(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0099987
(43) 공개일자 2023년07월05일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) *G06N 3/04* (2023.01)
H04B 7/0413 (2017.01)
(52) CPC특허분류
G06N 3/08 (2023.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2021-0189477
(22) 출원일자 2021년12월28일
심사청구일자 2021년12월28일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이인규
서울특별시 강남구 영동대로 640, 웨스트동 3303호(삼성동, 아이파크삼성)
장정현
서울특별시 동대문구 제기로6길 9-3, 402호(제기동, 태양빌리지)
(74) 대리인
최윤서, 손제관

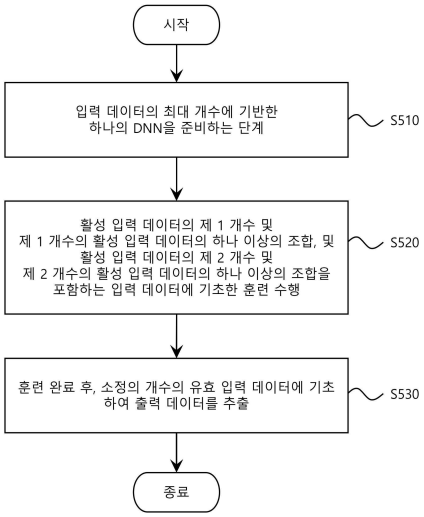
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **확장가능한 심층 신경망 훈련 방법 및 장치**

(57) 요 약

본 개시는 심층 신경망의 확장가능한 훈련 방법 및 장치에 대한 것이다. 본 개시의 일 실시예에 따른 확장가능한 심층신경망(DNN) 훈련 방법은, 입력 데이터의 최대 개수에 기반한 하나의 DNN을 준비하는 단계; 상기 DNN에 대해서, 상기 최대 개수 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 상기 최대 개수 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행하는 단계; 및 상기 DNN에 대한 훈련이 완료된 후, 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출하는 단계를 포함할 수 있다.

대 표 도 - 도5



(52) CPC특허분류
H04B 7/0413 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711134433
과제번호	2021-0-00467
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	6G 핵심 기술개발
연구과제명	지능형 6G 무선 액세스 시스템
기 여 율	1/1
과제수행기관명	고려대학교 산학협력단
연구기간	2021.04.01 ~ 2025.12.31

명 세 서

청구범위

청구항 1

확장가능한 심층신경망(DNN) 훈련 방법에 있어서, 상기 방법은:

입력 데이터의 최대 개수에 기반한 하나의 DNN을 준비하는 단계;

상기 DNN에 대해서, 상기 최대 개수 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 상기 최대 개수 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행하는 단계; 및

상기 DNN에 대한 훈련이 완료된 후, 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출하는 단계를 포함하는, 방법.

청구항 2

제 1 항에 있어서,

상기 제 1 개수의 활성 입력 데이터 및 상기 제 2 개수의 활성 입력 데이터에 기초하여 동일한 하나의 상기 DNN에 대한 훈련이 수행되는, 방법.

청구항 3

제 1 항에 있어서,

상기 제 1 개수 및 상기 제 2 개수는 랜덤하게 결정되는, 방법.

청구항 4

제 1 항에 있어서,

상기 제 1 개수의 활성 입력 데이터 및 상기 제 2 개수의 활성 입력 데이터 각각의 상기 하나 이상의 조합은 랜덤하게 결정되는, 방법.

청구항 5

제 1 항에 있어서,

상기 최대 개수 중 상기 제 1 개수의 활성 입력 데이터를 제외한 나머지 입력 데이터, 및 상기 최대 개수 중 상기 제 2 개수의 활성 입력 데이터를 제외한 나머지 입력 데이터는 0 행렬로 주어지는, 방법.

청구항 6

제 1 항에 있어서,

상기 최대 개수의 입력 데이터의 각각이 0 행렬을 가지는 확률에 기초하여, 상기 최대 개수의 입력 데이터의 각각이 결정되고,

상기 제 1 개수 및 상기 제 2 개수의 각각은, 각각의 훈련 수행에서 0 행렬이 아닌 입력 데이터의 개수에 해당하는, 방법.

청구항 7

제 1 항에 있어서,

최대 개수 이하의 활성 입력 데이터의 추가적인 하나 이상의 개수 및 상기 추가적인 하나 이상의 개수 각각에 대한 활성 입력 데이터의 하나 이상의 조합에 대한 추가적인 하나 이상의 오프라인 훈련이 수행되는, 방법.

청구항 8

제 7 항에 있어서,

상기 추가적인 하나 이상의 개수는, 상기 제 1 개수 및 상기 제 2 개수를 제외한 상기 최대 개수 이하의 나머지 모든 개수에 해당하는, 방법.

청구항 9

제 7 항에 있어서,

상기 소정의 개수는 상기 제 1 개수, 상기 제 2 개수 또는 상기 추가적인 하나 이상의 개수 중의 어느 하나에 해당하는, 방법.

청구항 10

제 1 항에 있어서,

상기 DNN은 소정의 시스템의 하나 이상의 파라미터의 최적화를 위해서 준비되는, 방법.

청구항 11

제 10 항에 있어서,

상기 입력 데이터의 개수는, 상기 소정의 시스템을 구성하는 개체 또는 자원 중의 하나 이상의 개수에 대응하는, 방법.

청구항 12

제 10 항에 있어서,

상기 소정의 시스템은 다중입출력(MIMO) 무선 통신 시스템인, 방법.

청구항 13

확장가능한 심층신경망 장치에 있어서, 상기 장치는:

메모리; 및

프로세서를 포함하고,

상기 프로세서는,

입력 데이터의 최대 개수에 기반한 하나의 DNN을 준비하고;

상기 DNN에 대해서, 상기 최대 개수 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 상기 최대 개수 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행하고; 및

상기 DNN에 대한 훈련이 완료된 후, 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출하도록 설정되는, 장치.

발명의 설명

기 술 분 야

[0001] 본 개시는 심층 신경망 훈련에 대한 것이며, 구체적으로는 무선 통신 시스템에서 심층 신경망을 확장가능한 설계에 기반하여 훈련하는 방법 및 장치에 대한 것이다.

배 경 기 술

[0002] MIMO(Multi-Input Multi-Output) 무선 통신 시스템 또는 다중 안테나 시스템의 성능 최적화 기법에 있어서, 시스템의 높은 복잡도 및 비선형 입출력 관계에 기반한 문제를 해결하기 위해, 심층 신경망(DNN)을 이용한 입출력

사이의 비선형 관계를 유추하는 딥러닝 방식이 연구되고 있다. 예를 들어, 최적의 기존 솔루션의 지도 학습(supervised learning)을 통한 근사치 매핑 함수를 훈련하는 방법, 훈련 데이터 분포로부터 비지도 학습(unsupervised learning)을 통한 시스템 성능을 나타내는 손실 함수 최소화 방법 등이 연구되고 있다.

[0003] 이러한 기존의 DNN을 이용한 시스템 성능 최적화 기법은, 오프라인 훈련 단계에서 고정된 규모의 신경망에 기초한 훈련을 진행한다. 고정된 규모의 신경망은, 최적화 대상이 되는 개체(entity), 자원(resource) 등에 대한 고정된 개수만을 설정한 신경망을 설계하는 것을 포함한다. 따라서, X 개의 입력을 가지는 시스템 설정에 따른 특정 규모의 신경망에 대한 오프라인 훈련이 완료되더라도, X와 상이한 Y 개의 입력을 가지는 시스템 설정에 대한 최적화 솔루션을 찾기 위해서는 상기 특정 규모와 상이한 규모의 신경망을 설계 및 훈련하는 과정이 요구된다. 즉, 기존의 심층신경망 오프라인 훈련 방법에서는, 입력과 출력을 나타내는 행렬의 차원이 고정되므로, 변화하는 시스템 설정에 대응하기 어려웠다. 즉, 상이한 시스템 설정의 각각에 대응하는 신경망의 설계가 필요하고, 각각의 신경망에 대한 훈련이 필요한 문제가 있다.

발명의 내용

해결하려는 과제

[0004] 본 개시의 기술적 과제는 가변적인 시스템 설정에 대응하는 심층 신경망 훈련 방법 및 장치를 제공하는 것이다.

[0005] 본 개시의 추가적인 기술적 과제는 복수의 시스템 설정에 대응하는 하나의 범용적인(versatile) 심층 신경망 모델을 설계하고 이를 훈련하는 방법 및 장치를 제공하는 것이다.

[0006] 본 개시에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0007] 본 개시의 일 양상에 따른 확장가능한 심층신경망(DNN) 훈련 방법은, 입력 데이터의 최대 개수에 기반한 하나의 DNN을 준비하는 단계; 상기 DNN에 대해서, 상기 최대 개수 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 상기 최대 개수 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행하는 단계; 및 상기 DNN에 대한 훈련이 완료된 후, 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출하는 단계를 포함할 수 있다.

[0008] 본 개시의 추가적인 양상에 따른 확장가능한 심층신경망 장치는, 메모리; 및 프로세서를 포함하고, 상기 프로세서는, 입력 데이터의 최대 개수에 기반한 하나의 DNN을 준비하고; 상기 DNN에 대해서, 상기 최대 개수 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 상기 최대 개수 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행하고; 및 상기 DNN에 대한 훈련이 완료된 후, 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출하도록 설정될 수 있다.

[0009] 본 개시에 대하여 위에서 간략하게 요약된 특징들은 후술하는 본 개시의 상세한 설명의 예시적인 양상일 뿐이며, 본 개시의 범위를 제한하는 것은 아니다.

발명의 효과

[0010] 본 개시에 따르면, 추가적인 시스템 설정 마다 심층 신경망에 대한 추가적인 오프라인 훈련 없이 대응할 수 있는 심층 신경망 훈련 방법 및 장치가 제공될 수 있다.

[0011] 본 개시에 따르면, 복수의 시스템 설정에 대응하는 하나의 범용적인(versatile) 심층 신경망 모델을 설계하고 이를 훈련하는 방법 및 장치가 제공될 수 있다.

[0012] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0013] 도 1은 본 개시에 따른 심층 신경망의 입출력 데이터를 나타내는 도면이다.

도 2는 본 개시가 적용될 수 있는 무선 통신 시스템의 최적화를 위한 심층 신경망의 예시를 나타낸다.

도 3은 본 개시에 따른 확장가능한 심층 신경망 훈련 및 적용의 예시를 설명하기 위한 도면이다.

도 4는 본 개시의 예시들에 따른 시뮬레이션 결과를 나타내는 도면이다.

도 5는 본 개시에 따른 심층신경망 훈련 방법을 설명하기 위한 흐름도이다.

도 6은 본 개시에 따른 심층신경망 장치의 구성을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하에서는 첨부한 도면을 참고로 하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나, 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다.

[0015] 본 개시의 실시예를 설명함에 있어서 공지 구성 또는 기능에 대한 구체적인 설명이 본 개시의 요지를 흐릴 수 있다고 판단되는 경우에는 그에 대한 상세한 설명은 생략한다. 그리고, 도면에서 본 개시에 대한 설명과 관계없는 부분은 생략하였으며, 유사한 부분에 대해서는 유사한 도면 부호를 붙인다.

[0016] 본 개시에 있어서, 어떤 구성요소가 다른 구성요소와 "연결", "결합" 또는 "접속"되어 있다고 할 때, 이는 직접적인 연결관계 뿐만 아니라, 그 중간에 또 다른 구성요소가 존재하는 간접적인 연결관계도 포함할 수 있다. 또한 어떤 구성요소가 다른 구성요소를 "포함한다" 또는 "가진다"고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 배제하는 것이 아니라 또 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0017] 본 개시에 있어서, 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용되며, 특별히 언급되지 않는 한 구성요소들 간의 순서 또는 중요도 등을 한정하지 않는다. 따라서, 본 개시의 범위 내에서 일 실시예에서의 제1 구성요소는 다른 실시예에서 제2 구성요소라고 칭할 수도 있고, 마찬가지로 일 실시예에서의 제2 구성요소를 다른 실시예에서 제1 구성요소라고 칭할 수도 있다.

[0018] 본 개시에 있어서, 서로 구별되는 구성요소들은 각각의 특징을 명확하게 설명하기 위한 것이며, 구성요소들이 반드시 분리되는 것을 의미하지는 않는다. 즉, 복수의 구성요소가 통합되어 하나의 하드웨어 또는 소프트웨어 단위로 이루어질 수도 있고, 하나의 구성요소가 분산되어 복수의 하드웨어 또는 소프트웨어 단위로 이루어질 수도 있다. 따라서, 별도로 언급하지 않더라도 이와 같이 통합된 또는 분산된 실시예도 본 개시의 범위에 포함된다.

[0019] 본 개시에 있어서, 다양한 실시예에서 설명하는 구성요소들이 반드시 필수적인 구성요소들은 의미하는 것은 아니며, 일부는 선택적인 구성요소일 수 있다. 따라서, 일 실시예에서 설명하는 구성요소들의 부분집합으로 구성되는 실시예도 본 개시의 범위에 포함된다. 또한, 다양한 실시예에서 설명하는 구성요소들에 추가적으로 다른 구성요소를 포함하는 실시예도 본 개시의 범위에 포함된다.

[0020] 본 개시는 무선 통신 시스템에서 네트워크 노드들 간의 통신에 대한 것이다. 네트워크 노드는, 기지국, 단말 또는 릴레이(relay) 중의 하나 이상을 포함할 수 있다. 기지국(Base Station, BS)이라는 용어는, 고정국(fixed station), Node B, eNodeB(eNB), ng-eNB, gNodeB(gNB), 액세스 포인트(Access Point, AP) 등의 용어로 대체될 수 있다. 단말(terminal)은 UE(User Equipment), MS(Mobile Station), MSS(Mobile Subscriber Station), SS(Subscriber Station), 비-AP 스테이션(non-AP STA) 등의 용어로 대체될 수 있다.

[0021] 무선 통신 시스템은 기지국과 단말 간의 통신을 지원할 수도 있고, 단말간 통신을 지원할 수도 있다. 기지국과 단말 간의 통신에 있어서, 하향링크(Downlink, DL)는 기지국으로부터 단말로의 통신을 의미한다. 상향링크(Uplink, UL)은 단말로부터 기지국으로의 통신을 의미한다. 단말간 통신은 D2D(Device-to-Device), V2X(Vehicle-to-everything), ProSe(Proximity Service), 사이드링크(sidelink) 통신 등의 다양한 통신 방식 또는 서비스를 포함할 수 있다. 단말간 통신에 있어서 단말은 센서 노드, 차량, 재난 경보기 등의 형태로 구현될 수도 있다.

[0022] 또한, 무선 통신 시스템은 릴레이(relay) 또는 릴레이 노드(RN)를 포함할 수 있다. 기지국과 단말 간의 통신에

릴레이가 적용되는 경우, 릴레이는 단말에 대해서 기지국으로서 기능할 수 있고, 릴레이는 기지국에 대해서 단말로서 기능할 수 있다. 한편, 단말간 통신에 릴레이가 적용되는 경우, 릴레이는 각각의 단말에 대해서 기지국으로서 기능할 수 있다.

[0023] 본 개시는 무선 통신 시스템의 다양한 다중 액세스 방식에 적용될 수 있다. 예를 들어, 다중 액세스 방식은 CDMA(Code Division Multiple Access), TDMA(Time Division Multiple Access), FDMA(Frequency Division Multiple Access), OFDMA(Orthogonal Frequency Division Multiple Access), SC-FDMA(Single Carrier-FDMA), OFDM-FDMA, OFDM-TDMA, OFDM-CDMA, NOMA(Non-Orthogonal Multiple Access) 등을 포함할 수 있다. 또한, 본 개시가 적용될 수 있는 무선 통신 시스템은, 상향링크 및 하향링크 통신이 서로 구별되는 시간 자원을 이용하는 TDD(Time Division Duplex) 방식을 지원할 수도 있고, 서로 구별되는 주파수 자원을 이용하는 FDD(Frequency Division Duplex) 방식을 지원할 수도 있다.

[0024] 본 개시에서, 채널을 전송 또는 수신한다는 것은 해당 채널을 통해서 정보 또는 신호를 전송 또는 수신한다는 의미를 포함한다. 예를 들어, 제어 채널을 전송한다는 것은, 제어 채널을 통해서 제어 정보 또는 신호를 전송한다는 것을 의미한다. 유사하게, 데이터 채널을 전송한다는 것은, 데이터 채널을 통해서 데이터 정보 또는 신호를 전송한다는 것을 의미한다.

[0025] 이하에서는 본 개시에 따른 확장가능한 심층 신경망 훈련에 대해서 설명한다.

[0026] 본 개시의 예시들은 무선 통신 시스템(예를 들어, MIMO 무선 통신 시스템)에서 개체의 개수를 기반으로 설명하지만, 본 개시의 범위가 이에 제한되는 것은 아니다. 즉, 무선 통신 시스템이 아닌 다른 타입의 시스템에서도 본 개시의 예시들이 적용될 수 있으며, 임의의 시스템을 구성하는 임의의 개체 및/또는 자원의 개수가 가변적인 경우에 추가적인 오프라인 훈련 없이 모두 대응할 수 있는 확장가능한 및/또는 범용적인 심층 신경망 훈련에 대해 본 개시의 예시들이 적용될 수 있다.

[0027] 이하의 예시들에서 무선 통신 시스템의 개체의 개수의 대표적인 예시로서 사용자 수를 기반으로 설명한다. 그러나, 본 개시의 예시들은 무선 통신 시스템에서 다른 타입의 개체(예를 들어, 기지국, 송신 안테나, 수신 안테나 등)의 개수를 기반으로 하는 DNN 훈련에도 적용될 수 있고, 무선 통신 시스템의 자원(예를 들어, 시간, 주파수, 코드 도메인 등)의 개수를 기반으로 하는 DNN 훈련에도 적용될 수 있다. 예를 들어, 이하의 설명에서 활성(active) 사용자 수 K 및 사용자 수의 최대값인 K_{max}는 임의의 시스템의 개체 및/또는 자원의 활성 개수 및 그 최대값으로 대체될 수 있다.

[0028] 도 1은 본 개시에 따른 심층 신경망의 입출력 데이터를 나타내는 도면이다.

[0029] 확장가능한 다중안테나 시스템의 최적화를 위해서 최대 사용자 수 K_{max}의 값이 결정될 수 있다. 사용자 수의 최대값과 별도로 활성(active) 사용자 수 K가 정의될 수 있다. 여기서, $K \leq K_{max}$이며, K는 임의의(또는 모든 가능한) 값을 가질 수 있다. 이에 따라, 심층 신경망의 입력으로 사용될 K_{max} 개의 기본 입력 데이터가 생성되고, 이와 함께 임의의 K 값에 해당하는 활성 입력 데이터가 생성될 수 있다. 즉, K_{max} 개의 사용자에 대한 데이터 중에서 K 개의 활성 데이터가 임의로 선택될 수 있다. 나머지(즉, $K_{max} - K$) 개의 데이터는 비활성화(inactive) 데이터로 정의된다. 비활성화 데이터는 0으로 이뤄진 행렬(즉, 0 행렬)에 해당할 수 있다.

[0030] 도 1(a)는 심층신경망의 오프라인 훈련을 설명하기 위한 도면이다.

[0031] 도 1(a)의 예시에서 K_{max} 개의 입력 데이터 $X_1, X_2, \ldots, X_{Kmax}$ 중에서 활성(active) 데이터의 개수 K가 랜덤하게 결정될 수 있다. 또한, K_{max} 개의 입력 데이터 중에서 어떤 데이터가 활성 데이터에 해당하는지(즉, 활성 데이터의 지정)가 랜덤하게 결정될 수 있다. 나머지 $K_{max} - K$ 개의 입력 데이터는 비활성 데이터로서 0 행렬로 취급한다.

[0032] 랜덤하게 결정된 K 개의 활성 데이터에 대해서 오프라인 훈련을 실행하여, 출력 데이터 Y를 획득할 수 있다. 여기서, 출력 데이터의 최대 개수는 도 1(a)에서는 입력 데이터 개수인 K_{max}와 동일한 예시를 나타내지만, 입력 데이터의 최대 개수와 출력 데이터의 (최대) 개수가 반드시 동일할 필요는 없다. 즉, 시스템 특성에 따라서 적절한 크기의 최대 입력 데이터 개수 및 (최대) 출력 데이터 개수가 각각 설정될 수 있다.

[0033] 활성 데이터의 개수 K는 다양한 값으로 결정될 수 있다. 예를 들어, 활성 데이터의 개수 K는 $0, 1, \ldots, K_{max}$ 중의 하나의 값으로 결정될 수 있다. 또한, 제 1 의 K 값에 대해서 신경망 오프라인 훈련이 수행되고, 제 2 의 K 값

에 대해서 신경망 오프라인 훈련이 수행되고, ..., K=Kmax인 경우에 대해서도 신경망 오프라인 훈련이 수행될 수 있다. 이러한 신경망 오프라인 훈련은 가능한 모든 K 값(즉, 0,1,...,K_{max})에 대해서 수행될 수 있다.

[0034] 활성 데이터의 개수가 동일하더라도, 다양한 조합의 활성 데이터들이 결정될 수 있다. 예를 들어, 본 개시에서 K 값에 대응하는 하나 이상의 활성 데이터는, ${}_{Kmax}C_K$ 개의 조합 중의 하나에 해당할 수 있다. 또한, 제 1 의 K 개의 조합에 대해서 신경망 오프라인 훈련이 수행되고, 제 2 의 K 개의 조합에 대해서 신경망 오프라인 훈련이 수행되고, ..., 제 ${}_{Kmax}C_K$ 의 K 개의 조합에 대해서 신경망 오프라인 훈련이 수행될 수 있다. 이러한 신경망 오프라인 훈련은 가능한 모든 조합의 경우의 수(즉, ${}_{Kmax}C_K$)에 대해서 수행될 수 있다.

[0035] 추가적으로 또는 대안적으로, 활성 데이터의 개수의 가능한 모든 경우와, 활성 데이터의 조합의 가능한 모든 경우 중에서, 주로 활용될 것으로 예측되는 일부의 경우에 대해서 신경망 훈련이 수행될 수도 있다.

[0036] 추가적으로 또는 대안적으로, 최대 사용자 수 K_{max}를 설정한 후, 각 사용자에 대한 입력 값을 확률적으로 부여하거나 제외할 수도 있다. 즉, 사용자 입력 데이터 $X_1,X_2,...,X_{Kmax}$ 의 각각이 0 행렬일 확률(또는 0 행렬이 아닐 확률)이 설정되고, 그에 따라 각각의 사용자 입력 데이터의 값이 결정될 수도 있다.

[0037] 본 개시에서와 같이 신경망 훈련 단계에서 입력 데이터의 개수 및/또는 입력 데이터의 조합을 확률적으로 및/또는 랜덤하게 결정하여 훈련을 수행하는 경우, 불완전한 입력에 대한 강인한 출력을 얻을 수 있다. 특정 사용자 개수로만 훈련을 수행하는 경우에는 해당 사용자 개수에 편향된 훈련이 수행되므로, 가능한 많은 사용자 개수의 경우의 수 및 사용자 입력 데이터의 조합의 경우의 수에 대한 훈련을 수행함으로써, 보편적이며 강인한 출력 데이터를 기대할 수 있다.

[0038] 즉, 본 개시의 랜덤한 입력 데이터의 개수 및 조합에 기반한 훈련은, 일반적인 신호처리 기법에서 일부 입력을 0으로 처리하여 입력을 무시 및/또는 제거함으로써 불완전한 및/또는 대략적인 출력을 획득하게 되는 것과는 달리, 불완전한 입력으로부터도 강인한 및/또는 정밀한 출력을 신경망을 통하여 획득할 수 있다.

[0039] 또한, 본 개시는 다양한 개수 및 조합의 입력 데이터에 대한 훈련을 마친 하나의 고정된 규모의(즉, 입력 데이터의 개수가 K_{max}개인) 신경망을 제공함으로써, 하나의 신경망을 활용하면서도 기존의 고정된 개수의 입력을 요구하는 문제를 해결하는 효과도 제공할 수 있다.

[0040] 도 1(b)는 오프라인 훈련을 마친 심층 신경망의 온라인 실행을 설명하기 위한 도면이다.

[0041] K_{max} 개의 최대 입력 데이터 개수 중에서 K^* 개($K^* \leq K_{max}$)의 유효 사용자에 대한 최적화된 출력이 필요한 경우를 가정한다. 이 경우, 도 1(b)에서와 같이 K_{max}--K^* 개의 사용자에 대한 입력 데이터를 0 행렬로 취급할 수 있다. 즉, 본 개시에 따른 신경망은 K_{max} 개의 입력을 기반으로 설계 및 훈련되어 있으므로, 실제 온라인 실행 과정에서 K^* 개의 유효 입력 데이터에 덧붙여서 K_{max}-K^* 개의 0 행렬의 입력, 즉, 전체 K_{max} 개의 입력이 신경망에 입력될 수 있다.

[0042] 이에 따라, 신경망의 출력은 Y_{Kmax} 개가 도출되며, 이 중에서 K^* 개의 유효 사용자에 해당하는 출력 데이터만 추출하여 사용할 수 있다. Y_{Kmax} 개 중에서 K^*를 제외한 나머지 출력 데이터는 사용되지 않을 수 있다. 예를 들어, 추출되는 출력 데이터의 개수 또는 출력 데이터들의 조합은 시스템 특성 또는 운영자의 설정에 따르는 소정의 기준 및/또는 설정에 기초하여 결정될 수 있다.

[0043] 도 1(b)에서는 추출되는 출력 데이터의 조합이 첫 번째 출력 데이터(Y_1)부터 K^* 번째 출력 데이터(Y_{K*})인 예시를 나타내지만, Y_{Kmax} 개의 출력 데이터 중에서 소정의 기준을 만족하는 출력 데이터가 추출될 수 있다. 또한, 추출되는 출력 데이터의 개수가 유효(즉, 0 행렬이 아닌) 입력 데이터의 개수와 동일한 K^* 개인 것을 예시적으로 나타내지만, 유효 입력 데이터의 최대 개수와 추출되는 출력 데이터의 개수가 반드시 동일할 필요는 없다. 또한, 전체 출력 데이터의 개수는 반드시 K_{max} 개일 필요는 없으며, 입력 데이터의 최대 개수와 출력 데이터의 최대 개수가 반드시 동일할 필요는 없다. 즉, 시스템 특성에 따라서 적절한 크기의 (최대) 입력 데이터 개수 및 (최대)

출력 데이터 개수가 각각 설정될 수 있다.

[0044] 도 2는 본 개시가 적용될 수 있는 무선 통신 시스템의 최적화를 위한 심층 신경망의 예시를 나타낸다.

[0045] 예를 들어, 도 2의 예시는 다중사용자-다중안테나(MU-MIMO) 시스템의 선형 프리코더 설계를 위한 신경망 구조에 해당할 수 있다.

[0046] 도 2(a)의 구성요소에서 $\mathcal{G}^{\mathrm{BS}}(\cdot)$ 는 채널 입력 H에 대해서 프리코더 V를 출력하는 연산을 나타낸다. 프리코더 V는 프리코더 산출식 $\mathcal{V}(\cdot)$ 에 기초하여 획득되며, 산출식 $\mathcal{V}(\cdot)$ 에 입력되는 행렬 U 및 W는 각각 $\mathcal{F}_{\mathrm{U}}(\cdot;\theta_{\mathrm{U}})$ 및 $\mathcal{F}_{\mathrm{W}}(\cdot;\theta_{\mathrm{W}})$ 에 기초하여 추출될 수 있다. 또한, 기지국의 프리코더 획득 연산 $\mathcal{G}^{\mathrm{BS}}(\cdot)$ 에 사용되는 파라미터는 Θ_{W}, Θ_{U}, 및 Θ_{β} 를 포함할 수 있다.

[0047] 각 단말로부터 전달 받은 채널 정보를 모두 연계(concatenation)한 결과 H(즉, $H=[H_1^T,\ldots,H_K^T]^T$)를 입력으로 하고, 모든 단말을 위한 프리코더 V (즉, $V=[V_1,\ldots,V_K]$)라는 최종적인 출력을 구하는데 필요한 모든 연산을 $\mathcal{G}^{\mathrm{BS}}(\cdot)$ 으로 정의한다.

[0048] 본 개시에서 심층 신경망을 이용해 얻고자 하는 것은, 가중치 행렬 W, 수신 필터 행렬 U, 및 정규화 상수 β이다. 가중치 행렬 W, 및 수신 필터 행렬 U는 각각 완전 연결 레이어로 구성된 심층 신경망을 통해 얻어낸다.

[0049] 심층 신경망에 사용되는 입력은 J (즉, $J=[H^H,V_{RZF}]$)이며, 여기서 V_{RZF} 는 RZF(regularized zero forcing) 프리코더이다.

[0050] 도 2(a)의 $\mathcal{F}_{\mathrm{W}}(\cdot;\theta_{\mathrm{W}})$ 로 정의된 완전 연결 레이어의 출력을 $\hat{\mathbf{W}}=[\hat{\mathbf{W}}_1,\cdots,\hat{\mathbf{W}}_K]$ 라 할 수 있고, $\hat{\mathbf{W}}=\mathcal{F}_{\mathrm{W}}(\mathbf{J};\theta_{\mathrm{W}})$. 와 같이 표현될 수 있다.

[0051] 수신 필터 행렬 U는 $\mathcal{F}_{\mathrm{U}}(\cdot;\theta_{\mathrm{U}})$ 로 정의된 완전 연결 레이어의 출력으로 얻을 수 있으며, $\mathbf{U}=\mathcal{F}_{\mathrm{U}}(\mathbf{J};\theta_{\mathrm{U}})$. 로 표현된다.

[0052] 두 개의 완전 연결 레이어 신경망으로부터 얻은 행렬들을 바탕으로 프리코딩 행렬 V를 구할 수 있다. 여기서 채널 추정 에러, 제한된 채널 피드백 상황을 대비해 신경망 파라미터 Θ_{β} 를 추가적인 정규화 상수로 추가하였다.

[0053] 도 2(a)의 예시에서 사용되는 모든 심층 신경망 파라미터를 $\Theta_{\mathrm{BS}}\triangleq\{\theta_{\mathrm{W}},\theta_{\mathrm{U}},\theta_{\beta}\}$. 라 정의했을 때, 두 완전 연결 레이어 심층 신경망의 입력 J를 채널 H로 나타낼 수 있으므로 프리코더 행렬 계산을 위한 모든 연산은

$$\mathbf{V} \triangleq \mathcal{G}_{\mathrm{BS}}(\mathbf{H}; \Theta_{\mathrm{BS}})$$

와 같이 정의될 수 있다.

[0054] 위와 같이 구성된 MU-MIMO 선형 프리코더 획득을 위한 심층 신경망은 오프라인 훈련을 거칠 수 있다. 즉, 심층 신경망의 입력 데이터 $H_1,H_2,...,H_K$의 최대 개수를 K_{max}로 설정하고, 입력 데이터의 개수 K 및 입력 데이터의 조합을 랜덤하게 또는 확률적으로 결정하여 다양한 개수 및 조합에 대해서 심층신경망을 훈련시킬 수 있다.

[0055] 훈련은 M 회 수행될 수 있으며, M 번째 훈련에 사용된 파라미터들인 $\Theta_{BS}^{[M]}$ 가 훈련완료된 심층신경망의 파라미터로서 적용될 수 있다. 도 2(b)와 같이 훈련 완료된 심층 신경망은 K_{max} 이하의 다양한 개수 및 다양한 조합의 입력에 대해서 강인한 출력 데이터인 프리코더를 제공할 수 있다.

[0056] 도 3은 본 개시에 따른 확장가능한 심층 신경망 훈련 및 적용의 예시를 설명하기 위한 도면이다.

[0057] 도 3(a)에서 M 회의 훈련 단계(training batch)가 수행되는 것을 가정한다. m 번째 훈련 단계에서 입력되는 데이터는 $H_{(m)1},H_{(m)2},...,H_{(m)Kmax}$ 이다. 즉, 모든 훈련 단계에서 입력 데이터(활성 및 비활성 데이터 합산)의 개수는 K_{max} 개로 동일하다. 각각의 훈련 단계에서 K_{max} 개의 입력 데이터 중에서 K 개는 활성 데이터이고, 나머지는 0 행렬로 대체될 수 있다. K 개의 개수 및 K 개의 활성 데이터의 조합은 훈련 단계마다 상이하게 적용될 수 있다. 나아가 동일한 K 개의 활성 데이터의 동일한 조합에 대해서도 반복적인 훈련이 수행될 수도 있다. 소정의 손실함수 등에 기초하여, 오프라인 훈련 과정이 완료될 수 있다.

[0058] 도 3(b)에서 실제 온라인 실행의 경우, K_{max} 개의 입력 데이터 중에서 유효한 및/또는 활성인 K' 개의 입력 데이터 및 나머지 K_{max}-K' 개의 비활성(즉, 0 행렬) 입력 데이터가 함께 입력될 수 있다. 심층신경망은 이미 K'의 입력 데이터의 경우에 대해서도 훈련완료된 상태이므로, 이에 따른 최적의 프리코더 V를 출력할 수 있다.

[0059] 도 4는 본 개시의 예시들에 따른 시뮬레이션 결과를 나타내는 도면이다.

[0060] 도 4는 32개의 기지국 안테나와 단일 안테나를 사용하는 K 명의 사용자로 구성된 다중안테나 시스템의 전송률 합(sum-rate)을 나타내며, 채널 모델로 3 개의 경로를 가진 밀리미터파 대역(mmWave) 채널 모델이 사용됐다. 본 개시에서 제안하는 방법대로 훈련된 경우의 전송률 합 성능이, 각 사용자 수에 대해 훈련된 경우의 성능과 유사하게 나오는 것을 관찰할 수 있다.

[0061] 도 5는 본 개시에 따른 심층신경망 훈련 방법을 설명하기 위한 흐름도이다.

[0062] 단계 S510에서 입력 데이터의 최대 개수(K_{max})에 기초한 하나의 심층신경망(DNN)을 준비할 수 있다.

[0063] 단계 S520에서 상기 하나의 DNN에 대해서, 최대 개수(K_{max}) 이하의 활성 입력 데이터의 제 1 개수 및 상기 제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및 최대 개수(K_{max}) 이하의 활성 입력 데이터의 제 2 개수 및 상기 제 2 개수의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행할 수 있다.

[0064] 이에 추가적으로, 상기 하나의 DNN에 대한 상기 입력 데이터는, 최대 개수(K_{max}) 이하의 활성 입력 데이터의 하나 이상의 추가적인 개수(제3, 제4, ... 개수), 및 상기 하나 이상의 추가적인 개수(제3, 제4, ... 개수)의 활성 입력 데이터의 하나 이상의 조합을 더 포함할 수 있다.

[0065] 즉, 하나의 동일한 DNN에 대해서 다양한(예를 들어, 랜덤 또는 확률 기반) 활성 입력 데이터의 개수에 대한 훈련이 수행될 수 있다. 다양한 활성 입력 데이터의 개수는 최대 개수 이하의 가능한 모든 개수에 해당할 수도 있다.

[0066] 또한, 각각의 활성 입력 데이터의 개수에 대해서 다양한(예를 들어, 랜덤 또는 확률 기반) 조합의 활성 입력 데이터에 대한 훈련이 수행될 수 있다. 다양한 활성 입력 데이터의 조합의 개수는 해당 개수에서 가능한 모든 조합의 경우의 수에 해당할 수도 있다.

[0067] 이에 따라 상기 하나의 DNN의 훈련은 다양한 개수 및 조합의 활성 입력 데이터에 대해서 완료될 수 있다.

[0068] 단계 S530에서 상기 하나의 DNN에 대한 훈련이 완료된 후, 상기 최대 개수(K_{max}) 이하의 소정의 개수의 유효 입력 데이터에 기초하여 출력 데이터를 추출할 수 있다.

[0069] 이러한 본 개시의 예시들에 따르면, 심층신경망 훈련에 대한 확장가능한 새로운 기법을 통하여 심층신경망의 범용성이 증가할 수 있다. 이에 따라, 심층신경망의 재설계 및 재훈련에 소요되는 시간 및 비용을 크게 감소시킬 수 있다.

[0070] 도 6은 본 개시에 따른 심층신경망 장치의 구성을 나타내는 도면이다.

[0071] 심층신경망 장치(600)는 프로세서(610), 안테나부(620), 트랜시버(630), 메모리(640)를 포함할 수 있다.

[0072] 프로세서(610)는 베이스밴드 관련 신호 처리를 수행하며, 상위계층 처리부(611) 및 물리계층 처리부(615)를 포함할 수 있다. 상위계층 처리부(611)는 MAC 계층, RRC 계층, 또는 그 이상의 상위계층의 동작을 처리할 수 있다. 물리계층 처리부(615)는 PHY 계층의 동작(예를 들어, 상향링크/하향링크/사이드링크 상의 송신/수신 신호 처리 등)을 처리할 수 있다. 프로세서(610)는 베이스밴드 관련 신호 처리를 수행하는 것 외에도, 심층신경망 장치(600) 전반의 동작을 제어할 수도 있다.

[0073] 안테나부(620)는 하나 이상의 물리적 안테나를 포함할 수 있고, 복수개의 안테나를 포함하는 경우 MIMO 송수신을 지원할 수 있다. 트랜시버(630)는 RF 송신기와 RF 수신기를 포함할 수 있다. 메모리(640)는 프로세서(610)의 연산 처리된 정보, 심층신경망 장치(600)의 동작에 관련된 소프트웨어, 운영체제, 애플리케이션 등을 저장할 수 있으며, 버퍼 등의 구성요소를 포함할 수도 있다.

[0074] 심층신경망 장치(600)의 프로세서(610)는 본 발명에서 설명하는 실시예들에서의 심층신경망 오프라인 훈련 및 온라인 적용 동작을 구현하도록 설정될 수 있다.

[0075] 예를 들어, 심층신경망 장치(600)의 프로세서(610)의 상위계층 처리부(611)는 심층신경망 훈련부를 포함할 수 있다.

[0076] 심층신경망 훈련부는 입력 데이터의 최대 개수에 기반한 하나의 DNN를 제공받을 수 있다. 심층신경망 훈련부는 상기 DNN에 대해서 상기 최대 개수 이하의 활성 입력 데이터의 제 1, 제 2, 제 3, ... 개수, 및 상기 제 1, 제 2, 제 3, ... 개수의 각각의 활성 입력 데이터의 하나 이상의 조합을 포함하는 입력 데이터에 기초한 오프라인 훈련을 수행할 수 있다. 이에 따라, 상기 DNN에 대한 훈련이 완료될 수 있다. 상기 DNN에 대한 훈련이 완료된 후, 심층신경망 장치(600)는 상기 최대 개수 이하의 소정의 개수의 유효 입력 데이터를 상기 DNN에 입력하여, 출력 데이터를 추출할 수 있다.

[0077] 심층신경망 장치(600)의 동작에 있어서 본 발명의 예시들에서 설명한 사항이 동일하게 적용될 수 있으며, 중복되는 설명은 생략한다.

[0078] 본 개시의 예시적인 방법들은 설명의 명확성을 위해서 동작의 시리즈로 표현되어 있지만, 이는 단계가 수행되는 순서를 제한하기 위한 것은 아니며, 필요한 경우에는 각각의 단계가 동시에 또는 상이한 순서로 수행될 수도 있다. 본 개시에 따른 방법을 구현하기 위해서, 예시하는 단계에 추가적으로 다른 단계를 포함하거나, 일부의 단계를 제외하고 나머지 단계를 포함하거나, 또는 일부의 단계를 제외하고 추가적인 다른 단계를 포함할 수도 있다.

[0079] 본 개시의 다양한 실시예는 모든 가능한 조합을 나열한 것이 아니고 본 개시의 대표적인 양상을 설명하기 위한 것이며, 다양한 실시예에서 설명하는 사항들은 독립적으로 적용되거나 또는 둘 이상의 조합으로 적용될 수도 있다.

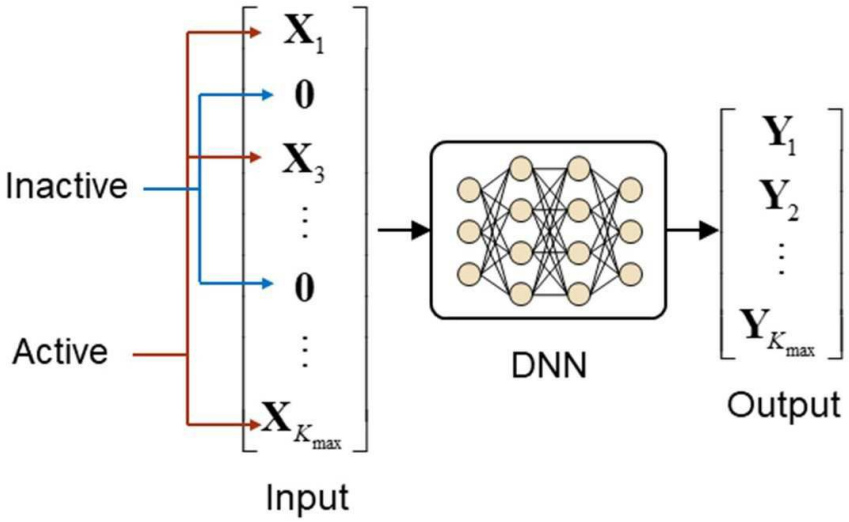
[0080] 또한, 본 개시의 다양한 실시예는 하드웨어, 펌웨어(firmware), 소프트웨어, 또는 그들의 결합 등에 의해 구현될 수 있다. 하드웨어에 의한 구현의 경우, 하나 또는 그 이상의 ASICs(Application Specific Integrated Circuits), DSPs(Digital Signal Processors), DSPDs(Digital Signal Processing Devices), PLDs(Programmable Logic Devices), FPGAs(Field Programmable Gate Arrays), 범용 프로세서(general processor), 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 등에 의해 구현될 수 있다.

[0081] 본 개시의 범위는 다양한 실시예의 방법에 따른 동작이 장치 또는 컴퓨터 상에서 실행되도록 하는 소프트웨어 또는 머신-실행가능한 명령들(예를 들어, 운영체제, 애플리케이션, 펌웨어(firmware), 프로그램 등), 및 이러한 소프트웨어 또는 명령 등이 저장되어 장치 또는 컴퓨터 상에서 실행 가능한 비-일시적 컴퓨터-판독가능 매체(non-transitory computer-readable medium)를 포함한다.

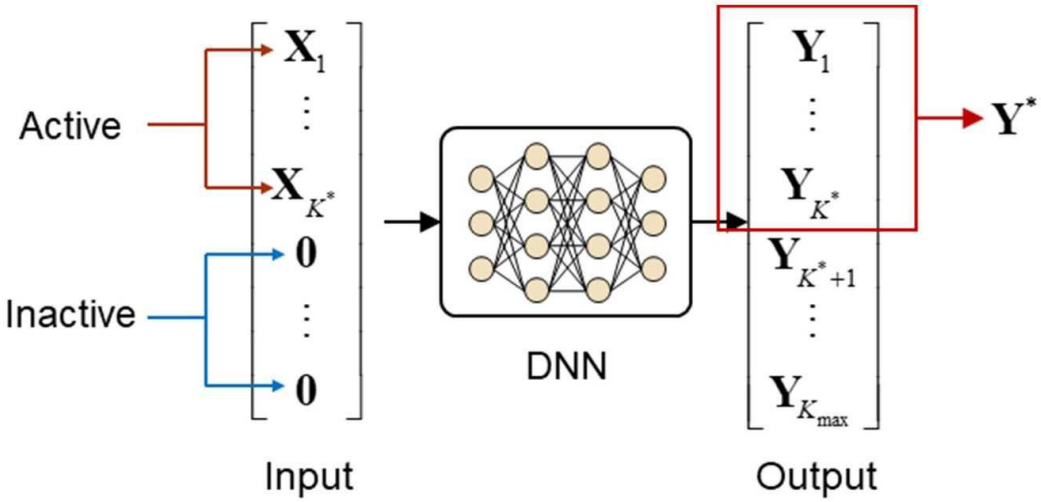
도면

도면1

(a)

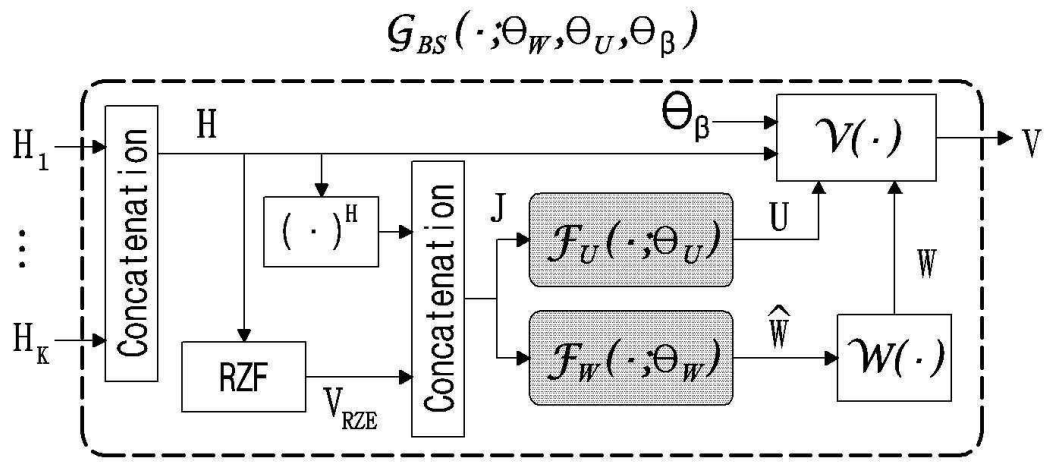

Inactive
Active
Input
DNN
Output

(b)

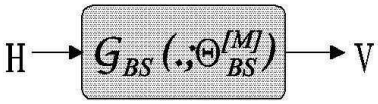

Active
Inactive
Input
DNN
Output

도면2

(a)

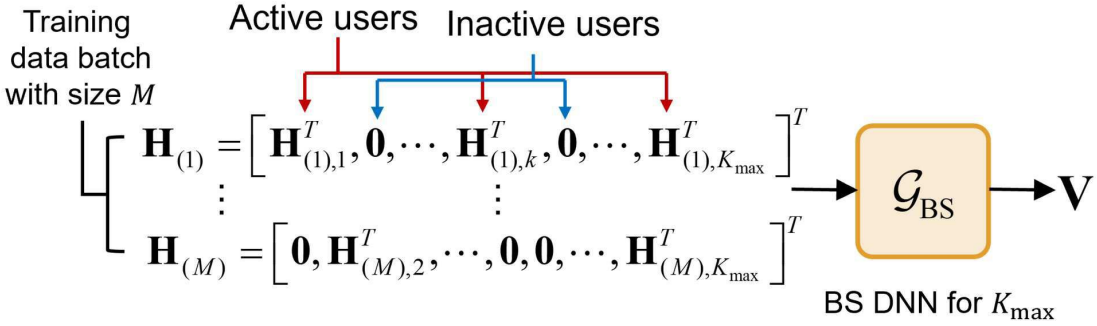

$\mathcal{G}_{BS}(\cdot;\Theta_W,\Theta_U,\Theta_\beta)$
H_1
H_K
Concatenation
H
$(\cdot)^H$
RZF
V_{RZE}
Concatenation
J
$\mathcal{F}_U(\cdot;\Theta_U)$
$\mathcal{F}_W(\cdot;\Theta_W)$
U
$\hat{W}$
$\mathcal{W}(\cdot)$
W
Θ_β
$\mathcal{V}(\cdot)$
V

(b)

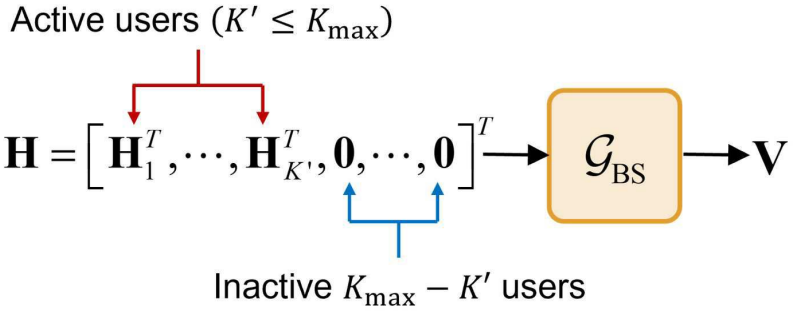

H
$\mathcal{G}_{BS}(\cdot;\Theta_{BS}^{[M]})$
V

도면3

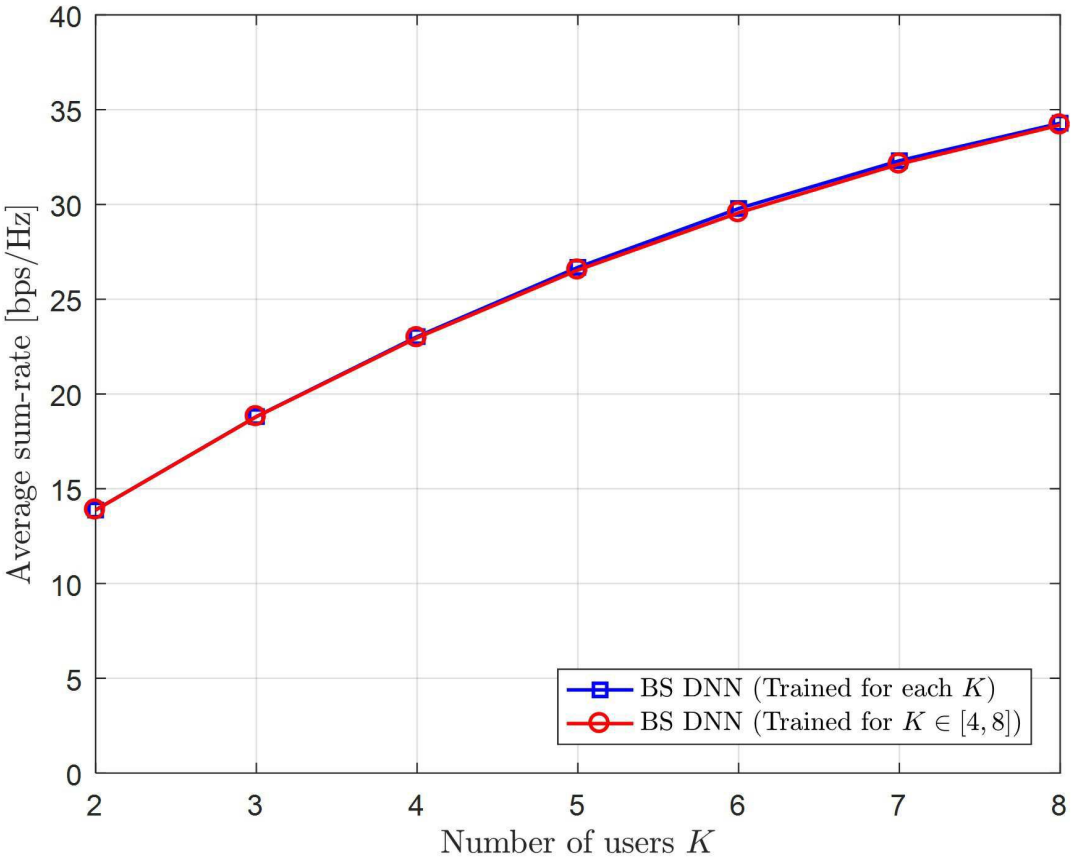
(a)


Training data batch with size M
Active users
Inactive users
$\mathbf{H}_{(1)} = \left[\mathbf{H}_{(1),1}^T, \mathbf{0}, \cdots, \mathbf{H}_{(1),k}^T, \mathbf{0}, \cdots, \mathbf{H}_{(1),K_{\max}}^T\right]^T$
$\mathbf{H}_{(M)} = \left[\mathbf{0}, \mathbf{H}_{(M),2}^T, \cdots, \mathbf{0}, \mathbf{0}, \cdots, \mathbf{H}_{(M),K_{\max}}^T\right]^T$
$\mathcal{G}_{\text{BS}}$
V
BS DNN for $K_{\max}$

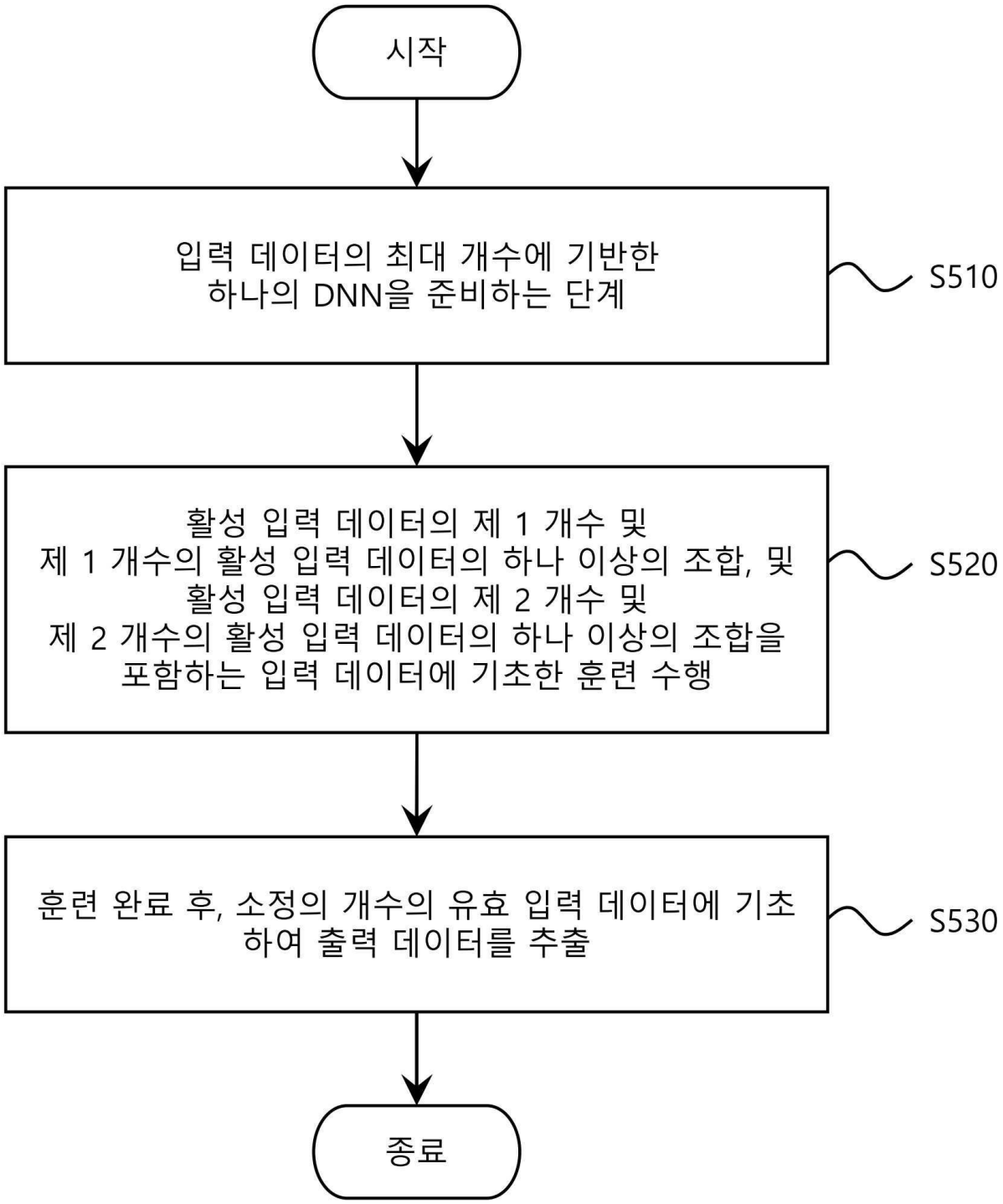
(b)


Active users ($K' \le K_{\max}$)
$\mathbf{H} = \left[\mathbf{H}_1^T, \cdots, \mathbf{H}_{K'}^T, \mathbf{0}, \cdots, \mathbf{0}\right]^T$
$\mathcal{G}_{\text{BS}}$
V
Inactive $K_{\max} - K'$ users

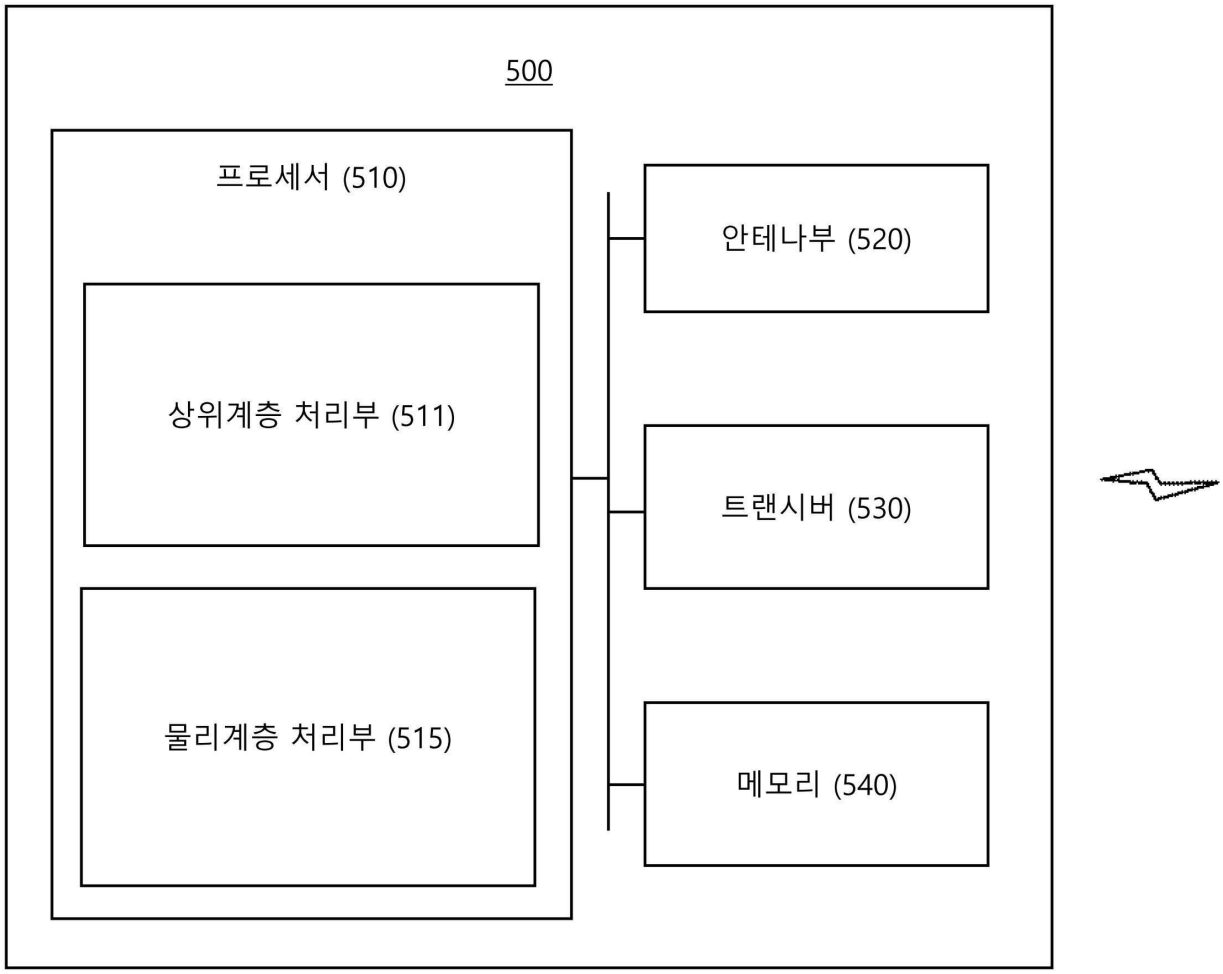
도면4


Average sum-rate [bps/Hz]
Number of users K
BS DNN (Trained for each K)
BS DNN (Trained for $K \in [4, 8]$)
0
5
10
15
20
25
30
35
40
2
3
4
5
6
7
8

도면5


시작
입력 데이터의 최대 개수에 기반한
하나의 DNN을 준비하는 단계
S510
활성 입력 데이터의 제 1 개수 및
제 1 개수의 활성 입력 데이터의 하나 이상의 조합, 및
활성 입력 데이터의 제 2 개수 및
제 2 개수의 활성 입력 데이터의 하나 이상의 조합을
포함하는 입력 데이터에 기초한 훈련 수행
S520
훈련 완료 후, 소정의 개수의 유효 입력 데이터에 기초
하여 출력 데이터를 추출
S530
종료

도면6


500
프로세서 (510)
상위계층 처리부 (511)
물리계층 처리부 (515)
안테나부 (520)
트랜시버 (530)
메모리 (540)