
x_{N-1}
x_N

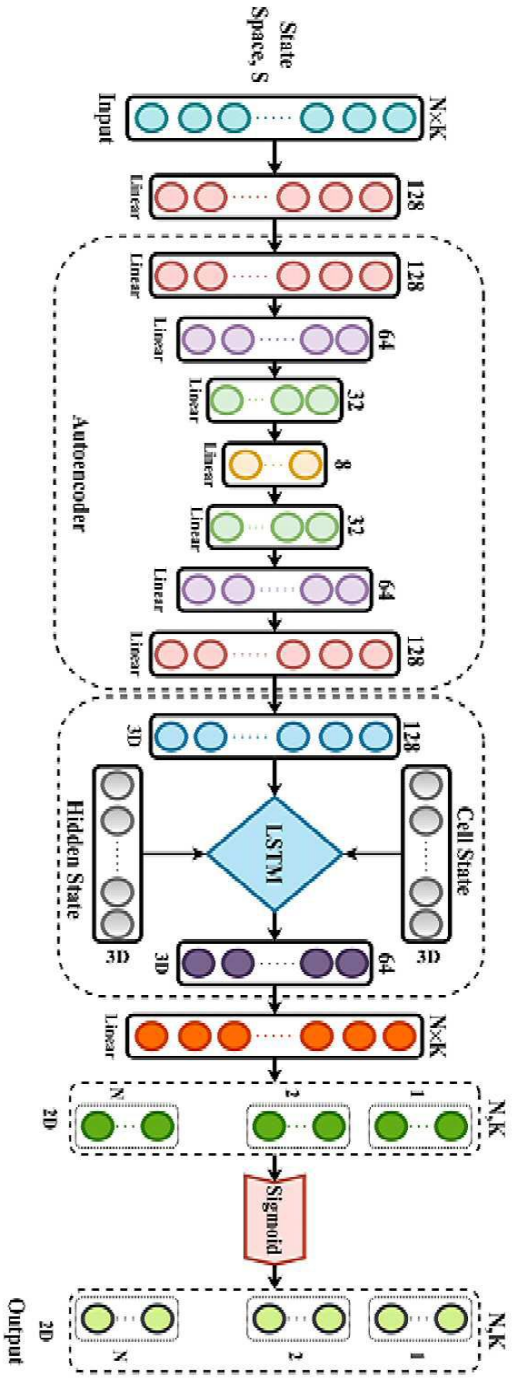
도면3


3-Device NOMA
(4 Channels)
Power
MC1
MC2
MC3
UE3
UE4
UE5
UL3
UL4
UE2
UL1
UL2
UE1
ch1
ch2
ch3
ch4
Fequency
Low
Priority
High

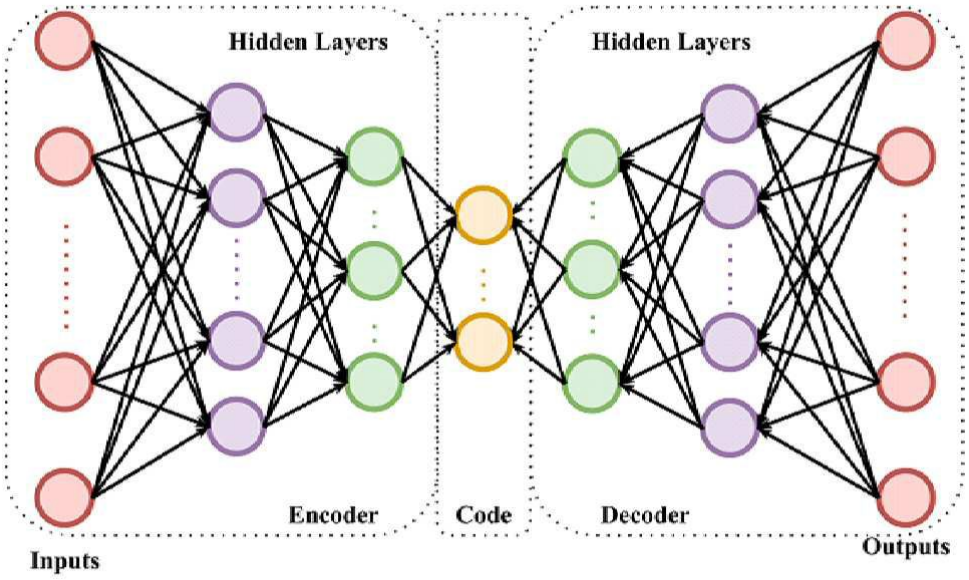
도면4


Reward, rt
Policy
Controller
Action, at
Agent
Environment
State, st
Next State, st+1

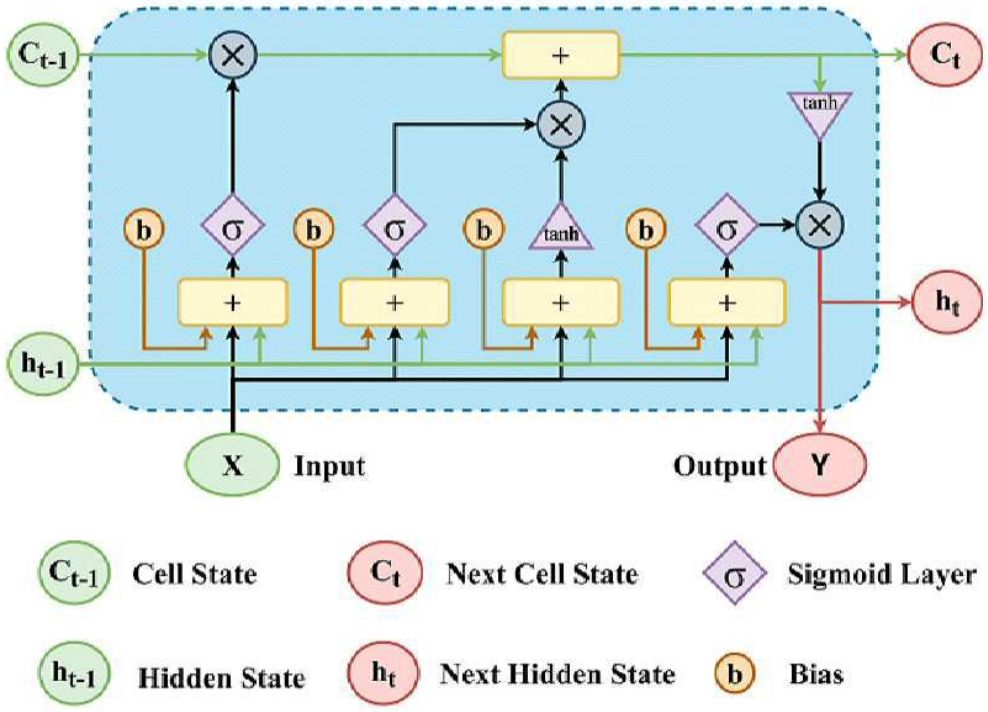
도면5


State Space, S
Input
N×K
Linear
128
Autoencoder
64
32
8
3D
LSTM
Hidden State
Cell State
Sigmoid
Output
N,K
2D

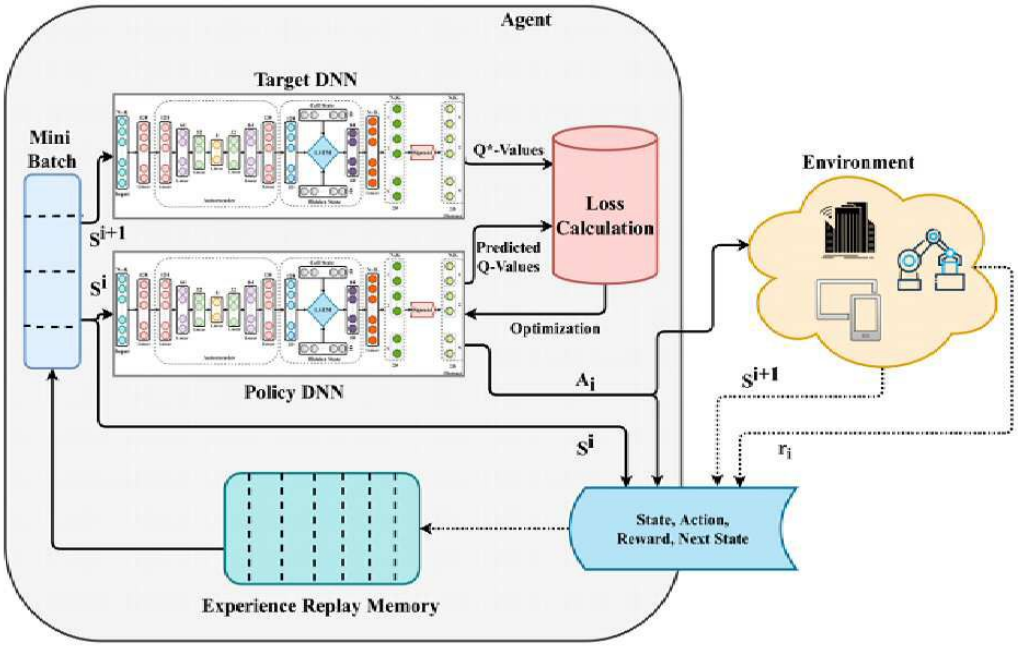
도면6


Hidden Layers
Hidden Layers
Encoder
Code
Decoder
Inputs
Outputs

도면7


C_{t-1}
C_t
tanh
b
σ
tanh
h_t
h_{t-1}
X
Input
Output
Y
C_{t-1} Cell State
C_t Next Cell State
σ Sigmoid Layer
h_{t-1} Hidden State
h_t Next Hidden State
b Bias

도면8

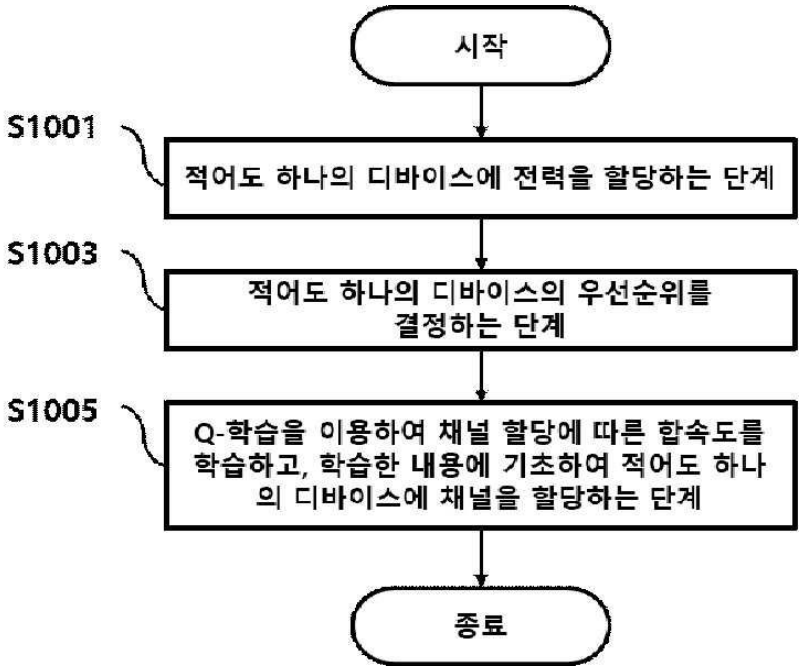

Agent
Target DNN
Mini Batch
Q*-Values
Loss Calculation
Environment
Predicted Q-Values
Optimization
Policy DNN
Ai
Si+1
Si
ri
State, Action, Reward, Next State
Experience Replay Memory

도면9

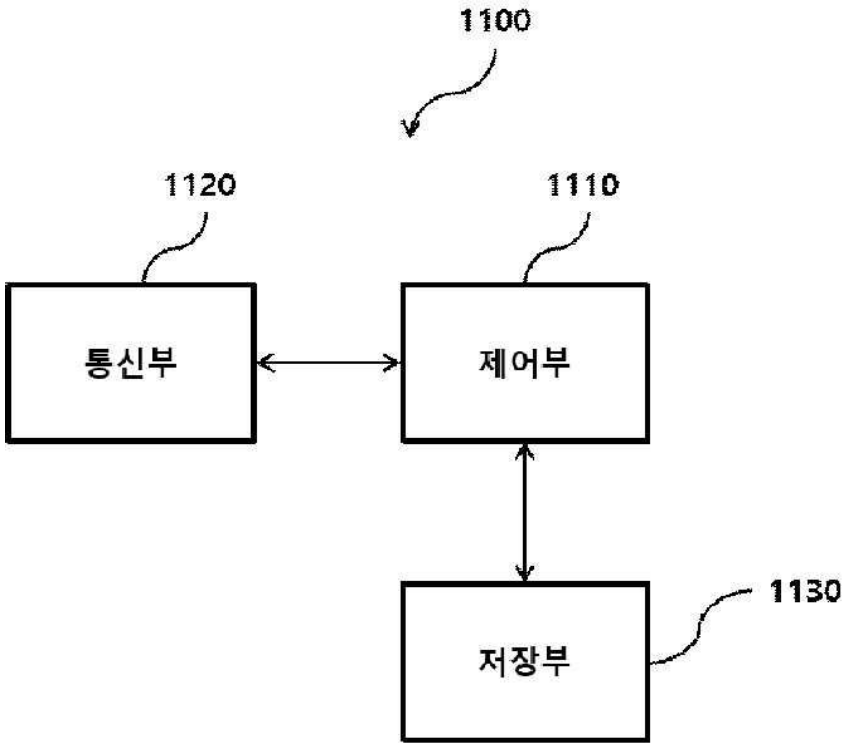
Algorithm 1 Proposed Deep Q-Learning Algorithm

Initialize policy and target DQL network with random parameters (p and p').
Initialize experience replay memory (ERM).
Initialize ϵ.
for each episode **do**
    for each instance **do**
        for each device **do**
            Select an channel and add to action space A_i for present state space S^i based on ϵ.
        end for
        Observe the immediate rewards r_i and next state space S^{i+1}.
        Insert (S^i, A_i, r_t, S^{i+1}) in ERM.
        Create a mini-batch with random sample of (S^i, A_i, r_t, S^{i+1}) from ERM.
        for each tuple in mini-batch **do**
            Obtain Q-values using policy DNN.
            Approximate Q^*-values using target DNN.
            Calculate the loss using Q an Q^*-values.
            Optimize the parameters p of the policy DNN using Adam optimizer.
        end for
    end for
    $p' \leftarrow p$ after certain number of episodes.
end for

도면10


시작
S1001
적어도 하나의 디바이스에 전력을 할당하는 단계
S1003
적어도 하나의 디바이스의 우선순위를 결정하는 단계
S1005
Q-학습을 이용하여 채널 할당에 따른 합속도를 학습하고, 학습한 내용에 기초하여 적어도 하나의 디바이스에 채널을 할당하는 단계
종료

도면11


1100
1120
1110
통신부
제어부
저장부
1130