

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구

국제사무국

(43) 국제공개일

2025 년 4 월 10 일 (10.04.2025)



(10) 국제공개번호

WO 2025/075223 A1

(51) 국제특허분류:

G06N 3/096 (2023.01)

G06N 3/042 (2023.01)

G06N 3/047 (2023.01)

G06N 5/04 (2006.01)

(21) 국제출원번호:

PCT/KR2023/015353

(22) 국제출원일:

2023 년 10 월 5 일 (05.10.2023)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(71) 출원인: 엘지전자 주식회사 (LG ELECTRONICS INC.) [KR/KR]; 07336 서울특별시 영등포구 여의대로 128 (KR).

(72) 발명자: 이경호 (LEE, Kyungho); 06772 서울특별시 서초구 양재대로11길 19 LG전자 특허센터 (KR). 이상림 (LEE, Sangrim); 06772 서울특별시 서초구 양재대로11길 19 LG전자 특허센터 (KR). 조민석 (JO, Minseok); 06772 서울특별시 서초구 양재대로11길 19 LG전자 특허센터 (KR).

(74) 대리인: 특허법인 로알 (ROYAL PATENT & LAW OFFICE); 06175 서울특별시 강남구 테헤란로108길 11 삼호빌딩 5층 (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의

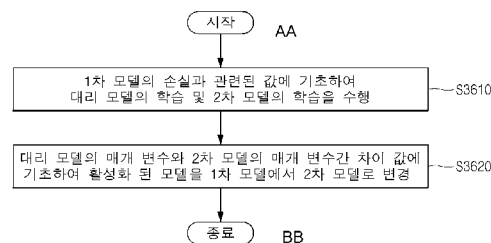
역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: METHOD FOR OPERATING AI/ML MODEL IN WIRELESS COMMUNICATION SYSTEM, AND APPARATUS THEREFOR

(54) 발명의 명칭: 무선 통신 시스템에서 AI/ML 모델의 운영 방법 및 그 장치



S3610 ... Train surrogate model and train secondary model on basis of value related to loss in primary model
S3620 ... Change activated model from primary model to secondary model on basis of difference value between parameter of surrogate model and parameter of secondary model
AA ... Start
BB ... End

(57) Abstract: This method performed by a first wireless device in a wireless communication system supporting an AI/ML model comprises the steps of: training a surrogate model and training a secondary model on the basis that a value related to loss in a primary model is greater than a first threshold value; and, on the basis of a value indicating the difference between a parameter of the surrogate model and a parameter of the secondary model, changing, from the primary model to the secondary model, a model activated for performing a task related to wireless communication.

(57) 요약서: AI/ML 모델을 지원하는 무선 통신 시스템에서 제1 무선 기기에 의해 수행되는 방법은 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델의 학습과 2차 모델의 학습을 수행하는 단계 및 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계를 포함한다.



WO 2025/075223 A1

명세서

발명의 명칭: 무선 통신 시스템에 서 AI/ML 모델의 운영 방법 및 그 장치

기술분야

- [1] 무선 통신 시스템에서 AI/ML 모델의 운영 방법 및 그 장치에 관한 것이다.

배경기술

- [2] 이동 통신 시스템은 사용자의 활동성을 보장하면서 음성 서비스를 제공하기 위해 개발되었다. 그러나 이동통신 시스템은 음성뿐 아니라 데이터 서비스까지 영역을 확장하였으며, 현재에는 폭발적인 트래픽의 증가로 인하여 자원의 부족 현상이 야기되고 사용자들이 보다 고속의 서비스를 요구하므로, 보다 발전된 이동 통신 시스템이 요구되고 있다.
- [3] 차세대 이동 통신 시스템의 요구 조건은 크게 폭발적인 데이터 트래픽의 수용, 사용자 당 전송률의 획기적인 증가, 대폭 증가된 연결 디바이스 개수의 수용, 매우 낮은 단대단 지연(End-to-End Latency), 고에너지 효율을 지원할 수 있어야 한다. 이를 위하여 이중 연결성(Dual Connectivity), 대규모 다중 입출력(Massive MIMO: Massive Multiple Input Multiple Output), 전이중(In-band Full Duplex), 비직교 다중접속(NOMA: Non-Orthogonal Multiple Access), 초광대역(Super wideband) 지원, 단말 네트워킹(Device Networking) 등 다양한 기술들이 연구되고 있다.
- [4] 한편, AI/ML (Artificial Intelligence/Machine Learning)을 지원하는 무선 통신 시스템에서 Deep Neural Network (DNN)가 활용될 수 있다. DNN은 성능이 좋은 반면, 설명 가능성(explainability)이 낮다. 따라서, DNN의 경우, 입력과 출력간의 상관 관계를 정의하기 어렵다. 이러한 문제는 DNN 뿐만 아니라 설명 가능성이 낮은 다른 AI/알고리즘/학습 방법에서도 동일하게 발생한다.

발명의 상세한 설명

기술적 과제

- [5] 상술한 바와 같이 무선 통신을 위해 성능은 뛰어나지만 설명 가능성이 낮은 AI/ML 모델이 활용될 수 있다. 해당 AI/ML 모델의 성능이 저하되는 경우, 성능 저하의 원인을 명확히 밝혀내기가 어려워 무선 통신 품질이 정상적으로 복구되기까지 오랜 시간이 소요될 수 있다.
- [6] 본 명세서의 목적은 AI/ML 모델의 성능 저하시 무선 통신 품질이 저하되는 정도를 최소화 하는 것이다.

과제 해결 수단

- [7] AI/ML (Artificial Intelligence/Machine Learning) 모델을 지원하는 본 명세서의 일 실시예에 따른 무선 통신 시스템에서 제1 무선 기기에 의해 수행되는 방법은 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계 및 상기 대리 모델

의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계를 포함한다. 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이고, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르다. 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초한다. 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이다.

- [8] 상기 1차 모델은 깊은 신경망(Deep Neural Network, DNN) 모델을 포함할 수 있다.
- [9] 상기 2차 모델은 깊은 신경망을 제외한 선형 모델(linear model), 커널 모델(kernel model), 그래프 모델(graph model) 및/또는 확률 모델(probabilistic model)을 포함할 수 있다.
- [10] 상기 1차 모델의 설명 가능성(explainability)은 상기 2차 모델의 설명 가능성보다 낮을 수 있다.
- [11] 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값은 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 유클리드 거리(Euclidean Distance)에 기초하여 계산된 평균 값일 수 있다.
- [12] i) 상기 1차 모델의 손실과 관련된 값이 상기 2차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 큰 것에 기초하여: 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이 상기 1차 모델에서 상기 2차 모델로 변경될 수 있다.
- [13] 상기 방법은 제2 무선 기기에 상기 활성화 된 모델의 변경과 관련된 메시지를 전송하는 단계를 더 포함할 수 있다. 상기 메시지는 상기 대리 모델의 매개 변수 및 상기 2차 모델의 매개 변수를 포함할 수 있다.
- [14] 상기 1차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 PREPARE 상태로 천이될 수 있다. 상기 2차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이될 수 있다.
- [15] 상기 1차 모델의 성능과 관련된 값이 상기 제1 임계값보다 큰 것에 기초하여, 상기 대리 모델 및 상기 2차 모델의 동작과 관련된 상태는 IDLE 상태에서 PREPARE 상태로 천이될 수 있다.
- [16] 상기 대리 모델의 학습 및 상기 2차 모델의 학습은 상기 PREPARE 상태에서 수행될 수 있다.
- [17] 상기 전이 학습은 지식 증류(knowledge distillation)와 관련될 수 있다.
- [18] 상기 무선 통신과 관련된 태스크(task)는 채널 상태 정보(Channel State Information)의 보고, 참조 신호의 측정(measurement of reference signal), 신호 변

조, 빔 트래킹(beam tracking) 또는 빔 탐색(beam search) 중 적어도 하나와 관련될 수 있다.

- [19] 상기 2차 모델의 학습을 위한 훈련 데이터(Training data)의 크기는 상기 1차 모델의 학습을 위한 훈련 데이터의 크기보다 작을 수 있다.
- [20] 상기 2차 모델의 학습을 위한 훈련 데이터는 i) 입력과 ii) 상기 입력에 대응되는 출력을 포함할 수 있다.
- [21] 상기 입력은 상기 제1 무선 기기의 위치, 속도 또는 자세 중 적어도 하나와 관련된 벡터를 포함할 수 있다. 상기 출력은 상기 벡터에 기초하여 결정된 정보를 포함할 수 있다.
- [22] 상기 벡터에 기초하여 결정된 정보는 상기 제1 무선 기기의 전송 또는 수신과 관련된 빔의 인덱스를 포함할 수 있다.
- [23] i) 상기 2차 모델의 손실과 관련된 값이 상기 1차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 작은 것에 기초하여: 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이 상기 2차 모델에서 상기 1차 모델로 변경될 수 있다.
- [24] 상기 1차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이될 수 있다. 상기 2차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 IDLE 상태로 천이될 수 있다.
- [25] 상기 제1 무선 기기는 송신자(Sender), 수신자(Receiver), 단말 또는 기지국일 수 있다.
- [26] AI/ML (Artificial Intelligence/Machine Learning) 모델을 지원하는 본 명세서의 다른 실시예에 따른 무선 통신 시스템에서 동작하는 제1 무선 기기는 하나 이상의 송수신기, 상기 하나 이상의 송수신기를 제어하는 하나 이상의 프로세서들 및 상기 하나 이상의 프로세서들에 동작 가능하게 접속 가능하고, 상기 하나 이상의 프로세서들에 의해 실행될 때 상기 하나 이상의 프로세서들이 동작들을 수행하도록 설정하는 지시(instruction)들을 저장하는 하나 이상의 메모리들을 포함한다.
- [27] 상기 동작들은 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계 및 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계를 포함한다. 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이고, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르다. 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초한다. 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이다.

- [28] 본 명세서의 또 다른 실시예에 따른 장치는 하나 이상의 메모리들 및 상기 하나 이상의 메모리들과 기능적으로 연결되어 있는 하나 이상의 프로세서들을 포함한다.
- [29] 상기 하나 이상의 메모리들은, 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들이 동작들을 수행하도록 설정하는 지시(instruction)들을 저장한다.
- [30] 상기 동작들은 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계 및 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계를 포함한다. 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이고, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르다. 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초한다. 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이다.
- [31] 본 명세서의 또 다른 실시예에 따른 하나 이상의 비일시적(non-transitory) 컴퓨터 판독 가능 매체는 하나 이상의 명령어를 저장한다.
- [32] 상기 하나 이상의 명령어는, 하나 이상의 프로세서에 의해 실행되는 경우, 상기 하나 이상의 프로세서가 동작들을 수행하도록 설정한다.
- [33] 상기 동작들은 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계 및 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계를 포함한다. 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이고, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르다. 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초한다. 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이다.

발명의 효과

- [34] 설명 가능성이 낮은 모델의 성능 저하시 해당 모델의 성능이 복구되기까지 무선 통신 품질이 심하게 열화되거나 통신이 단절될 수 있다. 반면, 본 명세서의 실시예에 의하면, 1차 모델의 성능 저하시 2차 모델에 기초하여 무선 통신 관련 태스크가 수행될 수 있다. 구체적으로 1차 모델의 손실과 관련된 값에 기초하여 대리 모델과 2차 모델의 학습이 수행되며, 대리 모델과 2차 모델의 매개 변수간 차

이 값에 기초하여 무선 통신과 관련된 태스크(task)를 위해 활성화 된 모델이 상기 1차 모델에서 상기 2차 모델로 변경된다.

- [35] 따라서, 설명 가능성이 낮은 AI/ML 모델만 사용하는 경우에 비해 무선 통신의 품질이 저하되는 빈도 내지 시간 구간을 줄일 수 있다. 구체적으로, AI/ML 모델 기반 무선 통신의 전반적인 신뢰성(reliability)이 개선될 수 있다. 또한 설명 가능성이 높은 AI/ML 모델은 입력과 출력의 인과 관계가 모델의 훈련된 매개 변수로 설명될 수 있으므로 그렇지 않은 모델보다 송신 및 수신자간 시스템 문제의 인과 관계를 좀 더 명확히 알 수 있다.
- [36] 기존 방식에 의하면, 단말 혹은 기지국이 부가적인 각종 센서에서 전달되는 부가 정보(side information data)과 연동되어 송수신기가 동작하였을 때 기존 통신 방식인 수학적 기반 모델을 통한 송신, 수신, 측정기로의 대체(fallback) 동작이 수행될 수 없다. 상술한 바와 같이 송수신기가 부가 정보 데이터와 연동하여 수학적 기반 모델링이 한정적이거나 불가능한 경우에도, 본 실시예에 의하면 상기 fallback 동작이 지원될 수 있다.
- [37] 본 명세서에서 얻을 수 있는 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 명세서가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [38] 이하에 첨부되는 도면들은 본 명세서에 관한 이해를 돕기 위한 것으로, 상세한 설명과 함께 본 명세서에 대한 실시 예들을 제공할 수 있다. 다만, 본 명세서의 기술적 특징이 특정 도면에 한정되는 것은 아니며, 각 도면에서 개시하는 특징들은 서로 조합되어 새로운 실시 예로 구성될 수 있다. 각 도면에서의 참조 번호(reference numerals)들은 구조적 구성요소(structural elements)를 의미할 수 있다.
- [39] 도 1은 본 명세서에 적용 가능한 통신 시스템 예시를 나타낸 도면이다.
- [40] 도 2는 본 명세서에 적용 가능한 무선 기기의 예시를 나타낸 도면이다.
- [41] 도 3은 본 명세서에 적용 가능한 전송 신호를 처리하는 방법을 나타낸 도면이다.
- [42] 도 4는 본 명세서에 적용 가능한 무선 기기의 다른 예시를 나타낸 도면이다.
- [43] 도 5는 본 명세서에 적용 가능한 휴대 기기의 예시를 나타낸 도면이다.
- [44] 도 6은 본 명세서에 적용 가능한 물리 채널들 및 이들을 이용한 신호 전송 방법을 나타낸 도면이다.
- [45] 도 7은 퍼셉트론 구조의 일례를 나타낸 도이다.
- [46] 도 8은 다층 퍼셉트론 구조의 일례를 나타낸 도이다.
- [47] 도 9는 심층 신경망의 일례를 나타낸 도이다.
- [48] 도 10은 컨볼루션 신경망의 일례를 나타낸 도이다.
- [49] 도 11은 컨볼루션 신경망에서의 필터 연산의 일례를 나타낸 도이다.
- [50] 도 12는 순환 루프가 존재하는 신경망 구조의 일례를 나타낸다.

- [51] 도 13은 순환 신경망의 동작 구조의 일례를 나타낸다.
- [52] 도 14는 AI/ML 모델의 기능적인 프레임 워크(functional framework)를 예시한다.
- [53] 도 15는 신경망을 포함하는 채널 디코더를 예시한다.
- [54] 도 16은 송신단과 수신단 모두에 신경망이 구성되는 오토 인코더 방식을 예시한다.
- [55] 도 17은 딥러닝 기반 채널 코딩 및 디코딩 동작을 예시한다.
- [56] 도 18은 무선채널 환경에서 수행되는 학습을 예시한다.
- [57] 도 19는 AI/ML의 설명 가능성과 성능의 상충 관계를 나타낸다.
- [58] 도 20은 본 명세서의 실시예에 따른 송신자(Sender)와 수신자(Receiver)의 구조를 나타낸다.
- [59] 도 21은 본 명세서의 실시예에 따른 primary-secondary AI/ML모델에 대한 지원 여부와 관련된 성능 정보의 전달을 위한 절차를 나타낸 흐름도이다.
- [60] 도 22는 본 명세서의 실시예에 따른 primary-secondary AI/ML모델의 셋업(setup)을 위한 절차를 나타낸 흐름도이다.
- [61] 도 23은 본 명세서의 실시예에 따른 primary-secondary AI/ML모델의 해제(release)을 위한 절차를 나타낸 흐름도이다.
- [62] 도 24는 본 명세서의 실시예에 따른 2차 모델(secondary model)의 상태 천이(state transition)를 나타낸다.
- [63] 도 25는 본 명세서의 실시예에 따른 대리 모델(surrogate model)의 상태 천이(state transition)를 나타낸다.
- [64] 도 26은 본 명세서의 실시예에 따른 1차 모델(primary model)의 상태 천이(state transition)를 나타낸다.
- [65] 도 27은 본 명세서의 실시예에 따른 송신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 이동하는 절차를 나타낸다.
- [66] 도 28은 본 명세서의 실시예에 따른 송신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 이동하는 절차를 나타낸다.
- [67] 도 29는 본 명세서의 실시예에 따른 송신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [68] 도 30은 본 명세서의 실시예에 따른 송신자의 2차 모델이 ACTIVE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [69] 도 31은 본 명세서의 실시예에 따른 수신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 이동하는 절차를 나타낸다.
- [70] 도 32는 본 명세서의 실시예에 따른 수신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 이동하는 절차를 나타낸다.
- [71] 도 33은 본 명세서의 실시예에 따른 수신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [72] 도 34는 본 명세서의 실시예에 따른 수신자의 2차 모델이 ACTIVE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.

[73] 도 35는 본 명세서의 실시예에 따라 수행되는 빔 추적 태스크를 나타낸다.

[74] 도 36은 본 명세서의 실시예에 따라 제1 무선 기기에 의해 수행되는 방법을 설명하기 위한 흐름도이다.

발명의 실시를 위한 형태

[75] 이하의 실시 예들은 본 명세서의 구성요소들과 특징들을 소정 형태로 결합한 것들이다. 각 구성요소 또는 특징은 별도의 명시적 언급이 없는 한 선택적인 것으로 고려될 수 있다. 각 구성요소 또는 특징은 다른 구성요소나 특징과 결합되지 않은 형태로 실시될 수 있다. 또한, 일부 구성요소들 및/또는 특징들을 결합하여 본 명세서의 실시 예를 구성할 수도 있다. 본 명세서의 실시 예들에서 설명되는 동작들의 순서는 변경될 수 있다. 어느 실시 예의 일부 구성이나 특징은 다른 실시 예에 포함될 수 있고, 또는 다른 실시 예의 대응하는 구성 또는 특징과 교체될 수 있다.

[76] 도면에 대한 설명에서, 본 명세서의 요지를 흐릴 수 있는 절차 또는 단계 등은 기술하지 않았으며, 당업자의 수준에서 이해할 수 있을 정도의 절차 또는 단계는 또한 기술하지 아니하였다.

[77] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함(comprising 또는 including)"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, "일(a 또는 an)", "하나(one)", "그(the)" 및 유사 관련어는 본 명세서를 기술하는 문맥에 있어서(특히, 이하의 청구항의 문맥에서) 본 명세서에 달리 지시되거나 문맥에 의해 분명하게 반박되지 않는 한, 단수 및 복수 모두를 포함하는 의미로 사용될 수 있다.

[78] 본 명세서의 실시예들은 기지국과 이동국 간의 데이터 송수신 관계를 중심으로 설명되었다. 여기서, 기지국은 이동국과 직접적으로 통신을 수행하는 네트워크의 종단 노드(terminal node)로서의 의미가 있다. 본 명세서에서 기지국에 의해 수행되는 것으로 설명된 특정 동작은 경우에 따라서는 기지국의 상위 노드(upper node)에 의해 수행될 수도 있다.

[79] 즉, 기지국을 포함하는 다수의 네트워크 노드들(network nodes)로 이루어지는 네트워크에서 이동국과의 통신을 위해 수행되는 다양한 동작들은 기지국 또는 기지국 이외의 다른 네트워크 노드들에 의해 수행될 수 있다. 이때, '기지국'은 고정국(fixed station), Node B, eNB(eNode B), gNB(gNode B), ng-eNB, 발전된 기지국(advanced base station, ABS) 또는 액세스 포인트(access point) 등의 용어에 의해 대체될 수 있다.

[80] 또한, 본 명세서의 실시 예들에서 단말(terminal)은 사용자 기기(user equipment, UE), 이동국(mobile station, MS), 가입자국(subscriber station, SS), 이동 가입자 단

말(mobile subscriber station, MSS), 이동 단말(mobile terminal) 또는 발전된 이동 단말(advanced mobile station, AMS) 등의 용어로 대체될 수 있다.

- [81] 또한, 송신단은 데이터 서비스 또는 음성 서비스를 제공하는 고정 및/또는 이동 노드를 말하고, 수신단은 데이터 서비스 또는 음성 서비스를 수신하는 고정 및/또는 이동 노드를 의미한다. 따라서, 상향링크의 경우, 이동국이 송신단이 되고, 기지국이 수신단이 될 수 있다. 마찬가지로, 하향링크의 경우, 이동국이 수신단이 되고, 기지국이 송신단이 될 수 있다.
- [82] 본 명세서의 실시 예들은 무선 접속 시스템들인 IEEE 802.xx 시스템, 3GPP(3rd Generation Partnership Project) 시스템, 3GPP LTE(Long Term Evolution) 시스템, 3GPP 5G(5th generation) NR(New Radio) 시스템 및 3GPP2 시스템 중 적어도 하나에 개시된 표준 문서들에 의해 뒷받침될 수 있으며, 특히, 본 명세서의 실시 예들은 3GPP TS(technical specification) 38.211, 3GPP TS 38.212, 3GPP TS 38.213, 3GPP TS 38.321 및 3GPP TS 38.331 문서들에 의해 뒷받침될 수 있다.
- [83] 또한, 본 명세서의 실시 예들은 다른 무선 접속 시스템에도 적용될 수 있으며, 상술한 시스템으로 한정되는 것은 아니다. 일 예로, 3GPP 5G NR 시스템 이후에 적용되는 시스템에 대해서도 적용 가능할 수 있으며, 특정 시스템에 한정되지 않는다.
- [84] 즉, 본 명세서의 실시 예들 중 설명하지 않은 자명한 단계들 또는 부분들은 상기 문서들을 참조하여 설명될 수 있다. 또한, 본 명세서에서 개시하고 있는 모든 용어들은 상기 표준 문서에 의해 설명될 수 있다.
- [85] 이하, 본 명세서에 따른 바람직한 실시 형태를 첨부된 도면을 참조하여 상세하게 설명한다. 첨부된 도면과 함께 이하에 개시될 상세한 설명은 본 명세서의 예시적인 실시 형태를 설명하고자 하는 것이며, 본 명세서의 기술 구성이 실시될 수 있는 유일한 실시 형태를 나타내고자 하는 것이 아니다.
- [86] 또한, 본 명세서의 실시 예들에서 사용되는 특정 용어들은 본 명세서의 이해를 돕기 위해서 제공된 것이며, 이러한 특정 용어의 사용은 본 명세서의 기술적 사상을 벗어나지 않는 범위에서 다른 형태로 변경될 수 있다.
- [87] 이하의 기술은 CDMA(code division multiple access), FDMA(frequency division multiple access), TDMA(time division multiple access), OFDMA(orthogonal frequency division multiple access), SC-FDMA(single carrier frequency division multiple access) 등과 같은 다양한 무선 접속 시스템에 적용될 수 있다.
- [88] 하기에서는 이하 설명을 명확하게 하기 위해, 3GPP 통신 시스템(e.g.(예, LTE, NR 등)을 기반으로 설명하지만 본 발명의 기술적 사상이 이에 제한되는 것은 아니다. LTE는 3GPP TS 36.xxx Release 8 이후의 기술을 의미할 수 있다. 세부적으로, 3GPP TS 36.xxx Release 10 이후의 LTE 기술은 LTE-A로 지칭되고, 3GPP TS 36.xxx Release 13 이후의 LTE 기술은 LTE-A pro로 지칭될 수 있다. 3GPP NR은 TS 38.xxx Release 15 이후의 기술을 의미할 수 있다. 3GPP 6G는 TS Release 17 및/

또는 Release 18 이후의 기술을 의미할 수 있다. "xxx"는 표준 문서 세부 번호를 의미한다. LTE/NR/6G는 3GPP 시스템으로 통칭될 수 있다.

[89] 본 명세서에 사용된 배경기술, 용어, 약어 등에 관해서는 본 발명 이전에 공개된 표준 문서에 기재된 사항을 참조할 수 있다. 일 예로, 36.xxx 및 38.xxx 표준 문서를 참조할 수 있다.

[90] 본 명세서에 적용 가능한 통신 시스템

[91] 이로 제한되는 것은 아니지만, 본 명세서에 개시된 다양한 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들은 기기들 간에 무선 통신/연결(예, 5G)을 필요로 하는 다양한 분야에 적용될 수 있다.

[92] 이하, 도면을 참조하여 보다 구체적으로 예시한다. 이하의 도면/설명에서 동일한 도면 부호는 다르게 기술하지 않는 한, 동일하거나 대응되는 하드웨어 블록, 소프트웨어 블록 또는 기능 블록을 예시할 수 있다.

[93] 도 1은 본 명세서에 적용되는 통신 시스템 예시를 도시한 도면이다. 도 1을 참조하면, 본 명세서에 적용되는 통신 시스템(100)은 무선 기기, 기지국 및 네트워크를 포함한다. 여기서, 무선 기기는 무선 접속 기술(예, 5G NR, LTE)을 이용하여 통신을 수행하는 기기를 의미하며, 통신/무선/5G 기기로 지칭될 수 있다. 이로 제한되는 것은 아니지만, 무선 기기는 로봇(100a), 차량(100b-1, 100b-2), XR(extended reality) 기기(100c), 휴대 기기(hand-held device)(100d), 가전(home appliance)(100e), IoT(Internet of Thing) 기기(100f), AI(artificial intelligence) 기기/서버(100g)를 포함할 수 있다. 예를 들어, 차량은 무선 통신 기능이 구비된 차량, 자율 주행 차량, 차량간 통신을 수행할 수 있는 차량 등을 포함할 수 있다. 여기서, 차량(100b-1, 100b-2)은 UAV(unmanned aerial vehicle)(예, 드론)를 포함할 수 있다. XR 기기(100c)는 AR(augmented reality)/VR(virtual reality)/MR(mixed reality) 기기를 포함하며, HMD(head-mounted device), 차량에 구비된 HUD(head-up display), 텔레비전, 스마트폰, 컴퓨터, 웨어러블 디바이스, 가전 기기, 디지털 사이니지(signage), 차량, 로봇 등의 형태로 구현될 수 있다. 휴대 기기(100d)는 스마트폰, 스마트패드, 웨어러블 기기(예, 스마트워치, 스마트글래스), 컴퓨터(예, 노트북 등) 등을 포함할 수 있다. 가전(100e)은 TV, 냉장고, 세탁기 등을 포함할 수 있다. IoT 기기(100f)는 센서, 스마트 미터 등을 포함할 수 있다. 예를 들어, 기지국(120), 네트워크(130)는 무선 기기로도 구현될 수 있으며, 특정 무선 기기(120a)는 다른 무선 기기에게 기지국/네트워크 노드로 동작할 수도 있다.

[94] 무선 기기(100a~100f)는 기지국(120)을 통해 네트워크(130)와 연결될 수 있다. 무선 기기(100a~100f)에는 AI 기술이 적용될 수 있으며, 무선 기기(100a~100f)는 네트워크(130)를 통해 AI 서버(100g)와 연결될 수 있다. 네트워크(130)는 3G 네트워크, 4G(예, LTE) 네트워크 또는 5G(예, NR) 네트워크 등을 이용하여 구성될 수 있다. 무선 기기(100a~100f)는 기지국(120)/네트워크(130)를 통해 서로 통신할 수도 있지만, 기지국(120)/네트워크(130)를 통하지 않고 직접 통신(예, 사이드링크 통신(sidelink communication))할 수도 있다. 예를 들어, 차량들(100b-1, 100b-2)은

직접 통신(예, V2V(vehicle to vehicle)/V2X(vehicle to everything) communication)을 할 수 있다. 또한, IoT 기기(100f)(예, 센서)는 다른 IoT 기기(예, 센서) 또는 다른 무선 기기(100a~100f)와 직접 통신을 할 수 있다.

- [95] 무선 기기(100a~100f)/기지국(120), 기지국(120)/기지국(120) 간에는 무선 통신/연결(150a, 150b, 150c)이 이뤄질 수 있다. 여기서, 무선 통신/연결은 상향/하향링크 통신(150a)과 사이드링크 통신(150b)(또는, D2D 통신), 기지국간 통신(150c)(예, relay, IAB(integrated access backhaul))과 같은 다양한 무선 접속 기술(예, 5G NR)을 통해 이뤄질 수 있다. 무선 통신/연결(150a, 150b, 150c)을 통해 무선 기기와 기지국/무선 기기, 기지국과 기지국은 서로 무선 신호를 송신/수신할 수 있다. 예를 들어, 무선 통신/연결(150a, 150b, 150c)은 다양한 물리 채널을 통해 신호를 송신/수신할 수 있다. 이를 위해, 본 명세서의 다양한 제안들에 기반하여, 무선 신호의 송신/수신을 위한 다양한 구성정보 설정 과정, 다양한 신호 처리 과정(예, 채널 인코딩/디코딩, 변조/복조, 자원 매핑/디매핑 등), 자원 할당 과정 등 중 적어도 일부가 수행될 수 있다.

[96] 본 명세서에 적용 가능한 통신 시스템

- [97] 도 2는 본 명세서에 적용될 수 있는 무선 기기의 예시를 도시한 도면이다.

- [98] 도 2를 참조하면, 제1 무선 기기(200a)와 제2 무선 기기(200b)는 다양한 무선 접속 기술(예, LTE, NR)을 통해 무선 신호를 송수신할 수 있다. 여기서, {제1 무선 기기(200a), 제2 무선 기기(200b)}은 도 1의 {무선 기기(100x), 기지국(120)} 및/또는 {무선 기기(100x), 무선 기기(100x)}에 대응할 수 있다.

- [99] 제1 무선 기기(200a)는 하나 이상의 프로세서(202a) 및 하나 이상의 메모리(204a)를 포함하며, 추가적으로 하나 이상의 송수신기(206a) 및/또는 하나 이상의 안테나(208a)를 더 포함할 수 있다. 프로세서(202a)는 메모리(204a) 및/또는 송수신기(206a)를 제어하며, 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들을 구현하도록 구성될 수 있다. 예를 들어, 프로세서(202a)는 메모리(204a) 내의 정보를 처리하여 제1 정보/신호를 생성한 뒤, 송수신기(206a)을 통해 제1 정보/신호를 포함하는 무선 신호를 전송할 수 있다. 또한, 프로세서(202a)는 송수신기(206a)를 통해 제2 정보/신호를 포함하는 무선 신호를 수신한 뒤, 제2 정보/신호의 신호 처리로부터 얻은 정보를 메모리(204a)에 저장할 수 있다. 메모리(204a)는 프로세서(202a)와 연결될 수 있고, 프로세서(202a)의 동작과 관련한 다양한 정보를 저장할 수 있다. 예를 들어, 메모리(204a)는 프로세서(202a)에 의해 제어되는 프로세스들 중 일부 또는 전부를 수행하거나, 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들을 수행하기 위한 명령들을 포함하는 소프트웨어 코드를 저장할 수 있다. 여기서, 프로세서(202a)와 메모리(204a)는 무선 통신 기술(예, LTE, NR)을 구현하도록 설계된 통신 모듈/회로/칩의 일부일 수 있다. 송수신기(206a)는 프로세서(202a)와 연결될 수 있고, 하나 이상의 안테나(208a)를 통해 무선 신호를 송신 및/또는 수신할 수 있다. 송수신기(206a)는 송신기 및/또는 수신기를 포함할 수 있다. 송수신기(206a)는 RF(radio

frequency) 유닛과 혼용될 수 있다. 본 명세서에서 무선 기기는 통신 모듈/회로/칩을 의미할 수도 있다.

- [100] 제2 무선 기기(200b)는 하나 이상의 프로세서(202b), 하나 이상의 메모리(204b)를 포함하며, 추가적으로 하나 이상의 송수신기(206b) 및/또는 하나 이상의 안테나(208b)를 더 포함할 수 있다. 프로세서(202b)는 메모리(204b) 및/또는 송수신기(206b)를 제어하며, 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들을 구현하도록 구성될 수 있다. 예를 들어, 프로세서(202b)는 메모리(204b) 내의 정보를 처리하여 제3 정보/신호를 생성한 뒤, 송수신기(206b)를 통해 제3 정보/신호를 포함하는 무선 신호를 전송할 수 있다. 또한, 프로세서(202b)는 송수신기(206b)를 통해 제4 정보/신호를 포함하는 무선 신호를 수신한 뒤, 제4 정보/신호의 신호 처리로부터 얻은 정보를 메모리(204b)에 저장할 수 있다. 메모리(204b)는 프로세서(202b)와 연결될 수 있고, 프로세서(202b)의 동작과 관련한 다양한 정보를 저장할 수 있다. 예를 들어, 메모리(204b)는 프로세서(202b)에 의해 제어되는 프로세스들 중 일부 또는 전부를 수행하거나, 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들을 수행하기 위한 명령들을 포함하는 소프트웨어 코드를 저장할 수 있다. 여기서, 프로세서(202b)와 메모리(204b)는 무선 통신 기술(예, LTE, NR)을 구현하도록 설계된 통신 모듈/회로/칩의 일부일 수 있다. 송수신기(206b)는 프로세서(202b)와 연결될 수 있고, 하나 이상의 안테나(208b)를 통해 무선 신호를 송신 및/또는 수신할 수 있다. 송수신기(206b)는 송신기 및/또는 수신기를 포함할 수 있다. 송수신기(206b)는 RF 유닛과 혼용될 수 있다. 본 명세서에서 무선 기기는 통신 모듈/회로/칩을 의미할 수도 있다.

- [101] 이하, 무선 기기(200a, 200b)의 하드웨어 요소에 대해 보다 구체적으로 설명한다. 이로 제한되는 것은 아니지만, 하나 이상의 프로토콜 계층이 하나 이상의 프로세서(202a, 202b)에 의해 구현될 수 있다. 예를 들어, 하나 이상의 프로세서(202a, 202b)는 하나 이상의 계층(예, PHY(physical), MAC(media access control), RLC(radio link control), PDCP(packet data convergence protocol), RRC(radio resource control), SDAP(service data adaptation protocol)와 같은 기능적 계층)을 구현할 수 있다. 하나 이상의 프로세서(202a, 202b)는 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들에 따라 하나 이상의 PDU(Protocol Data Unit) 및/또는 하나 이상의 SDU(service data unit)를 생성할 수 있다. 하나 이상의 프로세서(202a, 202b)는 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들에 따라 메시지, 제어정보, 데이터 또는 정보를 생성할 수 있다. 하나 이상의 프로세서(202a, 202b)는 본 명세서에 개시된 기능, 절차, 제안 및/또는 방법에 따라 PDU, SDU, 메시지, 제어정보, 데이터 또는 정보를 포함하는 신호(예, 베이스밴드 신호)를 생성하여, 하나 이상의 송수신기(206a, 206b)에게 제공할 수 있다. 하나 이상의 프로세서(202a, 202b)는 하나 이상의 송수신기(206a, 206b)로부터 신호(예, 베이스밴드 신호)를 수신할 수 있고, 본 명세서에 개시된

설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들에 따라 PDU, SDU, 메시지, 제어정보, 데이터 또는 정보를 획득할 수 있다.

- [102] 하나 이상의 프로세서(202a, 202b)는 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 또는 마이크로 컴퓨터로 지칭될 수 있다. 하나 이상의 프로세서(202a, 202b)는 하드웨어, 펌웨어, 소프트웨어, 또는 이들의 조합에 의해 구현될 수 있다. 일 예로, 하나 이상의 ASIC(application specific integrated circuit), 하나 이상의 DSP(digital signal processor), 하나 이상의 DSPD(digital signal processing device), 하나 이상의 PLD(programmable logic device) 또는 하나 이상의 FPGA(field programmable gate arrays)가 하나 이상의 프로세서(202a, 202b)에 포함될 수 있다. 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들은 펌웨어 또는 소프트웨어를 사용하여 구현될 수 있고, 펌웨어 또는 소프트웨어는 모듈, 절차, 기능 등을 포함하도록 구현될 수 있다. 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들은 수행하도록 설정된 펌웨어 또는 소프트웨어는 하나 이상의 프로세서(202a, 202b)에 포함되거나, 하나 이상의 메모리(204a, 204b)에 저장되어 하나 이상의 프로세서(202a, 202b)에 의해 구동될 수 있다. 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도들은 코드, 명령어 및/또는 명령어의 집합 형태로 펌웨어 또는 소프트웨어를 사용하여 구현될 수 있다.

- [103] 하나 이상의 메모리(204a, 204b)는 하나 이상의 프로세서(202a, 202b)와 연결될 수 있고, 다양한 형태의 데이터, 신호, 메시지, 정보, 프로그램, 코드, 지시 및/또는 명령을 저장할 수 있다. 하나 이상의 메모리(204a, 204b)는 ROM(read only memory), RAM(random access memory), EPROM(erasable programmable read only memory), 플래시 메모리, 하드 드라이브, 레지스터, 캐쉬 메모리, 컴퓨터 판독 저장 매체 및/또는 이들의 조합으로 구성될 수 있다. 하나 이상의 메모리(204a, 204b)는 하나 이상의 프로세서(202a, 202b)의 내부 및/또는 외부에 위치할 수 있다. 또한, 하나 이상의 메모리(204a, 204b)는 유선 또는 무선 연결과 같은 다양한 기술을 통해 하나 이상의 프로세서(202a, 202b)와 연결될 수 있다.

- [104] 하나 이상의 송수신기(206a, 206b)는 하나 이상의 다른 장치에게 본 명세서의 방법들 및/또는 동작 순서도 등에서 언급되는 사용자 데이터, 제어 정보, 무선 신호/채널 등을 전송할 수 있다. 하나 이상의 송수신기(206a, 206b)는 하나 이상의 다른 장치로부터 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도 등에서 언급되는 사용자 데이터, 제어 정보, 무선 신호/채널 등을 수신할 수 있다. 예를 들어, 하나 이상의 송수신기(206a, 206b)는 하나 이상의 프로세서(202a, 202b)와 연결될 수 있고, 무선 신호를 송수신할 수 있다. 예를 들어, 하나 이상의 프로세서(202a, 202b)는 하나 이상의 송수신기(206a, 206b)가 하나 이상의 다른 장치에게 사용자 데이터, 제어 정보 또는 무선 신호를 전송하도록 제어할 수 있다. 또한, 하나 이상의 프로세서(202a, 202b)는 하나 이상의 송수신기(206a, 206b)가 하나 이상의 다른 장치로부터 사용자 데이터, 제어 정보 또는 무선 신호

를 수신하도록 제어할 수 있다. 또한, 하나 이상의 송수신기(206a, 206b)는 하나 이상의 안테나(208a, 208b)와 연결될 수 있고, 하나 이상의 송수신기(206a, 206b)는 하나 이상의 안테나(208a, 208b)를 통해 본 명세서에 개시된 설명, 기능, 절차, 제안, 방법 및/또는 동작 순서도 등에서 언급되는 사용자 데이터, 제어 정보, 무선 신호/채널 등을 송수신하도록 설정될 수 있다. 본 명세서에서, 하나 이상의 안테나는 복수의 물리 안테나이거나, 복수의 논리 안테나(예, 안테나 포트)일 수 있다. 하나 이상의 송수신기(206a, 206b)는 수신된 사용자 데이터, 제어 정보, 무선 신호/채널 등을 하나 이상의 프로세서(202a, 202b)를 이용하여 처리하기 위해, 수신된 무선 신호/채널 등을 RF 밴드 신호에서 베이스밴드 신호로 변환(Convert)할 수 있다. 하나 이상의 송수신기(206a, 206b)는 하나 이상의 프로세서(202a, 202b)를 이용하여 처리된 사용자 데이터, 제어 정보, 무선 신호/채널 등을 베이스밴드 신호에서 RF 밴드 신호로 변환할 수 있다. 이를 위하여, 하나 이상의 송수신기(206a, 206b)는 (아날로그) 오실레이터 및/또는 필터를 포함할 수 있다.

[105] 도 3은 본 명세서에 적용되는 전송 신호를 처리하는 방법을 도시한 도면이다. 일 예로, 전송 신호는 신호 처리 회로에 의해 처리될 수 있다. 이때, 신호 처리 회로(300)는 스크램블러(310), 변조기(320), 레이어 매핑(330), 프리코더(340), 자원 매핑(350), 신호 생성기(360)를 포함할 수 있다. 이때, 일 예로, 도 3의 동작/기능은 도 2의 프로세서(202a, 202b) 및/또는 송수신기(206a, 206b)에서 수행될 수 있다. 또한, 일 예로, 도 3의 하드웨어 요소는 도 2의 프로세서(202a, 202b) 및/또는 송수신기(206a, 206b)에서 구현될 수 있다. 일 예로, 블록 310~350은 도 2의 프로세서(202a, 202b)에서 구현되고, 블록 360은 도 2의 송수신기(206a, 206b)에서 구현될 수 있으며, 상술한 실시 예로 한정되지 않는다.

[106] 코드워드는 도 3의 신호 처리 회로(300)를 거쳐 무선 신호로 변환될 수 있다. 여기서, 코드워드는 정보블록의 부호화된 비트 시퀀스이다. 정보블록은 전송블록(예, UL-SCH 전송블록, DL-SCH 전송블록)을 포함할 수 있다. 무선 신호는 도 6의 다양한 물리 채널(예, PUSCH, PDSCH)을 통해 전송될 수 있다. 구체적으로, 코드워드는 스크램블러(310)에 의해 스크램블된 비트 시퀀스로 변환될 수 있다. 스크램블에 사용되는 스크램블 시퀀스는 초기화 값에 기반하여 생성되며, 초기화 값은 무선 기기의 ID 정보 등이 포함될 수 있다. 스크램블된 비트 시퀀스는 변조기(320)에 의해 변조 심볼 시퀀스로 변조될 수 있다. 변조 방식은 pi/2-BPSK(pi/2-binary phase shift keying), m-PSK(m-phase shift keying), m-QAM(m-quadrature amplitude modulation) 등을 포함할 수 있다.

[107] 복소 변조 심볼 시퀀스는 레이어 매핑(330)에 의해 하나 이상의 전송 레이어로 매핑될 수 있다. 각 전송 레이어의 변조 심볼들은 프리코더(340)에 의해 해당 안테나 포트(들)로 매핑될 수 있다(프리코딩). 프리코더(340)의 출력 z 는 레이어 매핑(330)의 출력 y 를 $N \times M$ 의 프리코딩 행렬 W 와 곱해 얻을 수 있다. 여기서, N 은 안테나 포트의 개수, M 은 전송 레이어의 개수이다. 여기서, 프리코더(340)는 복소 변조 심볼들에 대한 트랜스폼(transform) 프리코딩(예, DFT(discrete fourier

transform) 변환)을 수행한 이후에 프리코딩을 수행할 수 있다. 또한, 프리코더(340)는 트랜스폼 프리코딩을 수행하지 않고 프리코딩을 수행할 수 있다.

[108] 자원 매핑(350)은 각 안테나 포트의 변조 심볼들을 시간-주파수 자원에 매핑할 수 있다. 시간-주파수 자원은 시간 도메인에서 복수의 심볼(예, CP-OFDMA 심볼, DFT-s-OFDMA 심볼)을 포함하고, 주파수 도메인에서 복수의 부반송파를 포함할 수 있다. 신호 생성기(360)는 매핑된 변조 심볼들로부터 무선 신호를 생성하며, 생성된 무선 신호는 각 안테나를 통해 다른 기기로 전송될 수 있다. 이를 위해, 신호 생성기(360)는 IFFT(inverse fast fourier transform) 모듈 및 CP(cyclic prefix) 삽입기, DAC(digital-to-analog converter), 주파수 상향 변환기(frequency uplink converter) 등을 포함할 수 있다.

[109] 무선 기기에서 수신 신호를 위한 신호 처리 과정은 도 3의 신호 처리 과정(310~360)의 역으로 구성될 수 있다. 일 예로, 무선 기기(예, 도 2의 200a, 200b)는 안테나 포트/송수신기를 통해 외부로부터 무선 신호를 수신할 수 있다. 수신된 무선 신호는 신호 복원기를 통해 베이스밴드 신호로 변환될 수 있다. 이를 위해, 신호 복원기는 주파수 하향 변환기(frequency downlink converter), ADC(analog-to-digital converter), CP 제거기, FFT(fast fourier transform) 모듈을 포함할 수 있다. 이후, 베이스밴드 신호는 자원 디-매핑 과정, 포스트코딩(postcoding) 과정, 복조 과정 및 디-스크램블 과정을 거쳐 코드워드로 복원될 수 있다. 코드워드는 복호(decoding)를 거쳐 원래의 정보블록으로 복원될 수 있다. 따라서, 수신 신호를 위한 신호 처리 회로(미도시)는 신호 복원기, 자원 디-매핑, 포스트코더, 복조기, 디-스크램블러 및 복호기를 포함할 수 있다.

[110] 본 명세서에 적용 가능한 무선 기기 구조

[111] 도 4는 본 명세서에 적용되는 무선 기기의 다른 예시를 도시한 도면이다.

[112] 도 4를 참조하면, 무선 기기(400)는 도 2의 무선 기기(200a, 200b)에 대응하며, 다양한 요소(element), 성분(component), 유닛/부(unit), 및/또는 모듈(module)로 구성될 수 있다. 예를 들어, 무선 기기(400)는 통신부(410), 제어부(420), 메모리부(430) 및 추가 요소(440)를 포함할 수 있다. 통신부는 통신 회로(412) 및 송수신기(들)(414)을 포함할 수 있다. 예를 들어, 통신 회로(412)는 도 2의 하나 이상의 프로세서(202a, 202b) 및/또는 하나 이상의 메모리(204a, 204b)를 포함할 수 있다. 예를 들어, 송수신기(들)(414)는 도 2의 하나 이상의 송수신기(206a, 206b) 및/또는 하나 이상의 안테나(208a, 208b)를 포함할 수 있다. 제어부(420)는 통신부(410), 메모리부(430) 및 추가 요소(440)와 전기적으로 연결되며 무선 기기의 제반 동작을 제어한다. 예를 들어, 제어부(420)는 메모리부(430)에 저장된 프로그램/코드/명령/정보에 기반하여 무선 기기의 전기적/기계적 동작을 제어할 수 있다. 또한, 제어부(420)는 메모리부(430)에 저장된 정보를 통신부(410)을 통해 외부(예, 다른 통신 기기)로 무선/유선 인터페이스를 통해 전송하거나, 통신부(410)를 통해 외부(예, 다른 통신 기기)로부터 무선/유선 인터페이스를 통해 수신된 정보를 메모리부(430)에 저장할 수 있다.

- [113] 추가 요소(440)는 무선 기기의 종류에 따라 다양하게 구성될 수 있다. 예를 들어, 추가 요소(440)는 파워 유닛/배터리, 입출력부(input/output unit), 구동부 및 컴퓨팅부 중 적어도 하나를 포함할 수 있다. 이로 제한되는 것은 아니지만, 무선 기기(400)는 로봇(도 1, 100a), 차량(도 1, 100b-1, 100b-2), XR 기기(도 1, 100c), 휴대 기기(도 1, 100d), 가전(도 1, 100e), IoT 기기(도 1, 100f), 디지털 방송용 단말, 홀로그램 장치, 공공 안전 장치, MTC 장치, 의료 장치, 핀테크 장치(또는 금융 장치), 보안 장치, 기후/환경 장치, AI 서버/기기(도 1, 140), 기지국(도 1, 120), 네트워크 노드 등의 형태로 구현될 수 있다. 무선 기기는 사용-예/서비스에 따라 이동 가능하거나 고정된 장소에서 사용될 수 있다.
- [114] 도 4에서 무선 기기(400) 내의 다양한 요소, 성분, 유닛/부, 및/또는 모듈은 전체가 유선 인터페이스를 통해 상호 연결되거나, 적어도 일부가 통신부(410)를 통해 무선으로 연결될 수 있다. 예를 들어, 무선 기기(400) 내에서 제어부(420)와 통신부(410)는 유선으로 연결되며, 제어부(420)와 제1 유닛(예, 430, 440)은 통신부(410)를 통해 무선으로 연결될 수 있다. 또한, 무선 기기(400) 내의 각 요소, 성분, 유닛/부, 및/또는 모듈은 하나 이상의 요소를 더 포함할 수 있다. 예를 들어, 제어부(420)는 하나 이상의 프로세서 집합으로 구성될 수 있다. 예를 들어, 제어부(420)는 통신 제어 프로세서, 어플리케이션 프로세서(application processor), ECU(electronic control unit), 그래픽 처리 프로세서, 메모리 제어 프로세서 등의 집합으로 구성될 수 있다. 다른 예로, 메모리부(430)는 RAM, DRAM(dynamic RAM), ROM, 플래시 메모리(flash memory), 휘발성 메모리(volatile memory), 비-휘발성 메모리(non-volatile memory) 및/또는 이들의 조합으로 구성될 수 있다.
- [115] 본 명세서가 적용 가능한 휴대 기기
- [116] 도 5는 본 명세서에 적용되는 휴대 기기의 예시를 도시한 도면이다.
- [117] 도 5는 본 명세서에 적용되는 휴대 기기를 예시한다. 휴대 기기는 스마트폰, 스마트패드, 웨어러블 기기(예, 스마트 워치, 스마트 글래스), 휴대용 컴퓨터(예, 노트북 등)를 포함할 수 있다. 휴대 기기는 MS(mobile station), UT(user terminal), MSS(mobile subscriber station), SS(subscriber station), AMS(advanced mobile station) 또는 WT(wireless terminal)로 지칭될 수 있다.
- [118] 도 5를 참조하면, 휴대 기기(500)는 안테나부(508), 통신부(510), 제어부(520), 메모리부(530), 전원공급부(540a), 인터페이스부(540b) 및 입출력부(540c)를 포함할 수 있다. 안테나부(508)는 통신부(510)의 일부로 구성될 수 있다. 블록 510~530/540a~540c는 각각 도 4의 블록 410~430/440에 대응한다.
- [119] 통신부(510)는 다른 무선 기기, 기지국들과 신호(예, 데이터, 제어 신호 등)를 송수신할 수 있다. 제어부(520)는 휴대 기기(500)의 구성 요소들을 제어하여 다양한 동작을 수행할 수 있다. 제어부(520)는 AP(application processor)를 포함할 수 있다. 메모리부(530)는 휴대 기기(500)의 구동에 필요한 데이터/파라미터/프로그램/코드/명령을 저장할 수 있다. 또한, 메모리부(530)는 입/출력되는 데이터/정보 등을 저장할 수 있다. 전원공급부(540a)는 휴대 기기(500)에게 전원을 공급하며,

유/무선 충전 회로, 배터리 등을 포함할 수 있다. 인터페이스부(540b)는 휴대 기기(500)와 다른 외부 기기의 연결을 지원할 수 있다. 인터페이스부(540b)는 외부 기기와의 연결을 위한 다양한 포트(예, 오디오 입/출력 포트, 비디오 입/출력 포트)를 포함할 수 있다. 입출력부(540c)는 영상 정보/신호, 오디오 정보/신호, 데이터, 및/또는 사용자로부터 입력되는 정보를 입력 받거나 출력할 수 있다. 입출력부(540c)는 카메라, 마이크로폰, 사용자 입력부, 디스플레이부(540d), 스피커 및/또는 햅틱 모듈 등을 포함할 수 있다.

- [120] 일 예로, 데이터 통신의 경우, 입출력부(540c)는 사용자로부터 입력된 정보/신호(예, 터치, 문자, 음성, 이미지, 비디오)를 획득하며, 획득된 정보/신호는 메모리부(530)에 저장될 수 있다. 통신부(510)는 메모리에 저장된 정보/신호를 무선 신호로 변환하고, 변환된 무선 신호를 다른 무선 기기에게 직접 전송하거나 기지국에게 전송할 수 있다. 또한, 통신부(510)는 다른 무선 기기 또는 기지국으로부터 무선 신호를 수신한 뒤, 수신된 무선 신호를 원래의 정보/신호로 복원할 수 있다. 복원된 정보/신호는 메모리부(530)에 저장된 뒤, 입출력부(540c)를 통해 다양한 형태(예, 문자, 음성, 이미지, 비디오, 햅틱)로 출력될 수 있다.

[121] 물리 채널들 및 일반적인 신호 전송

- [122] 도 6은 본 명세서에 적용되는 물리 채널들 및 이들을 이용한 신호 전송 방법을 도시한 도면이다.

- [123] 전원이 꺼진 상태에서 다시 전원이 켜지거나, 새로이 셀에 진입한 단말은 S611 단계에서 기지국과 동기를 맞추는 등의 초기 셀 탐색(initial cell search) 작업을 수행한다. 이를 위해 단말은 기지국으로부터 주 동기 채널(primary synchronization channel, P-SCH) 및 부 동기 채널(secondary synchronization channel, S-SCH)을 수신하여 기지국과 동기를 맞추고, 셀 ID 등의 정보를 획득할 수 있다.

- [124] 그 후, 단말은 기지국으로부터 물리 방송 채널(physical broadcast channel, PBCH) 신호를 수신하여 셀 내 방송 정보를 획득할 수 있다. 한편, 단말은 초기 셀 탐색 단계에서 하향링크 참조 신호(DL RS: Downlink Reference Signal)를 수신하여 하향링크 채널 상태를 확인할 수 있다. 초기 셀 탐색을 마친 단말은 S612 단계에서 물리 하향링크 제어 채널(physical downlink control channel, PDCCH) 및 물리 하향링크 제어 채널 정보에 따른 물리 하향링크 공유 채널(physical downlink control channel, PDSCH)을 수신하여 조금 더 구체적인 시스템 정보를 획득할 수 있다.

- [125] 이후, 단말은 기지국에 접속을 완료하기 위해 이후 단계 S613 내지 단계 S616과 같은 임의 접속 과정(random access procedure)을 수행할 수 있다. 이를 위해 단말은 물리 임의 접속 채널(physical random access channel, PRACH)을 통해 프리앰블(preamble)을 전송하고(S613), 물리 하향링크 제어 채널 및 이에 대응하는 물리 하향링크 공유 채널을 통해 프리앰블에 대한 RAR(random access response)를 수신할 수 있다(S614). 단말은 RAR 내의 스케줄링 정보를 이용하여 PUSCH(physical uplink shared channel)을 전송하고(S615), 물리 하향링크 제어채

널 신호 및 이에 대응하는 물리 하향링크 공유 채널 신호의 수신과 같은 충돌 해결 절차(contention resolution procedure)를 수행할 수 있다(S616).

[126] 상술한 바와 같은 절차를 수행한 단말은 이후 일반적인 상/하향링크 신호 전송 절차로서 물리 하향링크 제어 채널 신호 및/또는 물리 하향링크 공유 채널 신호의 수신(S617) 및 물리 상향링크 공유 채널(physical uplink shared channel, PUSCH) 신호 및/또는 물리 상향링크 제어 채널(physical uplink control channel, PUCCH) 신호의 전송(S618)을 수행할 수 있다.

[127] 단말이 기지국으로 전송하는 제어정보를 통칭하여 상향링크 제어정보(uplink control information, UCI)라고 지칭한다. UCI는 HARQ-ACK/NACK(hybrid automatic repeat and request acknowledgement/negative-ACK), SR(scheduling request), CQI(channel quality indication), PMI(precoding matrix indication), RI(rank indication), BI(beam indication) 정보 등을 포함한다. 이때, UCI는 일반적으로 PUCCH를 통해 주기적으로 전송되지만, 실시 예에 따라(예, 제어정보와 트래픽 데이터가 동시에 전송되어야 할 경우) PUSCH를 통해 전송될 수 있다. 또한, 네트워크의 요청/지시에 의해 단말은 PUSCH를 통해 UCI를 비주기적으로 전송할 수 있다.

[128] **6G 통신 시스템**

[129] 6G (무선통신) 시스템은 (i) 디바이스 당 매우 높은 데이터 속도, (ii) 매우 많은 수의 연결된 디바이스들, (iii) 글로벌 연결성(global connectivity), (iv) 매우 낮은 지연, (v) 배터리-프리(battery-free) IoT 디바이스들의 에너지 소비를 낮추고, (vi) 초고신뢰성 연결, (vii) 머신 러닝 능력을 가지는 연결된 지능 등에 목적이 있다. 6G 시스템의 비전은 "intelligent connectivity", "deep connectivity", "holographic connectivity", "ubiquitous connectivity"와 같은 4가지 측면일 수 있으며, 6G 시스템은 하기 표 1과 같은 요구 사항을 만족시킬 수 있다. 즉, 표 1은 6G 시스템의 요구 사항을 나타낸 표이다.

[130] [표1]

Per device peak data rate	1 Tbps
E2E latency	1 ms
Maximum spectral efficiency	100bps/Hz
Mobility support	Up to 1000km/hr
Satellite integration	Fully
AI	Fully
Autonomous vehicle	Fully
XR	Fully
Haptic Communication	Fully

[131] 이때, 6G 시스템은 향상된 모바일 브로드밴드(enhanced mobile broadband, eMBB), 초-저지연 통신(ultra-reliable low latency communications, URLLC),

mMTC (massive machine type communications), AI 통합 통신(AI integrated communication), 촉각 인터넷(tactile internet), 높은 스루풋(high throughput), 높은 네트워크 능력(high network capacity), 높은 에너지 효율(high energy efficiency), 낮은 백홀 및 접근 네트워크 혼잡(low backhaul and access network congestion) 및 향상된 데이터 보안(enhanced data security)과 같은 핵심 요소(key factor)들을 가질 수 있다.

[132] **6G 시스템의 핵심 구현 기술**

[133] 인공 지능(Artificial Intelligence)

[134] 6G 시스템에 가장 중요하며, 새로 도입될 기술은 AI이다. 4G 시스템에는 AI가 관여하지 않았다. 5G 시스템은 부분 또는 매우 제한된 AI를 지원할 것이다. 그러나, 6G 시스템은 완전히 자동화를 위해 AI가 지원될 것이다. 머신 러닝의 발전은 6G에서 실시간 통신을 위해 보다 지능적인 네트워크를 만들 것이다. 통신에 AI를 도입하면 실시간 데이터 전송이 간소화되고 향상될 수 있다. AI는 수많은 분석을 사용하여 복잡한 대상 작업이 수행되는 방식을 결정할 수 있다. 즉, AI는 효율성을 높이고 처리 지연을 줄일 수 있다.

[135] 핸드 오버, 네트워크 선택, 자원 스케줄링과 같은 시간 소모적인 작업은 AI를 사용함으로써 즉시 수행될 수 있다. AI는 M2M, 기계-대-인간 및 인간-대-기계 통신에서도 중요한 역할을 할 수 있다. 또한, AI는 BCI(Brain Computer Interface)에서 신속한 통신이 될 수 있다. AI 기반 통신 시스템은 메타 물질, 지능형 구조, 지능형 네트워크, 지능형 장치, 지능형 인지 라디오(radio), 자체 유지 무선 네트워크 및 머신 러닝에 의해 지원될 수 있다.

[136] 최근에는 AI를 무선 통신 시스템과 통합하려고 하는 시도들이 나타나고 있으나, 이는 application layer, network layer 특히, 딥러닝을 wireless resource management and allocation 분야에 집중되어 왔다. 그러나, 이러한 연구는 점점 MAC layer 와 Physical layer로 발전하고 있으며, 특히 물리계층에서 딥러닝을 무선 전송(wireless transmission)과 결합하고자 하는 시도들이 나타나고 있다. AI 기반의 물리계층 전송은, 근본적인 신호 처리 및 통신 메커니즘에 있어서, 전통적인 통신 프레임워크가 아니라 AI 드라이버에 기초한 신호 처리 및 통신 메커니즘을 적용하는 것을 의미한다. 예를 들어, 딥러닝 기반의 채널 코딩 및 디코딩(channel coding and decoding), 딥러닝 기반의 신호 추정(estimation) 및 검출(detection), 딥러닝 기반의 MIMO mechanism, AI 기반의 자원 스케줄링(scheduling) 및 할당(allocation) 등을 포함할 수 있다.

[137] 머신 러닝은 채널 추정 및 채널 트래킹을 위해 사용될 수 있으며, DL(downlink)의 물리 계층(physical layer)에서 전력 할당(power allocation), 간섭 제거(interference cancellation) 등에 사용될 수 있다. 또한, 머신 러닝은 MIMO 시스템에서 안테나 선택, 전력 제어(power control), 심볼 검출(symbol detection) 등에도 사용될 수 있다.

[138] 이하, 머신 러닝에 대해 보다 구체적으로 살펴본다.

- [139] 머신 러닝은 사람이 할 수 있거나 혹은 하기 어려운 작업을 대신해낼 수 있는 기계를 만들어내기 위해 기계를 학습시키는 일련의 동작을 의미한다. 머신 러닝을 위해서는 데이터와 러닝 모델이 필요하다. 머신 러닝에서 데이터의 학습 방법은 크게 3가지 즉, 지도 학습(supervised learning), 비지도 학습(unsupervised learning) 그리고 강화 학습(reinforcement learning)으로 구분될 수 있다.
- [140] 신경망 학습은 출력의 오류를 최소화하기 위한 것이다. 신경망 학습은 반복적으로 학습 데이터를 신경망에 입력시키고 학습 데이터에 대한 신경망의 출력과 타겟의 에러를 계산하고, 에러를 줄이기 위한 방향으로 신경망의 에러를 신경망의 출력 레이어에서부터 입력 레이어 방향으로 역전파(backpropagation) 하여 신경망의 각 노드의 가중치를 업데이트하는 과정이다.
- [141] 지도 학습은 학습 데이터에 정답이 라벨링된 학습 데이터를 사용하며 비지도 학습은 학습 데이터에 정답이 라벨링되어 있지 않을 수 있다. 즉, 예를 들어 데이터 분류에 관한 지도 학습의 경우의 학습 데이터는 학습 데이터 각각에 카테고리가 라벨링된 데이터 일 수 있다. 라벨링된 학습 데이터가 신경망에 입력되고 신경망의 출력(카테고리)과 학습 데이터의 라벨을 비교하여 오차(error)가 계산될 수 있다. 계산된 오차는 신경망에서 역방향(즉, 출력 레이어에서 입력 레이어 방향)으로 역전파 되며, 역전파에 따라 신경망의 각 레이어의 각 노드들의 연결 가중치가 업데이트 될 수 있다. 업데이트 되는 각 노드의 연결 가중치는 학습률(learning rate)에 따라 변화량이 결정될 수 있다. 입력 데이터에 대한 신경망의 계산과 에러의 역전파는 학습 사이클(epoch)을 구성할 수 있다. 학습률은 신경망의 학습 사이클의 반복 횟수에 따라 상이하게 적용될 수 있다. 예를 들어, 신경망의 학습 초기에는 높은 학습률을 사용하여 신경망이 빠르게 일정 수준의 성능을 확보하도록 하여 효율성을 높이고, 학습 후기에는 낮은 학습률을 사용하여 정확도를 높일 수 있다.
- [142] 데이터의 특징에 따라 학습 방법은 달라질 수 있다. 예를 들어, 통신 시스템 상에서 송신단에서 전송한 데이터를 수신단에서 정확하게 예측하는 것을 목적으로 하는 경우, 비지도 학습 또는 강화 학습 보다는 지도 학습을 이용하여 학습을 수행하는 것이 바람직하다.
- [143] 러닝 모델은 인간의 뇌에 해당하는 것으로서, 가장 기본적인 선형 모델을 생각할 수 있으나, 인공 신경망(artificial neural networks)와 같은 복잡성이 높은 신경망 구조를 러닝 모델로 사용하는 머신 러닝의 패러다임을 딥러닝(deep learning)이라 한다.
- [144] 학습(learning) 방식으로 사용하는 신경망 코어(neural network cord)는 크게 심층 신경망(DNN, deep neural networks), 합성곱 신경망(CNN, convolutional deep neural networks), 순환 신경망(RNN, Recurrent Boltzmann Machine) 방식이 있다.
- [145] 인공 신경망(artificial neural network)은 여러 개의 퍼셉트론을 연결한 예시이다.
- [146] 도 7을 참조하면, 입력 벡터 $x=(x_1, x_2, \dots, x_d)$ 가 입력되면 각 성분 x_i 에 가중치 $(W_1, W_2, \dots, W_d)$ 를 곱하고, 그 결과를 모두 합산한 후, 활성화함수 $\sigma()$ 를 적용하는 전

체 과정을 퍼셉트론(perceptron)이라 한다. 거대한 인공 신경망 구조는 도 7에 도시한 단순화된 퍼셉트론 구조를 확장하여 입력벡터를 서로 다른 다 차원의 퍼셉트론에 적용할 수도 있다. 설명의 편의를 위해 입력값 또는 출력값을 노드(node)라 칭한다.

- [147] 한편, 도 7에 도시된 퍼셉트론 구조는 입력값, 출력값을 기준으로 총 3개의 층(layer)로 구성되는 것으로 설명할 수 있다. 1st layer와 2nd layer 사이에는 $(d+1)$ 차원의 퍼셉트론 H 개, 2nd layer와 3rd layer 사이에는 $(H+1)$ 차원 퍼셉트론이 K 개 존재하는 인공신경망을 도 8와 같이 표현할 수 있다.
- [148] 입력벡터가 위치하는 층을 입력층(input layer), 최종 출력값이 위치하는 층을 출력층(output layer), 입력층과 출력층 사이에 위치하는 모든 층을 은닉층(hidden layer)라 한다. 도 8의 예시는 3개의 층이 개시되나, 실제 인공신경망 층의 개수를 카운트할 때는 입력층을 제외하고 카운트하므로 총 2개의 층으로 볼 수 있다. 인공신경망은 기본 블록의 퍼셉트론을 2차원적으로 연결되어 구성된다.
- [149] 전술한 입력층, 은닉층, 출력층은 다층 퍼셉트론 뿐 아니라 후술할 CNN, RNN 등 다양한 인공신경망 구조에서 공동적으로 적용될 수 있다. 은닉층의 개수가 많아질수록 인공신경망이 깊어진 것이며, 충분히 깊어진 인공신경망을 러닝모델로 사용하는 머신러닝 패러다임을 딥러닝(Deep Learning)이라 한다. 또한 딥러닝을 위해 사용하는 인공신경망을 심층 신경망(DNN: Deep neural network)라 한다.
- [150] 도 9에 도시된 심층 신경망은 은닉층+출력층이 8개로 구성된 다층 퍼셉트론이다. 상기 다층 퍼셉트론 구조를 완전 연결 신경망(fully-connected neural network)이라 표현한다. 완전 연결 신경망은 서로 같은 층에 위치하는 노드 간에는 연결 관계가 존재하지 않으며, 인접한 층에 위치한 노드들 간에만 연결 관계가 존재한다. DNN은 완전 연결 신경망 구조를 가지고 다수의 은닉층과 활성화함수들의 조합으로 구성되어 입력과 출력 사이의 상관관계 특성을 파악하는데 유용하게 적용될 수 있다. 여기서 상관관계 특성은 입출력의 결합확률(joint probability)을 의미할 수 있다. 도 9는 심층 신경망 예시를 나타낸 도이다.
- [151] 한편, 복수의 퍼셉트론을 서로 어떻게 연결하느냐에 따라 전술한 DNN과 다른 다양한 인공 신경망 구조를 형성할 수 있다.
- [152] DNN은 하나의 층 내부에 위치한 노드들이 1차원적의 세로 방향으로 배치되어 있다. 그러나, 도 10은 노드들이 2차원적으로 가로 w 개, 세로 h 개의 노드가 배치할 경우를 가정할 수 있다(도 10의 컨볼루션 신경망 구조). 이 경우, 하나의 입력 노드에서 은닉층으로 이어지는 연결과정에서 연결 하나당 가중치가 부가되므로 총 $h \times w$ 개의 가중치를 고려해야한다. 입력층에 $h \times w$ 개의 노드가 존재하므로 인접한 두 층 사이에는 총 $h^2 w^2$ 개의 가중치가 필요하다.
- [153] 도 10은 컨볼루션 신경망의 일례를 나타낸 도이다.
- [154] 도 10의 컨볼루션 신경망은 연결개수에 따라 가중치의 개수가 기하급수적으로 증가하는 문제가 있어 인접한 층 간의 모든 모드의 연결을 고려하는 대신, 크기

가 작은 필터(filter)가 존재하는 것으로 가정하여 도 10에서와 같이 필터가 겹치는 부분에 대해서는 가중합 및 활성화함수 연산을 수행하도록 한다.

- [155] 하나의 필터는 그 크기만큼의 개수에 해당하는 가중치를 가지며, 이미지 상의 어느 특정한 특징을 요인으로 추출하여 출력할 수 있도록 가중치의 학습이 이루어질 수 있다. 도 10에서는 3x3 크기의 필터가 입력층의 가장 좌측 상단 3x3 영역에 적용되고, 해당 노드에 대한 가중합 및 활성화함수 연산을 수행한 결과 출력값을 z_{22} 에 저장한다.
- [156] 상기 필터는 입력층을 스캔하면서 가로,세로 일정 간격 만큼 이동하면서 가중합 및 활성화함수 연산을 수행하고 그 출력값을 현재 필터의 위치에 위치시킨다. 이러한 연산 방식은 컴퓨터 비전(computer vision) 분야에서 이미지에 대한 컨볼루션(convolution) 연산과 유사하여 이러한 구조의 심층 신경망을 컨볼루션 신경망(CNN: convolutional neural network)라 하고, 컨볼루션 연산 결과 생성되는 은닉층을 컨볼루션 층(convolutional layer)라 한다. 또한, 복수의 컨볼루션 층이 존재하는 신경망을 심층 컨볼루션 신경망(DCNN: Deep convolutional)이라 한다.
- [157] 도 11은 컨볼루션 신경망에서의 필터 연산의 일례를 나타낸 도이다.
- [158] 컨볼루션 층에서는 현재 필터가 위치한 노드에서, 상기 필터가 커버하는 영역에 위치한 노드만을 포괄하여 가중합을 계산함으로써, 가중치의 개수를 줄여줄 수 있다. 이로 인해, 하나의 필터가 로컬(local) 영역에 대한 특징에 집중하도록 이용될 수 있다. 이에 따라 CNN은 2차원 영역 상의 물리적 거리가 중요한 판단 기준이 되는 이미지 데이터 처리에 효과적으로 적용될 수 있다. 한편, CNN은 컨볼루션 층의 직전에 복수의 필터가 적용될 수 있으며, 각 필터의 컨볼루션 연산을 통해 복수의 출력 결과를 생성할 수도 있다.
- [159] 한편, 데이터 속성에 따라 시퀀스(sequence) 특성이 중요한 데이터들이 있을 수 있다. 이러한 시퀀스 데이터들의 길이 가변성, 선후 관계를 고려하여 데이터 시퀀스 상의 원소를 매 시점(timestep) 마다 하나씩 입력하고, 특정 시점에 출력된 은닉층의 출력 벡터(은닉 벡터)를, 시퀀스 상의 바로 다음 원소와 함께 입력하는 방식을 인공 신경망에 적용한 구조를 순환 신경망 구조라 한다.
- [160] 도 12를 참조하면, 순환 신경망(RNN: recurrent neural network)은 데이터 시퀀스 상의 어느 시점 t 의 원소 ($x_1(t), x_2(t), \dots, x_d(t)$)를 완전 연결 신경망에 입력하는 과정에서, 바로 이전 시점 $t-1$ 은 은닉 벡터 ($z_1(t-1), z_2(t-1), \dots, z_H(t-1)$)을 함께 입력하여 가중합 및 활성화함수를 적용하는 구조이다. 이와 같이 은닉 벡터를 다음 시점으로 전달하는 이유는 앞선 시점들에서의 입력 벡터속 정보들이 현재 시점의 은닉 벡터에 누적된 것으로 간주하기 때문이다.
- [161] 도 12는 순환 루프가 존재하는 신경망 구조의 일례를 나타낸다.
- [162] 도 12를 참조하면, 순환 신경망은 입력되는 데이터 시퀀스에 대하여 소정의 시점 순서대로 동작한다.
- [163] 시점 1에서의 입력 벡터 ($x_1(1), x_2(1), \dots, x_d(1)$)가 순환 신경망에 입력되었을 때의 은닉 벡터 ($z_1(1), z_2(1), \dots, z_H(1)$)가 시점 2의 입력 벡터 ($x_1(2), x_2(2), \dots, x_d(2)$)와

함께 입력되어 가중합 및 활성화 함수를 통해 은닉층의 벡터 ($z_1(2), z_2(2), \dots, z_H(2)$)를 결정한다. 이러한 과정은 시점 2, 시점 3, ..., 시점 T 까지 반복적으로 수행된다.

[164] 도 13은 순환 신경망의 동작 구조의 일례를 나타낸다.

[165] 한편, 순환 신경망 내에서 복수의 은닉층이 배치될 경우, 이를 심층 순환 신경망(DRNN: Deep recurrent neural network)라 한다. 순환 신경망은 시퀀스 데이터(예를 들어, 자연어 처리(natural language processing)에 유용하게 적용되도록 설계되어 있다.

[166] 학습(learning) 방식으로 사용하는 신경망 코어로서 DNN, CNN, RNN 외에 제한 볼츠만 머신(RBM, Restricted Boltzmann Machine), 심층 신뢰 신경망(DBN, deep belief networks), 심층 Q-네트워크(Deep Q-Network)와 같은 다양한 딥 러닝 기법들을 포함하며, 컴퓨터비전, 음성인식, 자연어처리, 음성/신호처리 등의 분야에 적용될 수 있다.

[167] 도 14는 AI/ML 모델의 기능적인 프레임 워크(functional framework)를 예시한다.

[168] 도 14에 따른 프레임워크에서 각 용어의 정의 및 function별 동작은 다음 표 2에 기반할 수 있다.

[169] [표2]

Data collection: Data collected from the network nodes, management entity or UE, as a basis for ML model training, data analytics and inference.

ML Model: A data driven algorithm by applying machine learning techniques that generates a set of outputs consisting of predicted information, based on a set of inputs

ML Training: An online or offline process to train an ML model by learning features and patterns that best present data and get the trained ML model for inference.

ML Inference: A process of using a trained ML model to make a prediction or guide the decision based on collected data and ML model.

- Data Collection is a function that provides input data to Model training and Model inference functions. AI/ML algorithm specific data preparation (e.g., data pre-processing and cleaning, formatting, and transformation) is not carried out in the Data Collection function. Examples of input data may include measurements from UEs or different network entities, feedback from Actor, output from an AI/ML model.

- Training Data: Data needed as input for the AI/ML Model Training function.

- Inference Data: Data needed as input for the AI/ML Model Inference function.

- Model Training is a function that performs the ML model training, validation, and testing which may generate model performance metrics as part of the model testing procedure. The Model Training function is also responsible for data preparation (e.g. data pre-processing and cleaning, formatting, and transformation) based on Training Data delivered by a Data Collection function, if required.

- Model Deployment/Update: Used to initially deploy a trained, validated, and tested AI/ML model to the Model Inference function or to deliver an updated model to the Model Inference function.

- Note: Details of the Model Deployment/Update process as well as the use case specific AI/ML models transferred via this process are out of RAN3 Rel-17 study scope. The feasibility to single-vendor or multi-vendor environment has not been studied in RAN3 Rel-17 study.

- Model Inference is a function that provides AI/ML model inference output (e.g. predictions or decisions) . Model Inference function may provide model performance feedback to Model Training function when applicable. The Model inference function is also responsible for data preparation (e.g. data pre-processing and cleaning, formatting, and transformation) based on Inference Data delivered by a Data Collection function, if required.

- Output: The inference output of the AI/ML model produced by a Model Inference function.

- Note: Details of inference output are use case specific.

- Model Performance Feedback: It may be used for monitoring the performance of the AI/ML model.

- Note: Details of the Model Performance Feedback process are out of RAN3 scope.

[170]

- Actor is a function that receives the output from the Model inference function and triggers or performs corresponding actions. The Actor may trigger actions directed to other entities or to itself.

- Feedback: Information that may be needed to derive training or inference data or performance feedback.

[171]

Data Set

- [172] AI/ML에서 사용되는 Data set는 Training data, validation data, test data로 구분되며 그 정의는 다음과 같다.
- [173] - Training data
- [174] 모델을 학습하기 위한 Data set
- [175] - Validation data
- [176] 학습이 이미 완료된 모델을 검증하기 위한 Data set
- [177] Validation data는 보통 training data set의 over-fitting을 방지하기 위해서 사용되는 data set이다.
- [178] Validation data는 학습하는 과정에서 학습된 여러 가지 모델 중 best를 선택하기 위한 Data set에 해당한다. 따라서, Validation data는 학습과 관련된 data set로 볼 수 있다.
- [179] - Test data
- [180] 최종 평가를 위한 Data set. Test data는 학습과 무관하다.
- [181] 상기 Data Set의 경우, 상술한 data를 일정 비율로 포함하는 전체 training set가 사용될 수 있다.
- [182] 일 예로, training data와 validation data를 8:2 또는 7:3의 비율로 포함하는 training set가 사용될 수 있다.
- [183] 일 예로, training data, validation data 및 test data를 6:2:2의 비율로 포함하는 training set가 사용될 수 있다.
- [184] Collaboration level
- [185] 기지국과 단말사이의 AI/ML function의 capable여부에 따라 협력레벨은 다음 표 3과 같이 정의될 수 있다.
- [186] [표3]

Cat 0a) No collaboration framework: AI/ML algorithms purely implementation based and not requiring air-interface changes. -> Baseline Cat 0b) No collaboration framework with modified Air-Interface catering to efficient implementation-based AI/ML algorithms. Cat 1) Inter-node assistance to improve the respective nodes' AI/ML algorithms. This would apply to UEs getting assistance from gNBs (for training, adaptation, etc.) and vice-versa. This level does not require model exchange between network nodes. Cat 2) Joint ML operation between UEs and gNBs. This level requires AI/ML model instruction or exchange between network nodes.

- [187] 표 3에 따른 협력 레벨들은 일 예시이며, 구현 방식에 따라 예시된 바와 다르게 변형되어 활용될 수 있다. 예를 들어, 예시된 협력 레벨들 중 둘 이상이 결합된 협력 레벨이 정의/활용될 수 있다. 예를 들어, 예시된 협력 레벨들 중 하나 이상을 제외한 협력 레벨(들)이 활용될 수 있다.
- [188] 인공지능 기반 통신(AI4C, AI for communication)
- [189] 인공지능 기반 통신은 기존의 수학적 모델을 기반한 물리계층 통신에서 데이터 기반한 AI/ML기술을 기반한 통신으로의 전환이다. 이러한 전환은 실제 다루

고자 하는 데이터 분포를 잘 근사하는 딥러닝 기술의 발전에 의한 것이다. 물리 계층의 통신을 수행하는 채널 인코더, 디코더, 채널 추정과 보정, 동기, 파형 송수신기 등의 기능을 일부 대체하거나 범위를 넓혀서 최종적으로 모든 기존 통신이 AI/ML 기술에 의해 대체될 수 있다.

[190] 도 15는 신경망을 포함하는 채널 디코더를 예시한다.

[191] 도 15를 참조하면, 머신러닝을 물리 계층(physical layer)에 적용하려는 시도는 주로 수신단(receiver)의 특정 기능을 최적화하는 것에 중점을 둘 수 있다. 구체적인 일 예로, 채널 디코더(channel decoder)에 디코딩을 위해 신경망을 구성할 수 있다. 이때, 채널 디코더(channel decoder)를 신경망(neural network)으로 구성하는 경우, 채널 디코더의 성능이 향상될 수 있다. 일반적으로 송신단에서 전송되는 신호는 무선 채널(H)를 통과하여 노이즈(noise)가 추가된 수신 신호로 수신단으로 전달될 수 있다. 따라서, 수신단은 노이즈가 추가된 수신 신호에 대한 디코딩을 수행하여 실제 신호를 도출할 수 있다. 일 예로, 기존 채널 디코더는 기 설정된 방식에 따라 디코딩을 수행하여 성능 향상이 제한적일 수 있었다. 반면, 채널 디코더에 신경망을 구성하면 학습을 통해 디코딩 성능을 향상시킬 수 있으며 이를 통해 지연을 줄이고, 성능 향상을 고려할 수 있다.

[192] 또 다른 일 예로, 복수 개의 송수신 안테나를 가진 MIMO 시스템에서 MIMO 디텍터(MIMO detector)를 신경망으로 구현하는 경우, MIMO 시스템의 성능이 향상될 수 있다. 또 다른 일 예로, 오토 인코더(autoencoder)방식이 적용될 수 있다. 오토인코더는 송신단(transmitter) 및 수신단(receiver) 모두를 신경망(neural network)로 구성하여 엔드-투-엔드(end-to-end) 관점에서 최적화를 수행하여 성능 향상을 기하는 방식일 수 있다. 오토 인코더에 기초하여 통신이 수행되는 경우, 송신단 및 수신단 각각은 신경망을 포함할 수 있다. 이때, 송신단 및 수신단은 채널 환경을 포함하여 최적의 통신 설정 및 코딩 기법을 학습할 수 있다. 오토 인코더는 학습을 통해 획득한 정보를 이용하여 인코딩 및 디코딩을 수행할 수 있다. 구체적으로, 송신단 및 수신단 모두 신경망을 포함할 수 있으며, 인코딩 및 디코딩은 페어로서 함께 고려될 수 있다.

[193] 일 예로, 오토 인코더를 사용하는 경우, 통신 시스템은 연결(connection) 환경 또는 상황을 고려하여 최적화된 통신 체인을 형성할 수 있다. 즉, 통신 시스템은 단말 능력(UE capability)이나 채널 특성을 고려하여 최적의 통신 채널(communication channel)을 오토 인코더를 통해 만들 수 있다. 상술한 바에 기초하여 오토 인코더가 동작하는 경우, 통신 시스템은 더 적은 자원으로 더 짧은 시간에 통신을 수행할 수 있다. 일 예로, 대용량 전송(e.g. Tera-bps communication)의 경우, 송신단과 수신단은 초기 연결에 시간이 걸리더라도 최적의 통신환경에 도달하여 재전송 확률을 줄일 수 있다. 도 16은 송신단과 수신단 모두에 신경망이 구성되는 오토 인코더 방식을 예시한다.

[194] 도 17은 딥러닝 기반 채널 코딩 및 디코딩 동작을 예시한다.

- [195] 도 17을 참조하면, 채널 코딩 및 디코딩은 딥러닝 기반으로 수행될 수 있다. 송신기는 채널 환경에 대한 히스토리 정보에 기초하여 딥러닝을 기반으로 적용되는 채널 및 채널에 적용되는 코딩 방식을 결정할 수 있다. 일 예로, 송신기는 참조신호 및 위치정보를 수신기로 전송할 수 있다. 수신기는 참조신호 및 송신기의 위치정보를 이용하여 AI 코딩 초기 값을 생성하고, 이를 송신기로 전달할 수 있다. 송신기는 AI 코딩 초기 값을 고려하여 딥러닝 기반의 채널 및 채널에 적용되는 코딩 방식을 결정하고, 데이터에 대한 인코딩 수행 후 채널을 통해 수신기로 전달할 수 있다. 수신기는 디코딩을 수행하여 오류를 검출하고, 검출된 오류율에 기초하여 리워드(reward) 정보를 생성할 수 있다. 일 예로, 수신기는 한 개의 코드워드에 기초하여 리워드를 생성할 수 있다. 또 다른 일 예로, 수신기는 배치(Batch)의 형태로 여러 개의 블록(Block)을 수신한 이후에 리워드(reward)를 생성하여 송신기로 피드백 할 수 있으며, 상술한 실시예로 한정되지 않는다. 즉, 수신기는 디코딩을 수행한 후, 디코딩 기초한 결과 정보를 리워드 정보로서 송신기로 피드백 할 수 있다. 여기서, 송신기는 피드백된 리워드 정보를 딥러닝 기반 알고리즘에 반영하고, 상술한 과정을 반복하여 채널환경에 가장 적합한 채널 및 채널 코딩 방식을 선택할 수 있다.
- [196] 도 18은 무선채널 환경에서 수행되는 학습을 예시한다.
- [197] 도 18을 참조하면, 무선채널 환경에서 인공지능 시스템에 기초하여 학습이 수행될 수 있다. 일 예로, 인공지능 시스템 에이전트는 획득된 보상 값에 기초하여 강화 학습을 수행할 수 있다. 강화 학습은 두 개의 입력과 한 개의 출력으로 구성될 수 있다. 보다 상세하게는, 두 개의 입력은 상태(state) 정보와 보상(reward) 값일 수 있다. 이때, 상태(State) 정보는 무선 채널 환경에 기초하여 획득되는 인자일 수 있다. 일 예로, 상태 정보는 추정된 채널에 대한 정보일 수 있다. 또한, 보상 값(Reward)은 인공지능 시스템에 전달되는 값으로 인공지능 시스템 에이전트의 행동(action)에 기초하여 업데이트되어 피드백되는 값일 수 있다. 그 후, 인공지능 시스템 에이전트는 두 개의 입력에 기초하여 출력으로써 행동(action)을 도출할 수 있다.
- [198] 앞서 설명한 도 1 내지 도 18에 대한 설명/실시예들은 후술하는 실시예들을 구체화/구현하기 위해 적용될 수 있다.
- [199] 일 예로, 후술하는 실시예에 따른 방법을 수행하는 장치는 도 1 내지 도 5의 장치에 의해 구현될 수 있다.
- [200] 일 예로, 도 6에 기초한 물리 채널/신호 전송과 관련된 동작이 후술하는 실시예들의 동작과 함께 연계되어 수행될 수 있다.
- [201] 일 예로, 후술하는 실시예에 따른 AI/ML 모델은 도 7 내지 도 14에 대한 설명에 기초하여 구현될 수 있다.
- [202] 일 예로, 후술하는 실시예에 따른 수신자/송신자는 도 15 내지 도 16에 기초한 채널 디코더 및/또는 오토인코더를 포함할 수 있다.

- [203] 일 예로, 후술하는 실시예에 따른 방법에 도 17에 따른 채널 코딩 및 디코딩 동작이 결합될 수 있다.
- [204] 일 예로, 후술하는 실시예에 따른 학습은 도 18에 기반하여 수행될 수 있다.
- [205] 이하에서는 AI/ML모델이 쓰이는 통신에서 성능이 좋은 깊은 신경망 모델에 대한 견고한 온라인 학습 방법을 하기 위한 장치와 운용방법을 제안한다. 구체적인 설명에 앞서 후술하는 실시예에 따른 동작과 이에 따른 효과를 간략히 설명하면 다음과 같다.
- [206] 깊은 신경망 모델과 같이 성능은 좋지만 해석이 쉽지 않은 AI/ML모델을 1차 모델(primary model)로 정의한다. 2차 모델(secondary model)은 신경망과 같이 우수하지 않지만 적은 데이터로 훈련이 가능하고 해석이 쉬운 모델로 정의한다. 1차 모델의 성능 저하가 발생하면 채널의 지역성을 이용하여 2차 모델을 학습한다. 2차 모델과 매개변수만 다른 동일한 아키텍처를 가지는 대리 모델(surrogate model)에 1차 모델의 지식 증류를 수행한다. 대리 모델과 2차 모델과 비교를 통해 깊은 신경망 모델의 유효성을 판단하고 관리하여 태스크의 연속성 및 복원성을 향상한다.
- [207] 이하 상술한 실시예와 관련하여 기존 방식의 문제점을 구체적으로 설명한다.
- [208] 6G는 기존 통신 보다 더 다양하고 많은 유즈 케이스(use-case)를 지원해야 한다. 예를 들면 지상국 및 인공위성간의 통신, 무인항공기(UAV)와 지상국 혹은 무인항공기간의 통신, 자동차간의 통신, 작은 기기간의 IoT지원 등이 통합적으로 이루어진다.
- [209] 무선 통신은 6GHz이하 대역, 밀리미터 대역, 테라 헤르츠 대역 등 다양한 주파수 영역에서 이루어 지고 송신 및 수신기 아키텍처가 다양해 짐에 따라 무선 채널 환경이 매우 복잡해진다. 안테나 개수도 많아 지므로 채널의 차원은 높아진다. 채널 확률 분포는 차원이 높아지고 다양성으로 인하여 긴 꼬리 분포를 가질 수 있다.
- [210] 무선 통신 기술이 발전하는 AI/ML기술을 수용함에 따라 채널의 비선형적 환경에서 기존 보다 더 뛰어난 통신 성능을 가질 수 있을 것이다. AI/ML 모델 중 가장 뛰어나 성능을 보여주는 것은 깊은 신경망이다. 위에 설명한 여러가지 다양한 시나리오 아래에서 기존 수학적 모델이 고려하지 못한 비선형적 특징을 잘 학습하고 반영하기 때문이다.
- [211] AI/ML모델은 데이터 기반으로 훈련되어 동작한다. 하지만 단말이 이동함에 따라 훈련되어 있지 않은 통신 채널 환경을 만났을 때 성능이 저하될 것이다. 특히 메모리와 계산 용량이 작은 단말은 한정된 복잡도의 깊은 신경망을 가지므로 이동할 때 새롭게 한번도 학습하지 않은 채널 분포를 만날 확률이 높아진다.
- [212] 깊은 신경망이 이러한 새로운 채널을 만나면 데이터의 절단 등의 통신 품질 저하를 야기 할 수 있다. 이러한 통신 품질 저하의 문제는 물리 계층 태스크가 수행될 때 신호 대 잡음비(Signal to Noise Ratio, SNR), 비트 오류율(bit error rate), 심볼 오류율(symbol error rate), 블록 오류율(block error rate)등의 저하로 감지될 수 있

다. 깊은 신경망 모델이 통신 관련 태스크를 수행하였을 때 상술한 문제가 발생할 수 있다. 특히, 깊은 신경망은 현재 그 성능의 인과관계를 완전히 알 수 없다는 문제가 있다.

- [213] 상기의 문제를 해결하기 위하여 단말에 깊은 신경망 모델을 대체하는 설명 가능한 대리 모델(surrogate model)을 도입하여 송수신기의 운용하는 방법을 제안한다. 깊은 신경망 모델을 상황에 따라 대리 모델을 탄력적으로 운용하여 성능저하를 최소화하는 방법을 제안한다.
- [214] 이하 상술한 실시예를 1) 기술적 배경, 2) 본 실시예와 관련된 개념과 구조 및 3) 본관련 실시예와 관련된 절차 및 4) 본 실시예가 무선 통신 관련 태스크에 대한 적용된 경우 구체적인 동작들 순으로 설명한다.
- [215] 본 실시예의 기술적 배경을 이하 구체적으로 설명한다.
- [216] 통신 채널 환경의 지역성(locality)
- [217] 통신 채널 환경의 지역성은 통신 전파의 특성이 지역별로 더 특색을 가지는 분포를 가지는 것을 의미한다. 밀리미터 전파나 혹은 테라 헤르츠 전파 등 주파수가 높아지면서 전파 직진성이 더 생긴다. 이러한 전파가 다양한 지역적인 환경과 회절, 산란, 반사 특성과 결합하여 지역에 따른 고유의 분포를 가지는 채널의 지역성이 생긴다. 이러한 특성으로 채널 전파 확률 분포 상에서 긴 꼬리를 가진다. 예를 들면 TR 38.901 사양에서는 아래와 같이 전파 채널의 지역성을 볼 수 있다.
- [218] a) O2O(Outdoor-to-Outdoor) 및 O2I(Outdoor-to-Indoor)가 있는 UMi(거리 및 빌딩 숲, 개방 지역)
- [219] b) UMa with O2O and O2I O2I
- [220] c) 실내: 환경, 쇼핑몰. 일반적인 사무실 환경은 개방형 큐비클 영역, 벽으로 둘러싸인 공간으로
- [221] d) 백홀, 도시 지역의 옥외 위 옥상 백홀 및 소형 셀이 있는 거리 협곡 시나리오 포함
- [222] e) D2D/V2V. 열린 공간, 거리 협곡 및 실내 시나리오에서 장치 간 액세스. V2V는 차량간 통신
- [223] f) 경기장(지붕 개방) 및 체육관(지붕 폐쇄)과 같은 기타 시나리오.
- [224] g) 실내 산업 시나리오
- [225] 지역성은 단말이 처한 전파 환경 상에서 통신 채널 확률 분포가 비슷한 영역으로 정의될 수도 있다. 지역성에 따라 밀리미터 파 혹은 테라 헤르츠 대역의 참조 신호 태스크에 영향을 미치는 전파 환경이 구분될 수 있다. 고주파 전파 경로는 저주파 대역의 전파 특성과 달리 직진성을 더 갖는다. 저주파와는 반사, 회절, 굴절이 다른 이러한 고주파 전파 경로 분포는 정적 상태에 있는 빌딩, 지형뿐만 아니라 움직이는 물체, 안테나의 배치, 단말의 움직임에 따라 많은 경우의 수를 가진다. 특정 단말이 위치하는 지형적 분포(geographical distribution)와 단말의 동적 상태가 지역성에 가장 큰 영향을 준다. 이는 지역성이 지형적 분포와 연관될 수 있음을 의미한다.

- [226] 도 19는 AI/ML의 설명 가능성과 성능의 상충 관계를 나타낸다. 도 19를 참조하면, 높은 성능(Performance)을 갖는 학습 방법/AI/기술은 낮은 설명 가능성(explainability)을 갖는다.
- [227] 설명 가능한 AI (explainable AI or XAI)기술은 AI/ML모델의 동작의 이해와 해석 가능성을 높인다. 특히 깊은 신경망을 이용한 기술은 뛰어난 성능이 보이는데 반해 신경망 내부의 동작은 수학적으로 완전히 해석 되지 않고 있다. 설명 가능한 AI 기술은 수학적 접근과 데이터 해석적 접근을 병행하여 이러한 해석 가능성 혹은 설명 가능성을 높이기 위한 기술이다. 설명 가능한 AI 기술은 다양한 방식들이 존재한다. 일 예로, 입력 조합에 대한 출력의 응답을 살펴 봄으로서 입출력의 상관 및 연관 관계성을 밝히는 방식이 존재한다.
- [228] 복잡도와 설명 가능성은 서로 상충(trade-off) 관계를 가진다. 깊은 신경망은 복잡하고 성능이 좋은 반면 설명 가능성이 낮고 선형 모델은 단순하지만 비선형 데이터를 설명하기 쉽지 않은 대신 수학적으로 설명 가능 하다.
- [229] 지식 증류 (Knowledge Distillation)
- [230] 머신 러닝에서 지식 증류는 큰 모델 에서 작은 모델로 지식을 이전하는 프로세스이다. 큰 모델(예: 매우 깊은 신경망 또는 많은 모델의 앙상블)은 작은 모델보다 용량이 더 크다는 점에서 해당 모델을 평가하거나 훈련하는 계산 비용이 많이 들 수 있다. 지식 증류는 모델의 지식의 큰 손실 없이 큰 모델에서 더 작은 모델로 지식을 이전한다. 더 작은 모델은 훈련 비용이 매우 적게 들기 때문에 설명 가능한 모델로 지식을 증류하는데 쓸 수 있다. 모델 $f_{\theta}(x)$ 를 $g_{\psi}(x)$ 로 지식 증류를 수행할 수 있다. 이 때, 지식 증류를 위한 손실 함수 $\mathcal{L}$ 는 지식 증류 손실 함수 $\mathcal{L}_{KD}$ (Knowledge Distillation loss)와 크로스 엔트로피 손실 함수 $\mathcal{L}_{CE}$ (Cross Entropy loss)로 이루어진다(아래 수학적 식 1). 지식 증류는 전이 학습의 일종이며 다른 전이 학습의 방법도 가능하다.
- [231] [수식1]
- $$\mathcal{L} = \mathbb{E}_{p(x,y)}\{\mathcal{L}_{KD}(f_{\theta}(x), g_{\psi}(x)) + \lambda \mathcal{L}_{CE}(y, g_{\psi}(x))\}$$
- [232] 본 실시예와 관련된 제안 개념과 구조를 이하 구체적으로 설명한다.
- [233] 상술한 바와 같이 본 실시예는 성능이 좋은 AI/ML모델인 깊은 신경망이 성능이 저하되었을 때 데이터의 끊임 없이 연속적으로 통신이 수행될 수 있도록 하는 것을 목적으로 한다. 구체적으로, 1차 모델로 동작하는 송신 및 수신자 사이의 태스크의 성능이 저하 되었을 때 비교적 적은 데이터 세트로 학습하고 설명이 가능한 대리 모델과 2차 모델을 통해 이러한 문제를 해결하는 방법 및 장치를 제안한다.
- [234] 기본 구조
- [235] 기존의 대체 동작(fallback operation)과 다른 점은 다음과 같다.
- [236] 기존의 fallback operation은 다음과 같이 수행된다. 깊은 신경망 모델을 대체하기 위해 주어진 태스크 별로 특별히 설계된 수학 기반 모델인 기존 모델(legacy

model)(즉, AI/ML 모델을 기반으로 하지 않는 기존 통신 모델)이 이용된다. 반면, 본 실시예에 의하면 기존 모델이 이용되지 않는다. 구체적으로, 깊은 신경망 모델은 범용적이고 작은 훈련 데이터 세트를 이용하여 설명 가능한 모델로 대체된다.

- [237] 본 실시예에 의하면, 성능이 가장 좋지만 설명하기 힘든 깊은 신경망 모델은 1차 모델(primary model)로 정의된다. 상기 1차 모델의 성능이 저하되는 경우, 데이터 기반의 설명 가능하고 복잡도가 낮은 2차 모델(secondary model)이 활용된다. 본 실시예는 성능과 설명가능성간의 상충 관계를 활용한다는 점에서 단순히 기존 legacy 모델로 fallback 하는 기존 방식과는 다르다.
- [238] 제안한 개념을 수행하기 위한 송신자 및 수신자 구조를 도 20을 참조하여 설명한다.
- [239] 도 20은 본 명세서의 실시예에 따른 송신자(Sender)와 수신자(Receiver)의 구조를 나타낸다.
- [240] 도 20을 참조하면, 송신자(Sender, 2010) 및 수신자(Receiver, 2020)는 양방향(two-sided) 모델일 수 있다. 이 때, 송신자(Sender, 2010) 또는 수신자(Receiver, 2020)만 AI/ML 모델을 가질 수도 있다.
- [241] 1차 모델(primary model)(2012, 2022)은 주로 깊은 신경망 혹은 성능이 좋지만 설명성(explainability)이 낮은 AI/ML 모델이다. 대리 모델(surrogate model)(2013, 2023)과 2차 모델(secondary model)(2014, 2024)은 성능은 보통이지만 설명성이 높은 AI/ML 모델이며 주로 선형 모델로 구성된다. 대리 모델(2013, 2023)은 1차 모델의 지식 증류(knowledge distillation) 혹은 전이 학습의 결과이다. 2차 모델(2014, 2024)은 1차 모델(2012, 2022)과 독립적으로 현재 데이터 셋을 기반으로 온라인 훈련이 수행되는 모델이다.
- [242] 성능 혹은 손실이 지정된 수준 이하로 진입을 하면 1차 모델의 지식 증류 혹은 전이 학습을 수행한 대리 모델과 현재 데이터 셋을 기반으로 온라인 학습을 수행한 2차 모델을 비교한다. 이를 통해 현재의 1차 모델의 동작의 이상 여부를 추정 및 관리를 할 수 있다. 대리 모델과 2차 모델은 같은 모델이며 매개 변수 값만 다르다. 선형 모델을 예로 들면, 대리 모델은 $y=ax+b$ 로 표현되고, 2차 모델은 $y=a'x+b'$ 로 표현될 수 있다. 여기서 대리 모델의 매개 변수(a, b)는 2차 모델의 매개 변수(a', b')와는 다르다. 후술하는 바에 따라 SET_ACT 조건(수학식 3 및 수학식 4) 및 PRI_ACT 조건(수학식 5 및 수학식 6)에 2차 모델과 대리 모델의 매개 변수가 언급된다. 이 때, 해당 매개 변수의 비교(송신자:

ϕ

,

ϕ'

, 수신자:

ξ

- ,
 ξ'
)는 a와 a'(및/또는 b와 b')의 비교에 대응될 수 있다.
- [243] 송신자(2010)는 전송을 위한 AI/ML모델 중 가장 높은 성능을 가진 1차 모델(
 f_{θ}
)(2012)과 1차 모델의 대리 모델 (surrogate model)(
 u_{ϕ}
)(2013) 및 2차 모델(
 $u_{\phi'}$
)(2014)을 포함하도록 구성될 수 있다. 각 모델의 아래 첨자는 해당 모델의 매개 변수이다.
- [244] TX entity(2011)에서 전달하는 송신 데이터는 1차 모델(2012) 및 2차 모델(2014)에 기초하여 전송될 수 있다. 이 때, 두 개의 모델들(2012, 2014) 중 하나에 기초하여 채널
 $h(t)$
 로의 전송이 수행될 수 있다.
- [245] 수신자(2020)는 AI/ML모델인 1차 모델(
 g_{ψ}
)(2022)와 1차 모델의 대리 모델(
 v_{ξ}
)(2023)과 2차 모델(
 $v_{\xi'}$
)(2024)을 포함하도록 구성될 수 있다. 각 모델의 아래 첨자는 모델의 매개 변수이다.
- [246] 수신자(2020)에 도착하는 신호는 1차 모델(2022) 및 2차 모델(2024) 모두에 입력될 수 있지만, 1차 모델(2022) 및 2차 모델(2023) 중 하나의 Task data만 RX entity(2021)에 입력될 수 있다.
- [247] 송수신 태스크의 종류에 따라 데이터 셋
 $D \in (x, y)$
 과 각 모델의 역할이 달라질 수 있다.
- [248] 데이터 송수신 태스크인 경우 송신자/수신자의 1차 모델들(
 f_{θ}
 ,
 g_{ψ}
)(2012, 2022)은 오토 인코더의 인코더, 디코더가 될 수도 있다. 이 경우
 x

와

y

는 송수신 데이터가 같아야 하므로 같은 값을 가진다.

[249] 채널 추정 및 데이터 복원 태스크의 경우, 송신자의 1차 모델(f_{θ})

f_{θ}

(2012)은 전송 측의 프리코더이고 수신자의 1차 모델(g_{ψ})

g_{ψ}

(2022)은 채널 추정기일 수 있다. 이 때

x

는 단말의 상태와 송신 기준 신호로 이루어진 벡터일 수 있고,

y

는 채널 추정 값일 수 있다.

[250] 빔 관리 태스크의 경우, 송신자의 1차 모델(f_{θ})

f_{θ}

(2012)은 전송 빔 선택기일 수 있다. 수신자의 1차 모델(g_{ψ})

g_{ψ}

(2022)은 수신 빔 선택기 일 수 있다. 이 때,

x

는 단말의 상태와 송신 빔 신호로 이루어진 벡터일 수 있고,

y

는 수신 빔일 수 있다. AI/ML모델의 학습 목표는 데이터 셋

$D \in (x, y)$

에 대한 손실 함수

$\mathcal{L}$

에 대한 모델의 매개 변수를 찾는 것이다. 이러한 AI/ML 모델의 학습 목표는 아래 수학식 2로 표현될 수 있다.

[251] [수식2]

$$\theta, \psi = \arg \min E_{p(x, y)} [\mathcal{L}(y, g_{\psi} \circ h \circ f_{\theta}(x))]$$

[252] 수학식 2는 x, y 에 대해 손실 함수

$\mathcal{L}$

의 값이 최소인 송신자의 1차 모델(2012)의 매개 변수

θ

와 수신자의 1차 모델(2022)의 매개 변수

ψ

를 나타낸다.

[253] 본 명세서에서 송신자는 제1 무선 장치(도 2의 200a 또는 200b), 단말 또는 기지국을 의미할 수 있고, 수신자는 제2 무선 장치(도 2의 200a 또는 200b), 단말 또는 기지국을 의미할 수 있다. 일 예로, 송신자는 단말이고 수신자는 기지국일 수 있다. 일 예로, 송신자는 기지국이고 수신자는 단말일 수 있다. 일 예로, 송신자는 단말이고 수신자는 다른 단말일 수 있다.

[254] 채널 지역성

[255] 채널

h

는 상술한 통신 채널 환경의 지역성(locality)이 유지되는 단말의 상태 집합에서 훈련 데이터

D

에 모두 반영된다. 이러한 지역성은 단말의 위치, 자세, 움직임 벡터로 구성된 상태 벡터

x

와 기지국 혹은 수신자의 상태

x'

에 따라 채널을 결정한다. 상기 채널의 결정은 결합 채널 분포

$p(h, x, x')$

를 이용하여 설명될 수 있다.

x'

가 정적인 수신 기지국일 경우 단말 상태

x

에 따라

x'

는 결정될 것이므로 해당 채널 분포는

$p(h, x)$

로 표현될 수 있다. 만약에

x

,

x'

가 주어진다면 지역성에 따라

$p(h|x, x')$

는 로그 노말(log-normal), 라이시안(Rician) 분포 혹은 지형에 따른 특별한 분포를 가질 수 있다. 지역성은 단말의 가능한 상태

x

의 집합

R

로 정의될 수 있다. 집합

R

에 속하는 임의의 두

x

,

x'

는

$$KL(p(x)||p(x')) < \delta$$

을 만족한다. 여기서

KL

은 두 분포 간의 거리인 쿨백-라이블러 발산(Kullback-Leibler divergence)이다.

δ

는 AI/ML 모델의 용량과 지역의 채널 분포의 복잡도에 따라 결정될 수 있다.

- [256] 성능이 가장 좋은 1차 모델인 깊은 신경망을 운용하면서 성능이 지정된 모니터링 문턱 값(threshold) 밑으로 내려 오거나 혹은 특정 타이머에 설정된 윈도우 안에서 문턱 값 밑으로 내려오는 횟수가 지정된 수준을 넘어 서면 송신자 및/또는 수신자는 1차 모델(2012, 2022)을 대리 모델(2014, 2024)로 지식 증류 혹은 전이 학습을 주기적으로 수행할 수 있다. 이러한 대리 모델(2014, 2024)은 학습은 상술한 지역성과 관련된

$$KL(p(x)||p(x')) < \delta$$

을 만족할 수 있도록 주기적으로 수행될 수 있다. 지역성의 성립은 1차 모델의 비선형적인 부분을 2차 모델의 선형 모델로 증류할 때 필요 조건이다. 동시에 2차 모델을 현재 데이터 셋을 기반으로 온라인 학습을 수행한다.

- [257] 본 실시예와 관련된 절차 및 방법을 이하 구체적으로 설명한다.
- [258] Primary and secondary model capability procedure
- [259] 이하에서는 도 21을 참조하여 1차 모델(primary model) 및 2차 모델(secondary model)의 지원 여부와 관련된 성능(capability)의 시그널링을 위한 절차를 구체적으로 설명한다.
- [260] 도 21은 본 명세서의 실시예에 따른 primary-secondary AI/ML 모델에 대한 지원 여부와 관련된 성능 정보의 전달을 위한 절차를 나타낸 흐름도이다.
- [261] 도 21은 송신자와 수신자간 태스크에 대한 협의 지원 가능한 primary-secondary AI/ML모델에 대한 지원 여부와 관련된 정보를 교환하는 과정이다.
- [262] S2110에서, 송신자(Sender)는 수신자(Receiver)에 제1 모델 및 제2 모델의 지원 여부와 관련된 성능 정보를 요청한다. 예를 들어, 송신자(Sender)는 수신자(Receiver)에 primary-secondary model capability request 메시지를 전송할 수 있다.

[263] S2120에서, 수신자(Receiver)는 송신자(Sender)에 제1 모델 및 제2 모델의 지원 여부와 관련된 성능 정보를 전송한다. 예를 들어, 수신자는 송신자에 제1 모델 및 제2 모델의 지원 여부와 관련된 성능 정보로서 primary-secondary model list를 전송할 수 있다.

[264] 일 예로, 상기 primary-secondary model list는 다음 표 4에 기초한 정보를 포함할 수 있다.

[265] [표4]

Information element	Description
Supported primary AI/ML model list	지원되는 Primary AI/ML 모델의 식별자 (ID)(들)을 포함한다. 지정된 태스크에 대해 성능이 가장 좋은 모델(들)의 리스트이다. 해당 리스트에 포함된 모델은 깊은 신경망 모델일 수 있다. 해당 리스트는 입력 및 출력간의 관계의 관계가 수식으로 표현하는 것이 불가능한/어려운 모델(들)을 포함할 수 있다.
Supported secondary AI/ML model list	지원되는 Secondary AI/ML 모델의 식별자 (ID)(들)을 포함한다. Secondary AI/ML 모델(들)의 리스트이다. 해당 리스트는 성능 보다는 입출력의 대응 관계를 쉽게 알 수 있거나 적은 훈련 데이터로 학습할 수 있는 모델(들)을 포함할 수 있다.

[266] AI/ML모델은 식별자로 구분될 수 있다. 식별자에 AI/ML 모델의 종류, 아키텍처, 각종 파라미터, 입력 특징(feature) 및 출력 특징의 구조가 대응될 수 있다. 이러한 정보는 무선 자원의 절약을 위해 사전에 수신자 및 송신자간에 공유될 수 있다. 이후 지원되는 AI/ML 모델에 대한 정보를 전달하기 위해 식별자(들)만을 포함하는 리스트 정보가 수신자로부터 송신자로 전달될 수 있다. 도 21은 송신자의 요청에 기초하여 수신자가 송신자에게 지원되는 모델의 리스트 정보를 전달하는 동작만을 도시하고 있으나, 수신자의 요청에 기초하여 송신자가 수신자에게 지원되는 모델의 리스트 정보를 전달하는 동작도 수행될 수 있다.

[267] Setup procedure for primary and secondary model

[268] 이하에서는 도 22를 참조하여 1차 모델(primary model) 및 2차 모델(secondary model)의 셋업(setup)을 위한 절차를 구체적으로 설명한다.

[269] 도 22는 본 명세서의 실시예에 따른 primary-secondary AI/ML모델의 셋업(setup)을 위한 절차를 나타낸 흐름도이다.

[270] 구체적으로 도 22는 송신자와 수신자간 AI/ML 모델 지원 리스트를 primary secondary model capability procedure에서 서로 확인한 후 무선 통신과 관련된 태스크가 시작되기 전에 1차 모델과 2차 모델을 송신자와 수신자 간에 설정하는 절차를 나타낸다.

[271] S2210에서, 송신자(Sender)는 수신자(Receiver)에 제1/제2 모델의 셋업을 위한 요청한다. 예를 들어, 송신자(Sender)는 수신자(Receiver)에 primary-secondary model setup request 메시지를 전송할 수 있다.

- [272] S2220에서, 수신자(Receiver)는 송신자(Sender)에 제1/제2 모델의 셋업을 확인하는 메시지를 전송한다. 예를 들어, 수신자는 송신자에 primary-secondary model setup request confirm 메시지를 전송할 수 있다.
- [273] 도 21에서 설명한 바와 같이 시그널링의 방향은 도 22와 반대로 수행될 수도 있다. 예를 들어, 수신자의 모델 셋업 요청에 기초하여 송신자가 모델 셋업 확인 메시지를 수신자에 전송할 수 있다.
- [274] 도 22의 절차에서 제1/제2 모델의 셋업을 위한 요청 메시지 또는 제1/제2 모델의 셋업을 확인하는 메시지는 다음 표 5에 기초한 정보를 포함할 수 있다. 이 때, 표 5에 기초한 정보의 일부는 제1/2 모델 온라인 학습 절차(primary and secondary model online learning procedure)에서 사용될 수 있다.
- [275] [표5]

Information element	Description
Task ID	Task에 대한 식별자. beam tracking, beam search, CSI report, channel estimation, data demodulation, reference signal measurement등의 task에 대응 되는 식별자 이다.
Primary AI/ML model	Task에 대한 1차 AI/ML 모델 식별자 (ID)
Secondary AI/ML model list	Task에 대한 2차 AI/ML 모델 식별자 (ID) 리스트
Secondary model prepare criteria	2차 모델의 상태가 IDLE 상태에서 PREPARE 상태로 이동하는 primary loss threshold값 PRE_{TH} 이다. Loss는 task의 종류에 따라 BER, BLER, Accuracy등이 될 수 있다.
Secondary model monitoring period	2차 모델 PREPARE상태에서 온라인 러닝을 수행하고 대리 모델은 지식 종류 혹은 전이학습을 수행하는 주기 이다.
Secondary model activate criteria	2차 모델이 PREPARE상태에서 ACTIVE상태로 이동하는 기준이며 M_{diff} 과 Δ 로 구성한다. ACTIVE상태는 1차 모델 대신 2차 모델로 task처리를 수행한다..
Primary model activate criteria	2차 모델이 ACTIVE상태에서 IDLE상태로 이동하고 1차 모델이 ACTIVE상태로 이동할 때 필요한 Δ 값이다.

- [276] Release procedure for primary and secondary model
- [277] 이하에서는 도 23을 참조하여 1차 모델(primary model) 및 2차 모델(secondary model)의 해제(release)를 위한 절차를 구체적으로 설명한다.
- [278] 도 23은 본 명세서의 실시예에 따른 primary-secondary AI/ML모델의 해제(release)을 위한 절차를 나타낸 흐름도이다. 구체적으로 도 23은 제1/제2 모델 및 제1/제2 모델과 관련된 모든 동작을 해제하기 위한 절차는 나타낸다.
- [279] S2310에서, 송신자(Sender)는 수신자(Receiver)에 제1/제2 모델의 해제(release)를 요청한다. 예를 들어, 송신자(Sender)는 수신자(Receiver)에 primary-secondary model release request 메시지를 전송할 수 있다.

- [280] S2320에서, 수신자(Receiver)는 송신자(Sender)에 제1/제2 모델의 해제(release)를 확인하는 메시지를 전송한다. 예를 들어, 수신자는 송신자에 primary-secondary model release confirm 메시지를 전송할 수 있다.
- [281] 도 21에서 설명한 바와 같이 시그널링의 방향은 도 23과 반대로 수행될 수도 있다. 예를 들어, 수신자의 모델 해제 요청에 기초하여 송신자가 모델 해제 확인 메시지를 수신자에 전송할 수 있다.
- [282] State of primary and secondary model
- [283] 이하에서는 각 모델들(1차 모델(primary model), 대리 모델(surrogate model), 2차 모델(secondary model))의 각 상태(state)와 상태 천이(state transition)에 대하여 구체적으로 설명한다.
- [284] AI/ML 모델의 상태는 IDLE 상태, PREPARE 상태 또는 ACTIVE 상태일 수 있다.
- [285] 송신자 1차 모델 f_θ 과 수신자 1차 모델 g_ψ 은 태스크 동작을 시작하면 ACTIVE 상태로 이동한다. 이 때, 송신자/수신자의 2차 모델은 IDLE 상태이다. IDLE 상태인 AI/ML 모델은 아무런 동작을 수행하지 않는다.
- [286] 이하에서 도 24 내지 도 26을 참조하여 각 모델들의 상태 천이(state transition)를 구체적으로 설명한다.
- [287] 도 24는 본 명세서의 실시예에 따른 2차 모델(secondary model)의 상태 천이(state transition)를 나타낸다. 도 25는 본 명세서의 실시예에 따른 대리 모델(surrogate model)의 상태 천이(state transition)를 나타낸다. 도 26은 본 명세서의 실시예에 따른 1차 모델(primary model)의 상태 천이(state transition)를 나타낸다.
- [288] State transition from IDLE to PREPARE for primary and secondary model
- [289] 도 24 및 도 25를 참조하면, 1차 모델의 성능이 PRETH보다 나빠지면(또는 손실이 PRETH보다 커진 경우, $\text{Loss} > \text{PRETH}$), 송신자의 2차 모델 u_ϕ 과 수신자의 2차 모델 v_ξ 과 대리 모델 u_ϕ , v_ξ 은 IDLE 상태에서 PREPARE 상태로 이동한다. 이 조건이 SEC_PRE이다.
- [290] 반대로, 1차 모델의 성능이 PRETH보다 좋아지면(또는 손실이 PRETH보다 같거나 작아진 경우, $\text{Loss} \geq \text{PRETH}$), 송신자의 2차 모델 u_ϕ 과 수신자의 2차 모델 v_ξ 과 대리 모델 u_ϕ , v_ξ 은 PREPARE 상태에서 IDLE 상태로 이동한다. 이 조건이 SEC_IDL이다.
- [291] 2차 모델은 PREPARE 상태에서 현재 주어진 task에 대한 online learning을 수행한다(도 24). 동시에 1차 모델의 대리 모델은 1차 모델로부터 지식 증류 학습 혹은 전이 학습을 수행한다(도 25).
- [292] State transition from PREPARE to ACTIVE for secondary model
- [293] 도 24를 참조하면, 송신자의 2차 모델 및 수신자의 2차 모델은 독립적으로 또는 함께 ACTIVE 상태로 이동할 수 있다(도 26에서 각 2차 모델에 대응되는 1차 모델은 PREPARE 상태로 이동).

[294] 도 26을 참조하면, 송신자의 1차 모델 및 수신자의 1차 모델은 독립적으로 또는 함께 ACTIVE 상태로 이동할 수 있다(도 24에서 각 1차 모델에 대응되는 2차 모델은 IDLE 상태로 이동).

[295] 2차 모델은 ACTIVE 상태가 되면 실제 태스크 처리를 수행한다. 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 이동하는 조건(1차 모델이 ACTIVE 상태에서 PREPARE 상태로 이동하는 조건)은 SEC_ACT이다.

[296] 송신자의 2차 모델의 SEC_ACT 조건은 다음 수학적 식 3과 같이 표현될 수 있다.

[297] [수식 3]

$$E_{p(x,y)}[\mathcal{L}(y, g_\psi \circ h \circ u_{\phi'}(x))] + \Delta < E_{p(x,y)}[\mathcal{L}(y, g_\psi \circ h \circ f_\theta(x))] \quad \text{and}$$

$$E_{p(\phi, \phi')}[\|\phi - \phi'\|_2] > M_{diff}$$

[298] 수학적 식 3에서

Δ

는 SEC_ACT 조건 성립 여부 판단에 필요한 오프셋(offset)이다.

$$g_\psi \circ h \circ u_{\phi'}(x)$$

는 훈련 데이터

x

를 송신자의 2차 모델

$u_{\phi'}$

를 거쳐 채널

h

를 통과하여 수신자의 1차 모델

g_ψ

로 처리하는 합성 함수이다. 이 함수의 출력은 훈련 데이터

y

와 손실 함수로 비교한다. 연산자

$E_{p(x,y)}[\cdot]$

는 훈련 데이터 셋

(x, y)

에 대한 확률 분포

$p(x, y)$

에 대한 평균을 구한다. 손실 함수를 훈련 데이터 셋 분포에 대해 평균을 구한 것이 최종 평균 손실이다. 상술한 설명과 마찬가지로

$$g_\psi \circ h \circ f_\theta(x)$$

는 훈련 데이터

x

를 송신자의 1차 모델

$$f_{\theta}$$

를 거쳐 채널

$$h$$

을 통과하여 수신자의 1차 모델

$$g_{\psi}$$

로 처리하는 합성 함수이다.

[299] 수학식 3에서

$$E_{p(\phi, \phi')} [\|\phi - \phi'\|_2]$$

은 송신자의 2차 모델

$$u_{\phi'}$$

와 대리 모델

$$u_{\phi}$$

의 매개 변수의 L2 Norm(Euclidean Distance)을 평균한 값이다. 이 값이

$$M_{diff}$$

이상인 경우(또는

$$M_{diff}$$

보다 크거나 같은 경우) 2차 모델과 대리 모델은 서로 상이한 모델로 판별될 수 있다.

$$\Delta$$

는 손실 오프셋으로서 손실 값의 변동에 의한 민감도를 감소하는 역할을 한다.

[300] 수신자의 2차 모델의 SEC_ACT 조건은 다음 수학식 4와 같이 표현될 수 있다.

[301] [수식4]

$$E_{p(x,y)} [\mathcal{L}(y, v_{\xi'} \circ h \circ f_{\theta}(x))] + \Delta < E_{p(x,y)} [\mathcal{L}(y, g_{\psi} \circ h \circ f_{\theta}(x))] \quad \text{and}$$

$$E_{p(\xi, \xi')} [\|\xi - \xi'\|_2] > M_{diff}$$

[302] 수학식 4에서

$$v_{\xi} \circ h \circ f_{\theta}(x)$$

는 훈련 데이터

$$x$$

를 송신자의 1차 모델

$$f_{\theta}$$

를 거쳐 채널

$$h$$

를 통과하여 수신자의 2차 모델

v_{ξ}

로 처리하는 합성 함수 이다. 이 함수의 출력은 훈련 데이터

 y

와 손실 함수로 비교한다. 연산자

 $E_{p(x,y)}[.]$

는 훈련 데이터 셋

 (x,y)

에 대한 확률 분포

 $p(x,y)$

에 대한 평균을 구한다. 손실 함수를 훈련 데이터 셋 분포에 대해 평균을 구한 것이 최종 평균 손실이다. 상술한 설명과 마찬가지로

 $g_{\psi} \circ h \circ f_{\theta}(x)$

는 훈련 데이터

 x

를 송신자의 1차 모델

 f_{θ}

를 거쳐 채널

 h

를 통과하여 수신자의 1차 모델

 g_{ψ}

로 처리하는 합성 함수이다.

[303] 수학식 4에서

 $E_{p(\xi,\xi')}[\|\xi - \xi'\|_2]$

은 2차 모델

 $v_{\xi'}$

와 대리 모델

 v_{ξ}

의 매개 변수의 L2 Norm(Euclidean Distance)을 평균한 값이다. 이 값이

 M_{diff}

이상인 경우 2차 모델과 대리 모델은 서로 상이한 모델로 판별될 수 있다. 송신자와 마찬가지로

 Δ

는 손실 오프셋으로서 손실 값의 변동에 의한 민감도를 감소하는 역할을 한다.

[304] State transition from ACTIVE to PREPARE for primary model

[305] 도 26을 참조하면, 상술한 SEC_ACT 조건에 기초하여 송신자/수신자의 1차 모델은 ACTIVE 상태에서 PREPARE상태로 이동한다.

[306] State transition from PREPARE to ACTIVE for primary model

[307] 도 26을 참조하면, 송신자/수신자의 1차 모델은 독립적으로 또는 함께 PREPARE 상태에서 ACTIVE 상태로 이동할 수 있다. 이 때, 각 1차 모델에 대응되는 2차 모델은 ACTIVE 상태에서 IDLE 상태로 이동한다(도 24). 해당 상태 천이와 관련된 조건은 PRI_ACT이다.

[308] 송신자의 PRI_ACT 조건은 다음 수학적식 5와 같이 표현될 수 있다.

[309] [수식5]

$$E_{p(x,y)}[\mathcal{L}(y, g_{\psi} \circ h \circ u_{\phi}(x))] > E_{p(x,y)}[\mathcal{L}(y, g_{\psi} \circ h \circ f_{\theta}(x))] + \Delta \quad \text{and}$$

$$E_{p(\phi, \phi')}[\|\phi - \phi'\|_2] < M_{diff}$$

[310] 수학적식 5에서

Δ

는 측정 값의 큰 변동폭에 의한 잦은 상태 이동을 방지하기 위한 오프셋이다. 송신자의 PRI_ACT 조건은 송신자의 SEC_ACT 조건(수학적식 3)과 반대로 정의될 수 있다. 수학적식 5에 대한 설명은 수학적식 3과 대응되므로 중복된 설명을 생략한다.

[311] 수신자의 PRI_ACT 조건은 다음 수학적식 6과 같이 표현될 수 있다.

[312] [수식6]

$$E_{p(x,y)}[\mathcal{L}(y, v_{\xi} \circ h \circ f_{\theta}(x))] > E_{p(x,y)}[\mathcal{L}(y, g_{\psi} \circ h \circ f_{\theta}(x))] + \Delta \quad \text{and}$$

$$E_{p(\xi, \xi')}[\|\xi - \xi'\|_2] < M_{diff}$$

[313] 수신자의 PRI_ACT 조건은 수신자의 SEC_ACT 조건(수학적식 4)과 반대로 정의될 수 있다. 수학적식 6에 대한 설명은 수학적식 4와 대응되므로 중복된 설명을 생략한다.

[314] 이하에서는 도 27 내지 도 34를 참조하여 송신자/수신자의 1차 모델(및/또는 2차 모델)의 온라인 학습(online learning) 및 상태 천이와 관련된 동작을 구체적으로 설명한다.

[315] Online learning procedure to PREPARE state of secondary model from IDLE for sender

[316] 도 27은 본 명세서의 실시예에 따른 송신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 이동하는 절차를 나타낸다.

[317] 도 27을 참조하면, S2710에서, 송신자(Sender)의 1차 모델은 ACTIVE 상태에 있고, 송신자(Sender)의 2차 모델은 IDLE 상태에 있다.

[318] 성능이 좋은 깊은 신경망인 1차 모델의 성능이 기대 수준보다 낮아지면(예: Loss < PRETH) 무선 통신 관련 태스크의 수행을 위한 모델을 1차 모델에서 설명 가능한 2차 모델로 교체/변경하기 위한 준비 동작이 수행된다. 즉, 1차 모델의 성

능에 기초하여 송신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 이동할 수 있다.

- [319] S2720에서, 송신자는 수신자에 SEC_PRE transition request 메시지를 전송한다. 성능 저하는 손실 함수 평가를 수신자만이 알 수 있기 때문에 수신자가 이러한 조건을 감시하여 SEC_PRE transition 요청 메시지를 전송한다.
- [320] S2730에서, 수신자는 송신자에 SEC_PRE transition confirm 메시지를 전송한다.
- [321] S2740에서, 송신자는 2차 모델의 상태 천이를 수행한다. 구체적으로 송신자의 2차 모델은 IDLE 상태에서 PREPARE 상태로 이동한다. 이 때, 수신자는 송신자가 2차/대리 모델들의 학습을 개시하기 전에 자신의 모델들의 매개 변수를 동결할 수 있다(Stop online learning).
- [322] S2750에서, 송신자는 각각 2차 모델에 대한 온라인 학습 및 대리 모델에 대한 지식 증류(혹은 전이 학습)을 수행한다. 구체적으로, 송신자는 해당 학습을 위한 훈련 데이터(Training data)를 수신자에 전송한다.
- [323] S2760에서, 수신자는 상기 온라인 학습에 대한 손실을 송신자에게 피드백 한다.
- [324] Online learning procedure to ACTIVE state of secondary model from PREPARE for sender
- [325] 도 28은 본 명세서의 실시예에 따른 송신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 이동하는 절차를 나타낸다.
- [326] 도 28을 참조하면, S2810에서, 송신자(Sender)의 2차 모델은 PREPARE 상태에 있다. 도면에 도시된 것은 아니나 송신자의 1차 모델은 ACTIVE 상태에 있다. 이 때, 수신자는 1차 모델(및/또는 2차 모델)의 온라인 학습을 정지한다.
- [327] S2820에서, 송신자는 2차 모델의 PREPARE 상태에서 SEC_ACT 조건이 충족되는 지 여부를 결정한다. SEC_ACT 조건이 충족되는 것에 기초하여, 송신자는 수신자에게 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 요청 메시지를 전송한다. 예를 들어, 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 요청 메시지는 모델 변경 여부가 TRUE로 설정된 정보를 포함하는 Request SEC_ACT transition for sender 메시지일 수 있다. 상기 Request SEC_ACT transition for sender 메시지는 송신자의 대리 모델과 2차 모델의 매개 변수를 포함할 수 있다.
- [328] S2830에서, 수신자는 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 확인 메시지(예: Confirm SEC_ACT transition for sender 메시지)를 전송한다.
- [329] S2840에서, 송신자는 2차 모델을 PREPARE 상태에서 ACTIVE 상태로 천이한다. 송신자는 1차 모델을 ACTIVE 상태에서 PREPARE상태로 천이한다. 수신자는 1차 모델(및/또는 2차 모델)의 온라인 학습을 재개한다.
- [330] Online learning procedure to IDLE state of secondary model from PREPARE for sender

- [331] 도 29는 본 명세서의 실시예에 따른 송신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [332] 도 29를 참조하면, S2910에서, 송신자(Sender)의 2차 모델은 PREPARE 상태에 있다. 도면에 도시된 것은 아니나 송신자의 1차 모델은 ACTIVE 상태에 있다. 이때, 수신자는 1차 모델(및/또는 2차 모델)의 온라인 학습을 정지한다.
- [333] S2920에서, 송신자는 2차 모델의 PREPARE 상태에서 SEC_ACT 조건이 충족되는 지 여부를 결정한다. SEC_ACT 조건이 충족되지 않는 것에 기초하여, 송신자는 수신자에게 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 요청 메시지를 전송한다. 예를 들어, 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 요청 메시지는 FALSE로 설정된 Request SEC_ACT transition for sender 메시지일 수 있다.
- [334] S2930에서, 수신자는 2차 모델의 SEC_ACT 조건에 기초한 상태 천이에 대한 확인 메시지(예: Confirm SEC_ACT transition for sender 메시지)를 전송한다.
- [335] S2940에서, 송신자는 2차 모델을 PREPARE 상태에서 IDLE 상태로 천이한다. 수신자는 1차 모델(및/또는 2차 모델)의 온라인 학습을 재개한다.
- [336] Online learning procedure to ACTIVE state of primary model from PREPARE for sender (or to IDLE state of secondary model from ACTIVE for sender)
- [337] 도 30은 본 명세서의 실시예에 따른 송신자의 2차 모델이 ACTIVE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [338] 도 30을 참조하면, S3010에서, 송신자(Sender)의 1차 모델은 PREPARE 상태에 있고, 송신자(Sender)의 2차 모델은 ACTIVE 상태에 있다.
- [339] S3020에서, 송신자는 1차 모델 및 2차 모델의 온라인 학습을 수행한다. 송신자는 해당 학습을 위한 훈련 데이터(Training data)를 수신자에 전송한다.
- [340] S3030에서, 수신자는 상기 온라인 학습에 대한 손실을 송신자에게 피드백 한다.
- [341] S3040에서, 송신자는 (2차 모델의 ACTIVE 상태에서) PRI_ACT 조건이 충족되는 지 여부를 평가한다. PRI_ACT 조건이 충족되는 것에 기초하여, 송신자는 수신자에게 1차 모델의 PRI_ACT 조건에 기초한 상태 천이에 대한 요청 메시지를 전송한다. 예를 들어, 1차 모델의 PRI_ACT 조건에 기초한 상태 천이에 대한 요청 메시지는 TRUE로 설정된 Request PRI_ACT transition for sender 메시지일 수 있다.
- [342] S3050에서, 수신자는 1차 모델의 PRI_ACT 조건에 기초한 상태 천이에 대한 확인 메시지(예: Confirm PRI_ACT transition for sender 메시지)를 전송한다.
- [343] S3060에서, 수신자는 1차 모델을 PREPARE 상태에서 ACTIVE 상태로 천이한다. 수신자는 2차 모델을 ACTIVE 상태에서 IDLE 상태로 천이한다.
- [344] Online learning procedure to PREPARE state of secondary model from IDLE for receiver

- [345] 도 31은 본 명세서의 실시예에 따른 수신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 이동하는 절차를 나타낸다.
- [346] 성능이 좋은 깊은 신경망인 1차 모델의 성능이 기대 수준 보다 낮아지면 설명 가능한 2차 모델로 교체하기 위한 준비 동작이 수행된다. 도 31에서 수신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 천이하는 동작과 관련된 절차들(S3110 내지 S3150)은 도 27에서 송신자의 2차 모델이 IDLE 상태에서 PREPARE 상태로 천이하는 동작과 관련된 절차들(S2710 내지 S2750)에 대응된다. 따라서, 중복된 설명을 생략한다.
- [347] Online learning procedure to ACTIVE state of secondary model from PREPARE for receiver
- [348] 도 32는 본 명세서의 실시예에 따른 수신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 이동하는 절차를 나타낸다.
- [349] 도 32에서 SET_ACT 조건이 충족되는 것에 기초하여 수신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 천이하는 동작과 관련된 절차들(S3210 내지 S3240)은 도 28에서 SET_ACT 조건이 충족되는 것에 기초하여 송신자의 2차 모델이 PREPARE 상태에서 ACTIVE 상태로 천이하는 동작과 관련된 절차들(S2810 내지 S2840)에 대응된다. 따라서, 중복된 설명을 생략한다.
- [350] Online learning procedure to IDLE state of secondary model from PREPARE for receiver
- [351] 도 33은 본 명세서의 실시예에 따른 수신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [352] 도 33에서 SET_ACT 조건이 충족되지 않는 것에 기초하여 수신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 천이하는 동작과 관련된 절차들(S3310 내지 S3340)은 도 29에서 SET_ACT 조건이 충족되지 않는 것에 기초하여 송신자의 2차 모델이 PREPARE 상태에서 IDLE 상태로 천이하는 동작과 관련된 절차들(S2910 내지 S2940)에 대응된다. 따라서, 중복된 설명을 생략한다.
- [353] Online learning procedure to ACTIVE state of primary model from PREPARE for receiver (or to IDLE state of secondary model from ACTIVE for receiver)
- [354] 도 34는 본 명세서의 실시예에 따른 수신자의 2차 모델이 ACTIVE 상태에서 IDLE 상태로 이동하는 절차를 나타낸다.
- [355] 도 34에서 PRI_ACT 조건이 충족되는 것에 기초하여 수신자의 1차 모델이 PERPARE 상태에서 ACTIVE 상태로 천이(2차 모델이 ACTIVE 상태에서 IDLE 상태로 천이)하는 동작과 관련된 절차들(S3410 내지 S3450)은 도 30에서 PRI_ACT 조건이 충족되는 것에 기초하여 송신자의 1차 모델이 PERPARE 상태에서 ACTIVE 상태로 천이(2차 모델이 ACTIVE 상태에서 IDLE 상태로 천이)하는 동작과 관련된 절차들(S3010 내지 S3060)에 대응된다. 따라서, 중복된 설명을 생략한다.
- [356] Priority of online learning procedure for sender and receiver task

- [357] 상술한 도 27 내지 도 34의 절차에 따른 동작에 있어서, 송신자는 수신자로 수신자는 송신자로 변경하여 해당 절차에 따른 동작이 수행될 수 있다. 일 예로, 도 27의 절차에 따른 동작에서 송신자에 의한 동작은 수신자에 의해 수행되고, 수신자에 의한 동작은 송신자에 의해 수행될 수 있다.
- [358] 상술한 도 27 내지 도 34의 절차에 따른 동작에 있어서, 수신자의 모델의 상태 천이와 관련된 동작 및 송신자의 모델의 상태 천이와 관련된 동작이 함께 수행될 수 있다. 이 경우, 수신자(또는 송신자)의 상태 천이와 관련된 과정/동작이 우선 순위를 가지고 수행될 수 있다.
- [359] 이하에서는 도 35를 참조하여 상술한 실시예가 무선 통신 관련 태스크(빔 추적 태스크)에 대한 적용된 경우 구체적인 동작들을 구체적으로 설명한다.
- [360] 도 35는 본 명세서의 실시예에 따라 수행되는 빔 추적 태스크를 나타낸다.
- [361] 구체적으로 도 35는 물리 계층 통신에서 기지국과 단말간에 수행되는 빔 추적 태스크를 나타낸다. 예측된 빔 인덱스의 분포를 알아내기 위해 AI/ML의 회귀 태스크가 활용될 수 있다. 빔 인덱스의 추적을 위해 기지국이 빔 기준 신호를 단말에 전송하는 것으로 가정한다. 이 때, 빔 추적은 i) 송신자 측에서 송신 빔을 결정하기 위한 송신 빔 추적 및 ii) 수신자 측에서 수신 빔을 결정하기 위한 수신 빔 추적으로 구분될 수 있다. 이하에서는 송신 빔 추적을 중심으로 설명한다.
- [362] 송신 빔을 결정하는 AI/ML의 1차 모델은 성능이 좋은 깊은 신경망에 기반하며, 2차 모델은 선형 회귀 모델에 기반할 수 있다. 기지국의 AI/ML 모델의 입력 특징 x 는 단말의 위치, 속도, 자세로 이루어진 단말의 특징 벡터이고 y 는 예측 빔 인덱스의 분포 정보가 될 수 있다. 이 온라인 학습에서 훈련 샘플은 다른 방법을 통해 단말/기지국이 서로 알고 있다고 가정한다.
- [363] 도 35의 (a)는 1차 모델에서 벡터 x 의 일부 요소에 대한 데이터 y 를 나타낸다. 1차 모델(깊은 신경망)은 큰 영역에 대해 비선형적인 입력 및 출력 관계를 보여준다(즉, 낮은 설명 가능성). 구체적으로, x 표시는 벡터 x 에 대한 데이터 y 의 실제 분포이고, 실선은 해당 데이터 y 에 대한 1차 모델의 예측을 나타낸다.

- [364] 채널의 지역성을 가질 수 있도록 온라인 학습이 이루어 진다면 도 35의 (a)에서 표현된 전역적인 모델(1차 모델)의 일부는 지역성 영역에 대해 도 35의 (b)와 같이 표현될 수 있다.
- [365] 도 35의 (b)는 채널 지역성이 성립하는 영역에서 단말의 시간에 따른 이동에 대한 예측 빔의 평균 대응을 보여준다.
- [366] 상술한 SEC_PRE 조건이 충족되는 것에 기초하여 기지국의 2차 모델 및 대리 모델은 IDLE 상태에서 PREPARE 상태로 이동한다. PREPARE 상태에서 2차 모델의 학습 및 대리 모델의 학습(1차 모델에 기초한 지식 증류)이 수행된다. 이 때, 각 모델의 학습이 수행되는 동시에 성능이 모니터링 된다. 도 35의 (b)에서 실제 분포는 실선(35B-1)이고 깊은 신경망 모델의 예측은 점선(35B-2)이다.
- [367] 여기서, i) 1차 모델의 성능이 2차 모델보다 크게 떨어지고, ii) 도 35의 (b)에서 실선(35B-1)과 점선(35B-2)의 차이만큼 2차 모델과 대리 모델간 매개 변수의 차이가 상술한

$$M_{diff}$$

값보다 크게 되어 SEC_ACT 조건이 충족될 수 있다. 구체적으로, i) 1차 모델의 손실이 2차 모델의 손실+

$$\Delta$$

보다 크고 (수학식 3의

$$E_{p(x,y)}[\mathcal{L}(y, g_{\psi} \circ h \circ u_{\phi}(x))] + \Delta < E_{p(x,y)}[\mathcal{L}(y, g_{\psi} \circ h \circ f_{\theta}(x))]$$

, ii) 2차 모델의 매개 변수와 대리 모델의 매개 변수간의 차이가

$$M_{diff}$$

값보다 크게 되어 SEC_ACT 조건이 충족될 수 있다.

- [368] SEC_ACT 조건이 충족되면 기지국은 빔 추적 태스크를 위해 활성화 된 1차 모델을 2차 모델로 변경한다. 구체적으로, 기지국의 1차 모델은 ACTIVE 상태에서 PREPARE상태로 천이하고, 기지국의 2차 모델은 PREPARE 상태에서 ACTIVE 상태로 천이한다. 2차 모델은 도 35의 (b)에서 실선(35B-1)에 대한 온라인 학습 및 추정을 수행한다. 2차 모델의 학습은 선형 모델의 학습에 기반하는 바, 훈련에 필요한 데이터 수에 대한 대략적인 상한이 존재한다. 따라서, 2차 모델의 학습은 데이터 수에 기초하여 예측되는 시간 내에 빠르게 수행될 수 있다.
- [369] 성능이 저하된 1차 모델(예: 깊은 신경망)은 백그라운드 훈련 데이터로 긴 시간 훈련을 수행하거나 해당 채널의 지역성에 맞는 깊은 신경망 모델을 다운로드 받아 성능을 복원할 수 있다. 만약 1차 모델의 성능이 복원되면(예를 들어, 수학식 5에 기초한 PRI_ACT 조건이 충족되면), 1차 모델은 PREPARE 상태에서 ACTIVE 상태로 천이하고, 2차 모델은 ACTIVE 상태에서 IDLE 상태로 천이한다. 이 때, 대리 모델도 2차 모델의 상태 천이에 따라 IDLE 상태로 천이할 수 있다. 상술한 모

텔의 상태 천이에 기초하여 1차 모델이 다시 송신 빔 추적 태스크를 수행할 수 있다.

- [370] 도 35에서 설명의 편의를 위해 송신 빔 추적 태스크를 기준으로 기술하였으나, 상술한 실시예는 수신 빔 추적 태스크에도 확장 적용될 수 있다. 나아가, 상술한 실시예는 빔 추적 캐스트 외의 물리 계층의 다른 태스크에도 적용될 수 있다.
- [371] 상술한 실시예에 따른 동작 및 효과를 요약하면 다음과 같다.
- [372] AI/ML 1차 모델인 성능이 좋은 신경망이 학습하지 못한 채널 혹은 통신 환경을 겪을 때 채널의 지역성을 활용하여 설명성이 높은 2차 모델의 학습이 수행되고 동일한 모델인 대리 모델에 1차 모델의 지식 증류 혹은 전이학습이 수행된다.
- [373] 1차 모델의 일시적 성능 저하 또는 복원만을 기초로 모델 교체가 수행되는 경우에는 빈번한 모델 교체 동작으로 인해 무선 통신의 latency가 더욱 증가할 수 있다. 본 실시예에 의하면, 단순히 1차 모델의 성능 저하/복원시 모델을 변경하는 것이 아니라 2차 모델과 대리 모델과의 비교를 통해 태스크를 위해 활성화 된 모델을 1차 모델(2차 모델)에서 2차 모델(1차 모델)로 교체한다. 따라서, 보다 신뢰성 있는 판단을 기반으로 모델 교체의 효율을 제고하면서 온라인 학습과 태스크를 연속적으로 수행할 수 있다.
- [374] 2차 모델이 활성화 되어 동작하는 동안 1차 모델은 백그라운드로 온라인 학습을 성공적으로 수행할 수 있다. 이 때 1차 모델의 성능이 원래대로 돌아 오면 다시 1차 모델이 태스크 동작을 2차 모델 대신 수행한다. 이러한 방식으로 물리 계층의 송수신 태스크의 성능이 최상을 유지할 수 있다.
- [375] 구현적인 측면에서 상술한 실시예들에 따른 동작(예: 시멘틱 통신을 지원하기 위한 동작)들은 상술한 도 1 내지 도 5의 장치(예: 도 2의 프로세서(202a, 202b))에 의해 처리될 수 있다.
- [376] 또한 상술한 실시예에 따른 상술한 실시예들에 따른 동작(예: 시멘틱 통신을 지원하기 위한 동작)들은 적어도 하나의 프로세서(예: 도 2의 프로세서(202a, 202b))를 구동하기 위한 명령어/프로그램(예: instruction, executable code)형태로 메모리(예: 도 2의 204a, 204b)에 저장될 수도 있다.
- [377] 이하 상술한 실시예들을 무선 기기(예: 도 2의 제1 무선 기기(200a), 제2 무선 기기(200b))의 동작 측면에서 도 36을 참조하여 구체적으로 설명한다. 이하 설명되는 방법들은 설명의 편의를 위하여 구분된 것일 뿐, 어느 한 방법의 일부 구성이 다른 방법의 일부 구성과 치환되거나, 상호 간에 결합되어 적용될 수 있음은 물론이다.
- [378] 도 36은 본 명세서의 실시예에 따라 제1 무선 기기에 의해 수행되는 방법을 설명하기 위한 흐름도이다.
- [379] 도 36을 참조하면, 본 명세서의 일 실시예에 따른 AI/ML (Artificial Intelligence/ Machine Learning) 모델을 지원하는 무선 통신 시스템에서 제1 무선 기기에 의해 수행되는 방법은 1차 모델의 손실과 관련된 값에 기초하여 대리 모델의 학습 및 2차 모델의 학습을 수행하는 단계(S3610) 및 대리 모델의 매개 변수와 2차 모델

의 매개 변수간 차이 값에 기초하여 활성화 된 모델을 1차 모델에서 2차 모델로 변경하는 단계(S3620)를 포함한다.

- [380] S3610에서, 제1 무선 기기는 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행한다.
- [381] 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델일 수 있다. 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다를 수 있다. 일 실시예에 의하면, 상기 무선 통신과 관련된 태스크(task)는 채널 상태 정보(Channel State Information)의 보고, 참조 신호의 측정(measurement of reference signal), 신호 변조, 빔 트래킹(beam tracking) 또는 빔 탐색(beam search) 중 적어도 하나와 관련될 수 있다. 다만, 상기 무선 통신과 관련된 태스크는 상기 열거된 태스크들에 한정되는 것은 아니다. 상기 무선 통신과 관련된 태스크는 무선 통신과 관련된 각 계층(예: physical, MAC 및/또는 RRC) 별로 수행되는 하나 이상의 동작들과 관련될 수 있다.
- [382] 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초할 수 있다. 일 실시예에 의하면, 상기 전이 학습은 지식 증류(knowledge distillation)와 관련될 수 있다. 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습일 수 있다.
- [383] 일 실시예에 의하면, 상기 대리 모델 및 상기 2차 모델의 학습은 PREPARE 상태에서 수행될 수 있다. 구체적으로, 상기 1차 모델의 성능과 관련된 값이 상기 제1 임계값보다 큰 것에 기초하여, 상기 대리 모델 및 상기 2차 모델의 동작과 관련된 상태는 IDLE 상태에서 PREPARE 상태로 천이될 수 있다. 상기 대리 모델의 학습 및 상기 2차 모델의 학습은 상기 PREPARE 상태에서 수행될 수 있다.
- [384] 일 실시예에 의하면, 상기 1차 모델은 깊은 신경망(Deep Neural Network, DNN) 모델을 포함할 수 있다. 상기 1차 모델이 상기 DNN 모델에 한정되는 것은 아니며 도 19의 설명 가능성(Explainability)이 낮은 모델들을 포함할 수 있다. 일 예로, 상기 1차 모델은 서포트 벡터 머신(support vector machine, SVM)에 기반하는 모델을 포함할 수 있다. 일 예로, 상기 1차 모델은 랜덤 포레스트(Random Forest)에 기초한 모델을 포함할 수 있다. 일 예로, 상기 1차 모델은 선형 모델(linear model)보다 성능이 좋고 설명 가능성은 낮은 모델을 포함할 수 있다. 상기 1차 모델은 위에서 열거된 모델들로 제한되는 것은 아니다. 상기 1차 모델은 후술하는 2차 모델보다 (성능이 좋지만) 설명 가능성(Explainability)은 낮은 모델을 지칭하는 것일 수 있다.
- [385] 후술하는 바와 같이 상기 무선 통신 관련 태스크를 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하기 전에 상기 2차 모델을 위한 학습이 완료될 필요가 있다. 상기 2차 모델은 상기 1차 모델에 비해 설명 가능성이 높다는 점에서 학습에 요구되는 데이터의 양이 한정적일 수 있다. 일 실시예에 의하면, 상기 2차 모델의 학습을 위한 훈련 데이터(Training data)의 크기는 상기 1차 모델의 학습을 위한 훈련 데이터의 크기보다 작을 수 있다.

- [386] 일 실시예에 의하면, 상기 2차 모델의 학습을 위한 훈련 데이터는 i) 입력과 ii) 상기 입력에 대응되는 출력을 포함할 수 있다. 상기 입력은 상기 제1 무선 기기의 위치, 속도 또는 자세 중 적어도 하나와 관련된 벡터를 포함할 수 있다. 상기 출력은 상기 벡터에 기초하여 결정된 정보를 포함할 수 있다. 상기 벡터에 기초하여 결정된 정보는 상기 제1 무선 기기의 전송 또는 수신과 관련된 빔의 인덱스를 포함할 수 있다. 일 예로, 상기 벡터에 기초하여 결정된 정보는 상기 제1 무선 기기의 수신 빔(spatial domain receive filter) 또는 상기 제1 무선 기기의 수신 빔(spatial domain transmission filter)와 관련될 수 있다. 이 경우, 상기 무선 통신과 관련된 태스크는 상기 제1 무선 기기의 빔 추적과 관련된 태스크일 수 있다.
- [387] S3620에서, 제1 무선 기기는 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경한다.
- [388] 일 실시예에 의하면, 상기 2차 모델은 선형 모델(linear model)을 포함할 수 있다. 상기 1차 모델이 상기 선형 모델에 한정되는 것은 아니며 도 19의 설명 가능성(Explainability)이 높은 모델들을 포함할 수 있다. 일 예로, 상기 2차 모델은 결정 트리(decision tree)에 기초한 모델을 포함할 수 있다. 상기 2차 모델은 위에서 열거된 모델들로 제한되는 것은 아니다. 상기 2차 모델은 상기 1차 모델보다 설명 가능성(explainability)은 높은 모델을 지칭하는 것일 수 있다. 일 예로, 상기 2차 모델은 깊은 신경망을 제외한 선형 모델(linear model), 커널 모델(kernel model), 그래프 모델(graph model) 및/또는 확률 모델(probabilistic model)을 포함할 수 있다.
- [389] 일 실시예에 의하면, 상기 1차 모델의 설명 가능성(explainability)은 상기 2차 모델의 설명 가능성보다 낮을 수 있다.
- [390] 일 실시예에 의하면, 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값은 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 유클리드 거리(Euclidean Distance)에 기초하여 계산된 평균 값일 수 있다. 상기 평균 값은 수학식 3(또는 수학식 4)에 기반할 수 있다. 상기 유클리드 거리는 수학식 3(또는 수학식 4)의 L2 Norm을 의미할 수 있다.
- [391] 일 실시예에 의하면, 상술한 활성화 된 모델의 변경(1차 -> 2차)은 SEC_ACT 조건이 충족되는 것에 기초하여 수행될 수 있다. 이하 구체적으로 설명한다.
- [392] i) 상기 1차 모델의 손실과 관련된 값이 상기 2차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 큰 것에 기초하여:
- [393] 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이 상기 1차 모델에서 상기 2차 모델로 변경될 수 있다. 본 실시예는 수학식 3(또는 수학식 4)에 기반할 수 있다.
- [394] 상기 SEC_ACT 조건이 충족되는 것에 기초하여 상기 1차/2차 모델의 동작과 관련된 상태는 다음과 같이 천이될 수 있다.

- [395] 상기 1차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 PREPARE 상태로 천이될 수 있다. 상기 2차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이될 수 있다.
- [396] 설명 가능성이 낮은 상기 1차 모델의 성능이 상기 2차 모델의 성능보다 좋다는 점에서, 활성화 된 모델이 변경된 후에도 상기 1차 모델의 성능이 복구된다면 다시 해당 모델에 기초하여 무선 통신 관련 태스크를 수행할 필요가 있다. 활성화 된 모델을 상기 2차 모델에서 상기 1차 모델로 변경하는 동작은 PRI_ACT 조건이 충족되는 것에 기초하여 수행될 수 있다. 이하 구체적으로 설명한다.
- [397] 일 실시예에 의하면, i) 상기 2차 모델의 손실과 관련된 값이 상기 1차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 작은 것에 기초하여:
- [398] 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이 상기 2차 모델에서 상기 1차 모델로 변경될 수 있다. 본 실시예는 수학적식 5(또는 수학적식 6)에 기반할 수 있다.
- [399] 상기 PRI_ACT 조건이 충족되는 것에 기초하여 상기 1차/2차 모델의 동작과 관련된 상태는 다음과 같이 천이될 수 있다.
- [400] 상기 1차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이될 수 있다. 상기 2차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 IDLE 상태로 천이될 수 있다.
- [401] 일 실시예에 의하면, 상기 방법은 제2 무선 기기에 상기 활성화 된 모델의 변경과 관련된 메시지를 전송하는 단계를 더 포함할 수 있다. 상기 메시지는 상기 대리 모델의 매개변수 및 상기 2차 모델의 매개변수를 포함할 수 있다.
- [402] 일 실시예에 의하면, 상기 제1 무선 기기는 송신자(Sender)(예: 도 20의 2010, 도 21 내지 도 23, 도 27 내지 도 34), 수신자(Receiver) (예: 도 20의 2020, 도 21 내지 도 23, 도 27 내지 도 34), 단말(User Equipment, UE) 또는 기지국(Base station)일 수 있다. 상기 제2 무선 기기는 수신자(Receiver)(예: 도 20의 2020, 도 21 내지 도 23, 도 27 내지 도 34), 송신자(Sender)(예: 도 20의 2010, 도 21 내지 도 23, 도 27 내지 도 34), 기지국(Base station) 또는 단말(User Equipment, UE)일 수 있다.
- [403] 일 실시예에 의하면, 상기 방법은 도 21에 기초한 제1 무선 기기의 성능과 관련된 시그널링 절차(S2110 및/또는 S2120)를 더 포함할 수 있다. 상기 S2110 및/또는 S2120은 S3610 이전에 수행될 수 있다. 일 예로, (제2 무선 기기의 요청에 기초하여) 제1 무선 기기는 제2 무선 기기에 제1 무선 기기의 성능과 관련된 정보를 전송할 수 있다. 상기 제1 무선 기기의 성능과 관련된 정보는 제1 무선 기기에 의해 지원되는 1차 모델(primary model) 및/또는 2차 모델(secondary model)과 관련된 정보를 포함할 수 있다(예: 표 4).
- [404] 일 실시예에 의하면, 상기 방법은 도 22에 기초한 모델 셋업과 관련된 시그널링 절차(S2210 및/또는 S2220)를 더 포함할 수 있다. 상기 S2210 및/또는 S2220은

S3610 이전에 수행될 수 있다. 일 예로, (제1 무선 기기의 요청에 기초하여) 제1 무선 기기는 제2 무선 기기로부터 제1 무선 기기의 모델 셋업과 관련된 확인 메시지를 수신할 수 있다.

[405] 일 실시예에 의하면, 상기 방법은 도 23에 기초한 모델 해제(release)와 관련된 시그널링 절차(S2310 및/또는 S2320)를 더 포함할 수 있다. 상기 S2310 및/또는 S2320은 S3610 이전에 수행될 수 있다. 일 예로, (제1 무선 기기의 요청에 기초하여) 제1 무선 기기는 제2 무선 기기로부터 제1 무선 기기의 모델 셋업과 관련된 확인 메시지를 수신할 수 있다.

[406] 상술한 S3610 및 S3620, 도 21 내지 도 23에 기초한 동작은 도 1 내지 도 5의 장치에 의해 구현될 수 있다. 예를 들어, 단말(예: 도 2의 200a)은 상술한 S3610 및 S3620, 도 21 내지 도 23에 기초한 동작을 수행하도록 하나 이상의 트랜시버(206a) 및/또는 하나 이상의 메모리(204a)를 제어할 수 있다. 예를 들어, 기지국(예: 도 2의 200b)은 상술한 S3610 및 S3620, 도 21 내지 도 23에 기초한 동작을 수행하도록 하나 이상의 트랜시버(206b) 및/또는 하나 이상의 메모리(204b)를 제어할 수 있다.

[407] 여기서, 본 명세서의 무선 기기(200a, 200b)에서 구현되는 무선 통신 기술은 LTE, NR 및 6G뿐만 아니라 저전력 통신을 위한 Narrowband Internet of Things를 포함할 수 있다. 이때, 예를 들어 NB-IoT 기술은 LPWAN(Low Power Wide Area Network) 기술의 일례일 수 있고, LTE Cat NB1 및/또는 LTE Cat NB2 등의 규격으로 구현될 수 있으며, 상술한 명칭에 한정되는 것은 아니다. 추가적으로 또는 대체적으로, 본 명세서의 무선 기기(XXX, YYY)에서 구현되는 무선 통신 기술은 LTE-M 기술을 기반으로 통신을 수행할 수 있다. 이때, 일 예로, LTE-M 기술은 LPWAN 기술의 일례일 수 있고, eMTC(enhanced Machine Type Communication) 등의 다양한 명칭으로 불릴 수 있다. 예를 들어, LTE-M 기술은 1) LTE CAT 0, 2) LTE Cat M1, 3) LTE Cat M2, 4) LTE non-BL(non-Bandwidth Limited), 5) LTE-MTC, 6) LTE Machine Type Communication, 및/또는 7) LTE M 등의 다양한 규격 중 적어도 어느 하나로 구현될 수 있으며 상술한 명칭에 한정되는 것은 아니다. 추가적으로 또는 대체적으로, 본 명세서의 무선 기기(XXX, YYY)에서 구현되는 무선 통신 기술은 저전력 통신을 고려한 지그비(ZigBee), 블루투스(Bluetooth) 및 저전력 광역 통신망(Low Power Wide Area Network, LPWAN) 중 적어도 어느 하나를 포함할 수 있으며, 상술한 명칭에 한정되는 것은 아니다. 일 예로 ZigBee 기술은 IEEE 802.15.4 등의 다양한 규격을 기반으로 소형/저-파워 디지털 통신에 관련된 PAN(personal area networks)을 생성할 수 있으며, 다양한 명칭으로 불릴 수 있다.

[408] 이상에서 설명된 실시 예들은 본 명세서의 구성요소들과 특징들이 소정 형태로 결합된 것들이다. 각 구성요소 또는 특징은 별도의 명시적 언급이 없는 한 선택적인 것으로 고려되어야 한다. 각 구성요소 또는 특징은 다른 구성요소나 특징과 결합되지 않은 형태로 실시될 수 있다. 또한, 일부 구성요소들 및/또는 특징들을

결합하여 본 명세서의 실시 예를 구성하는 것도 가능하다. 본 명세서의 실시 예들에서 설명되는 동작들의 순서는 변경될 수 있다. 어느 실시예의 일부 구성이나 특징은 다른 실시 예에 포함될 수 있고, 또는 다른 실시예의 대응하는 구성 또는 특징과 교체될 수 있다. 특허청구범위에서 명시적인 인용 관계가 있지 않은 청구항들을 결합하여 실시 예를 구성하거나 출원 후의 보정에 의해 새로운 청구항으로 포함시킬 수 있음은 자명하다.

- [409] 본 명세서에 따른 실시 예는 다양한 수단, 예를 들어, 하드웨어, 펌웨어(firmware), 소프트웨어 또는 그것들의 결합 등에 의해 구현될 수 있다. 하드웨어에 의한 구현의 경우, 본 발명의 일 실시 예는 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), 프로세서, 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 등에 의해 구현될 수 있다.
- [410] 펌웨어나 소프트웨어에 의한 구현의 경우, 본 명세서의 일 실시 예는 이상에서 설명된 기능 또는 동작들을 수행하는 모듈, 절차, 함수 등의 형태로 구현될 수 있다. 소프트웨어 코드는 메모리에 저장되어 프로세서에 의해 구동될 수 있다. 상기 메모리는 상기 프로세서 내부 또는 외부에 위치하여, 이미 공지된 다양한 수단에 의해 상기 프로세서와 데이터를 주고 받을 수 있다.
- [411] 본 명세서는 본 명세서의 필수적 특징을 벗어나지 않는 범위에서 다른 특정한 형태로 구체화될 수 있음은 당업자에게 자명하다. 따라서, 상술한 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니 되고 예시적인 것으로 고려되어야 한다. 본 명세서의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 명세서의 등가적 범위 내에서의 모든 변경은 본 명세서의 범위에 포함된다.

청구범위

- [청구항 1] AI/ML (Artificial Intelligence/Machine Learning) 모델을 지원하는 무선 통신 시스템에서 제1 무선 기기에 의해 수행되는 방법에 있어서,
1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계,
상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이며, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르고,
상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초하며, 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이고; 및
상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계;
를 포함하는 방법.
- [청구항 2] 제1 항에 있어서,
상기 1차 모델은 깊은 신경망(Deep Neural Network, DNN) 모델을 포함하는 것을 특징으로 하는 방법.
- [청구항 3] 제1 항에 있어서,
상기 2차 모델은 깊은 신경망을 제외한 선형 모델(linear model), 커널 모델(kernel model), 그래프 모델(graph model) 및/또는 확률 모델(probabilistic model)을 포함하는 것을 특징으로 하는 방법.
- [청구항 4] 제1 항에 있어서,
상기 1차 모델의 설명 가능성(explainability)은 상기 2차 모델의 설명 가능성보다 낮은 것을 특징으로 하는 방법.
- [청구항 5] 제1 항에 있어서,
상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값은 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 유클리드 거리(Euclidean Distance)에 기초하여 계산된 평균 값인 것을 특징으로 하는 방법.
- [청구항 6] 제1 항에 있어서,
i) 상기 1차 모델의 손실과 관련된 값이 상기 2차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 큰 것에 기초하여:
상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이 상기 1차 모델에서 상기 2차 모델로 변경되는 것을 특징으로 하는 방법.

- [청구항 7] 제1 항에 있어서,
제2 무선 기기에 상기 활성화 된 모델의 변경과 관련된 메시지를 전송하는 단계를 더 포함하고,
상기 메시지는 상기 대리 모델의 매개변수 및 상기 2차 모델의 매개변수를 포함하는 것을 특징으로 하는 방법.
- [청구항 8] 제6 항에 있어서,
상기 1차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 PREPARE 상태로 천이되고, 상기 2차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이되는 것을 특징으로 하는 방법.
- [청구항 9] 제1 항에 있어서,
상기 1차 모델의 성능과 관련된 값이 상기 제1 임계값보다 큰 것에 기초하여, 상기 대리 모델 및 상기 2차 모델의 동작과 관련된 상태는 IDLE 상태에서 PREPARE 상태로 천이되는 것을 특징으로 하는 방법.
- [청구항 10] 제9 항에 있어서,
상기 대리 모델의 학습 및 상기 2차 모델의 학습은 상기 PREPARE 상태에서 수행되는 것을 특징으로 하는 방법.
- [청구항 11] 제1 항에 있어서,
상기 전이 학습은 지식 증류(knowledge distillation)와 관련된 것을 특징으로 하는 방법.
- [청구항 12] 제1 항에 있어서,
상기 무선 통신과 관련된 태스크(task)는 채널 상태 정보(Channel State Information)의 보고, 참조 신호의 측정(measurement of reference signal), 신호 변조, 빔 트래킹(beam tracking) 또는 빔 탐색(beam search) 중 적어도 하나와 관련된 것을 특징으로 하는 방법.
- [청구항 13] 제1 항에 있어서,
상기 2차 모델의 학습을 위한 훈련 데이터(Training data)의 크기는 상기 1차 모델의 학습을 위한 훈련 데이터의 크기보다 작은 것을 특징으로 하는 방법.
- [청구항 14] 제13 항에 있어서,
상기 2차 모델의 학습을 위한 훈련 데이터는 i) 입력과 ii) 상기 입력에 대응되는 출력을 포함하고,
상기 입력은 상기 제1 무선 기기의 위치, 속도 또는 자세 중 적어도 하나와 관련된 벡터를 포함하며,
상기 출력은 상기 벡터에 기초하여 결정된 정보를 포함하는 것을 특징으로 하는 방법.
- [청구항 15] 제14 항에 있어서,
상기 벡터에 기초하여 결정된 정보는 상기 제1 무선 기기의 전송 또는 수신과 관련된 빔의 인덱스를 포함하는 것을 특징으로 하는 방법.

- [청구항 16] 제1 항에 있어서,
 i) 상기 2차 모델의 손실과 관련된 값이 상기 1차 모델의 손실과 관련된 값에 오프셋 값을 더한 값보다 크고, ii) 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값이 제2 임계값보다 작은 것에 기초하여;
 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이 상기 2차 모델에서 상기 1차 모델로 변경되는 것을 특징으로 하는 방법.
- [청구항 17] 제16 항에 있어서,
 상기 1차 모델의 동작과 관련된 상태는 PREPARE 상태에서 ACTIVE 상태로 천이되고, 상기 2차 모델의 동작과 관련된 상태는 ACTIVE 상태에서 IDLE 상태로 천이되는 것을 특징으로 하는 방법.
- [청구항 18] 제1 항에 있어서,
 상기 제1 무선 기기는 송신자(Sender), 수신자(Receiver), 단말 또는 기지국인 것을 특징으로 하는 방법.
- [청구항 19] AI/ML (Artificial Intelligence/Machine Learning) 모델을 지원하는 무선 통신 시스템에서 동작하는 제1 무선 기기에 있어서,
 하나 이상의 송수신기;
 상기 하나 이상의 송수신기를 제어하는 하나 이상의 프로세서들; 및
 상기 하나 이상의 프로세서들에 동작 가능하게 접속 가능하고, 상기 하나 이상의 프로세서들에 의해 실행될 때 상기 하나 이상의 프로세서들이 동작들을 수행하도록 설정하는 지시(instruction)들을 저장하는 하나 이상의 메모리들을 포함하며,
 상기 동작들은,
 1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계,
 상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델이며, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르고,
 상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초하며, 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이고; 및
 상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화 된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계;를 포함하는 제1 무선 기기.
- [청구항 20] 하나 이상의 메모리들 및 상기 하나 이상의 메모리들과 기능적으로 연결되어 있는 하나 이상의 프로세서들을 포함하는 장치에 있어서,

상기 하나 이상의 메모리들은, 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들이 동작들을 수행하도록 설정하는 지시(instruction)들을 저장하며,

상기 동작들은,

1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계,

상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이며, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르고,

상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초하며, 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이고; 및

상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계;를 포함하는 장치.

[청구항 21]

하나 이상의 명령어를 저장하는 하나 이상의 비일시적(non-transitory) 컴퓨터 판독 가능 매체에 있어서,

상기 하나 이상의 명령어는, 하나 이상의 프로세서에 의해 실행되는 경우, 상기 하나 이상의 프로세서가 동작들을 수행하도록 설정하며,

상기 동작들은,

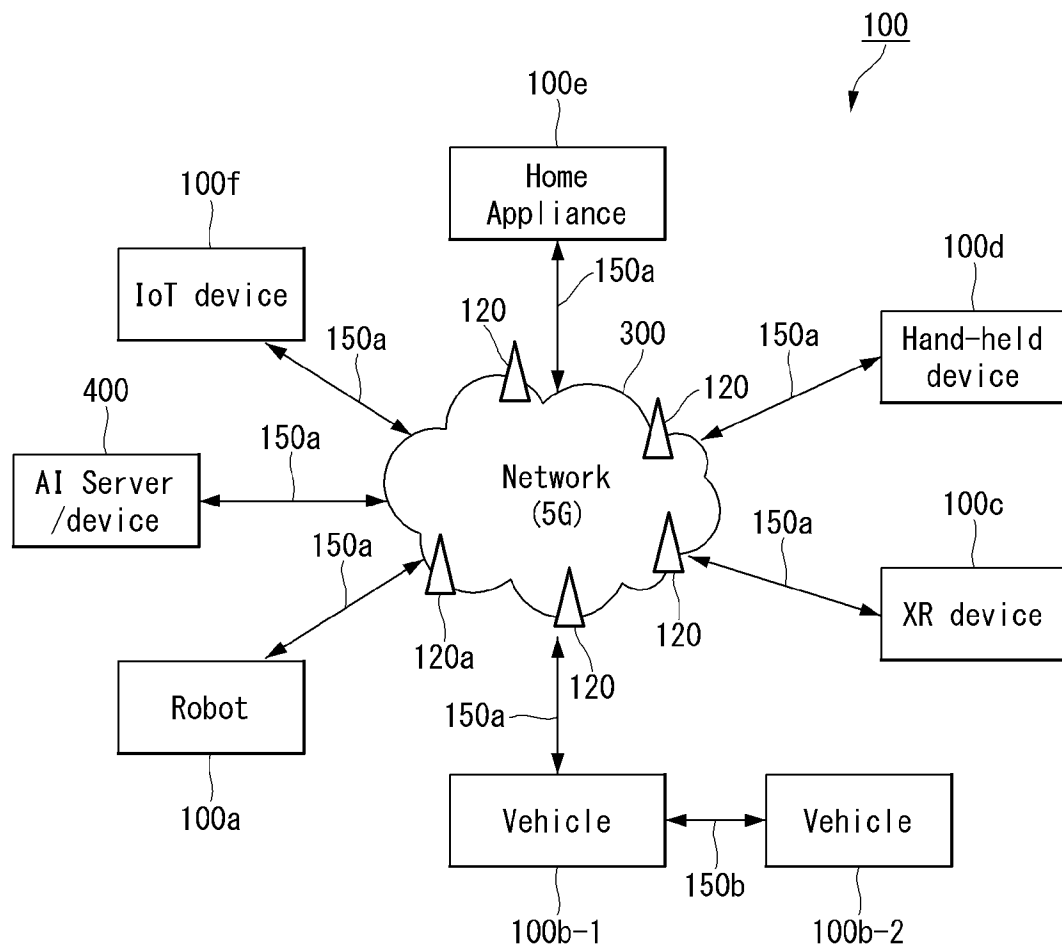
1차 모델의 손실과 관련된 값이 제1 임계값보다 큰 것에 기초하여 대리 모델(surrogate model)의 학습과 2차 모델의 학습을 수행하는 단계,

상기 1차 모델은 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델이며, 상기 대리 모델과 2차 모델은 모델 종류 및 구조가 동등하며 학습에 기초한 매개 변수가 서로 다르고,

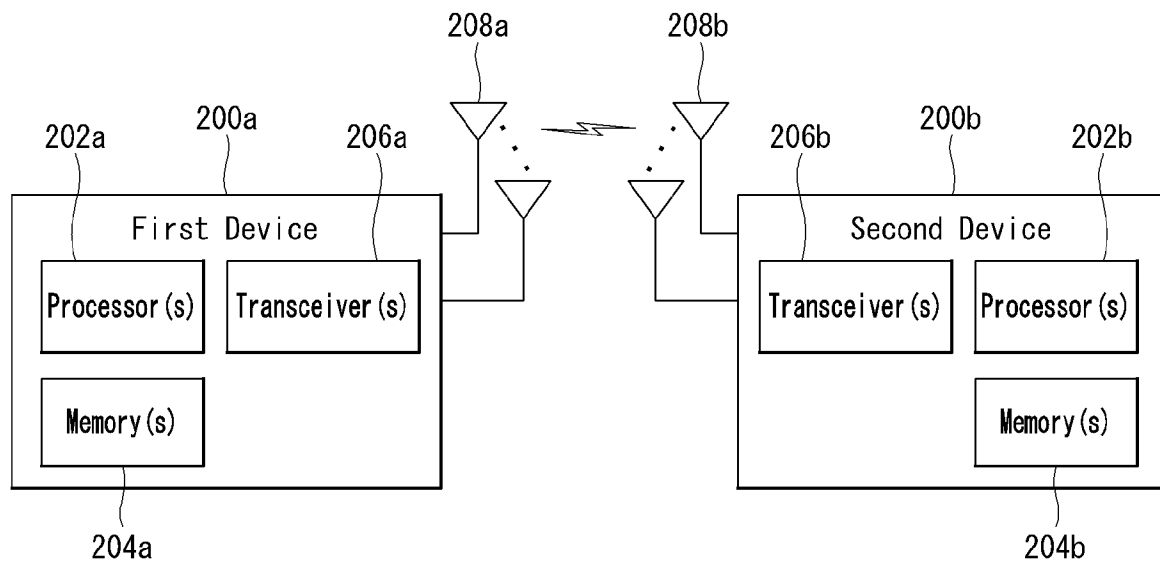
상기 대리 모델의 학습은 상기 1차 모델의 전이 학습(Transfer learning)에 기초하며, 상기 2차 모델의 학습은 현재 송수신하고 있는 무선 신호를 기반으로 하는 온라인 학습이고; 및

상기 대리 모델의 매개 변수와 상기 2차 모델의 매개 변수간의 차이를 나타내는 값에 기초하여, 상기 무선 통신과 관련된 태스크(task)의 수행을 위해 활성화된 모델을 상기 1차 모델에서 상기 2차 모델로 변경하는 단계;를 포함하는 비일시적(non-transitory) 컴퓨터 판독 가능 매체.

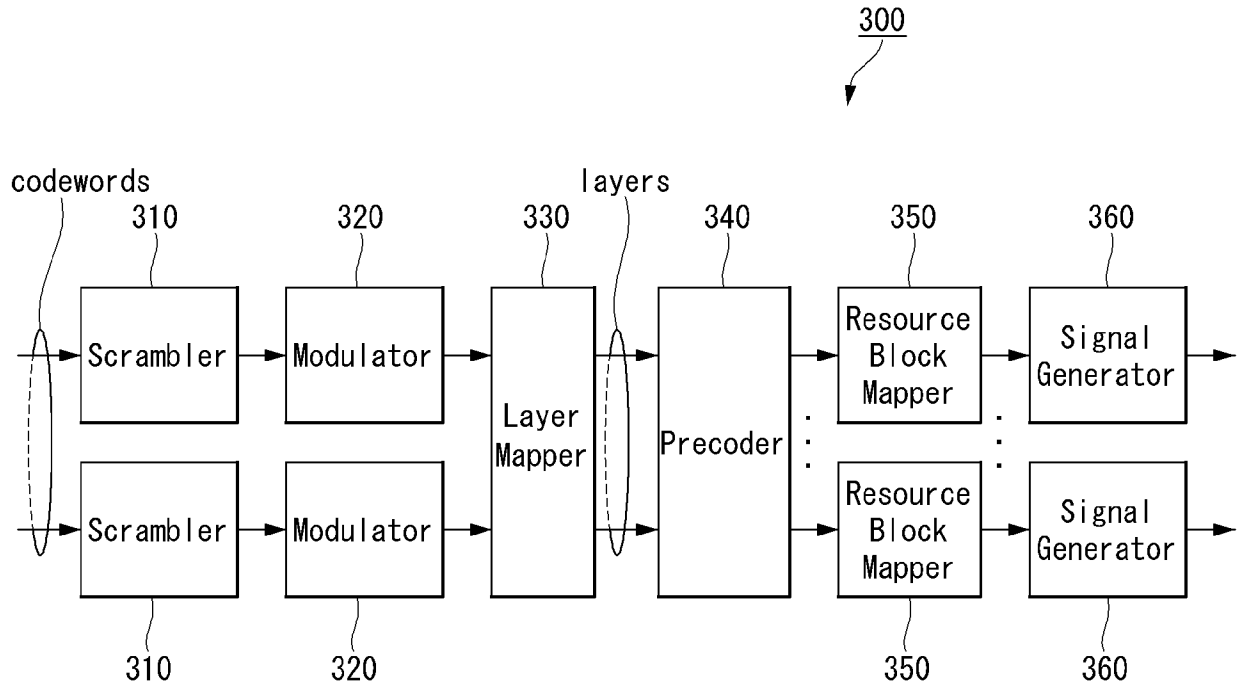
[도1]



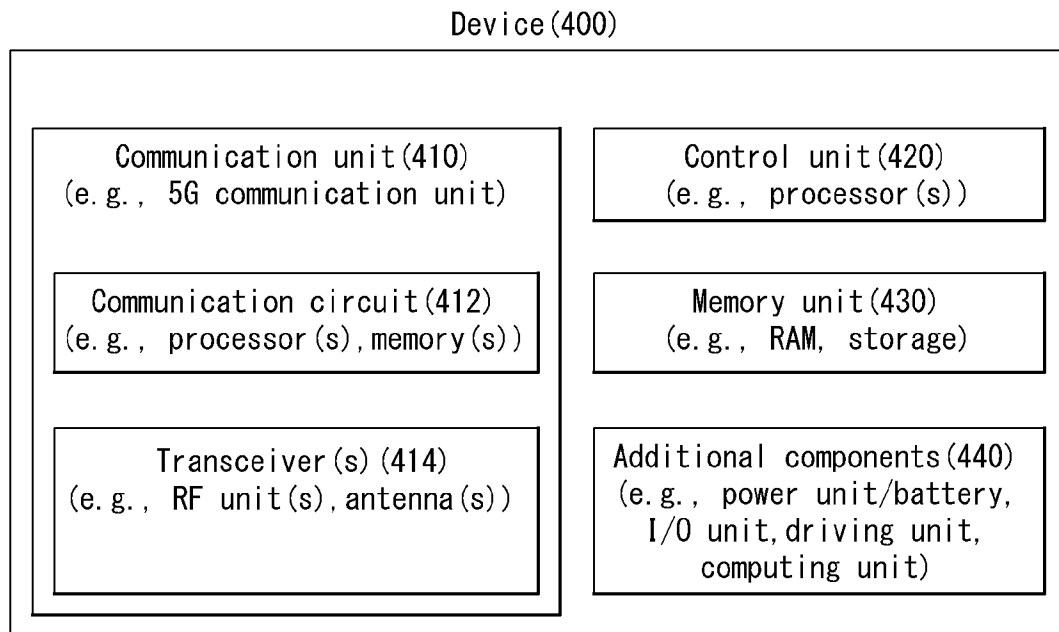
[도2]



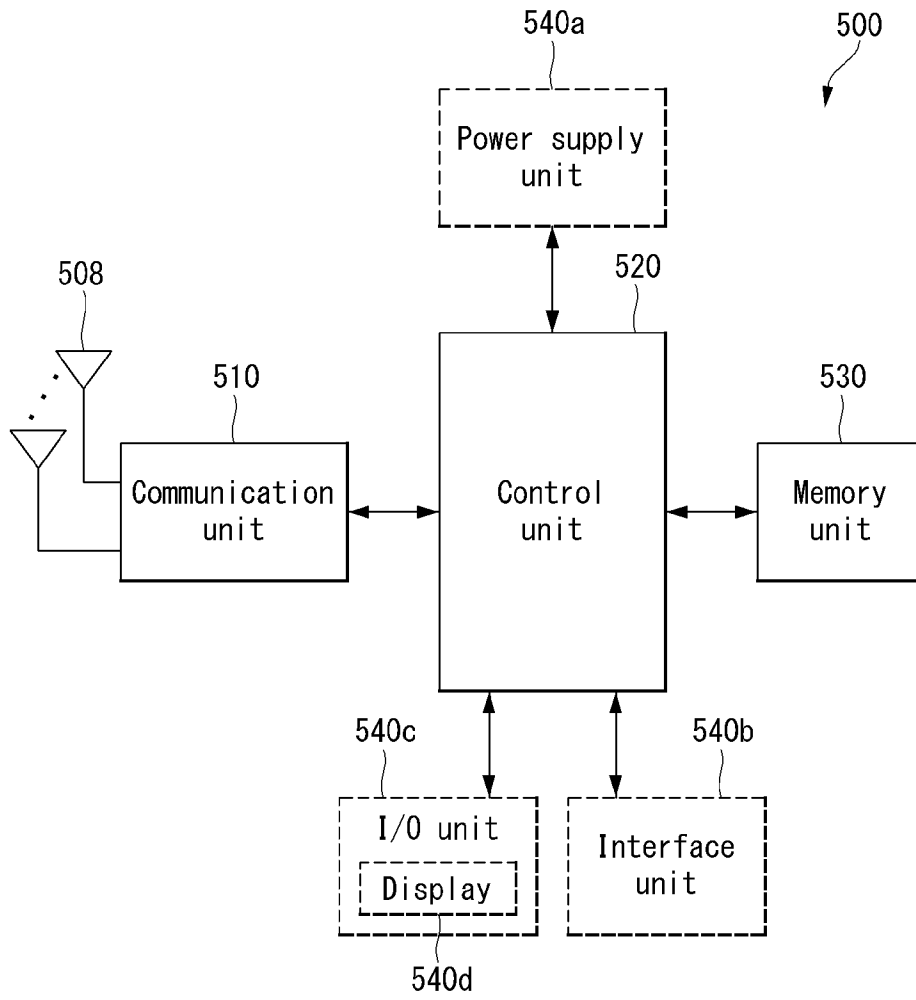
[도3]



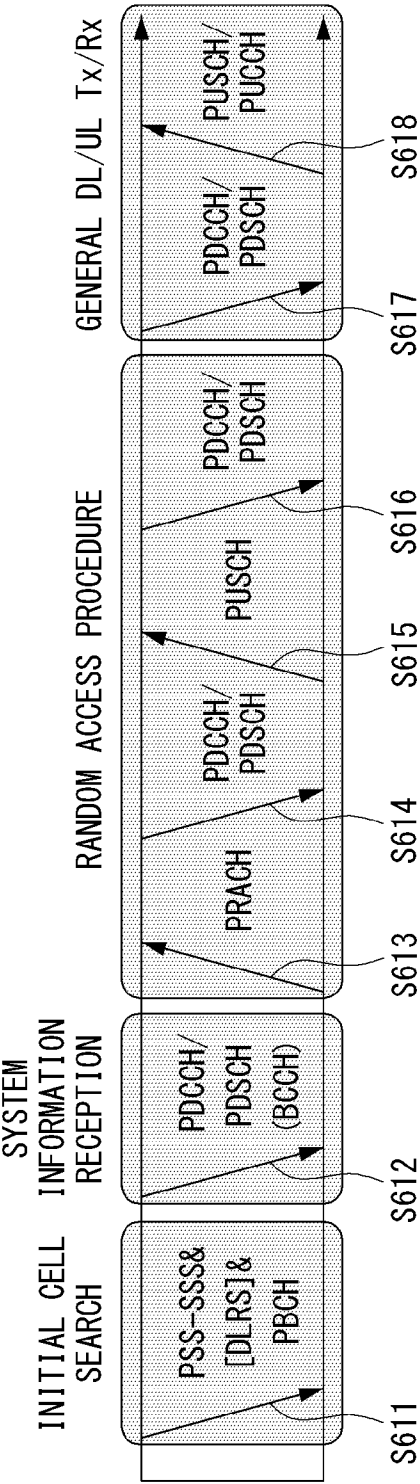
[도4]



[도5]

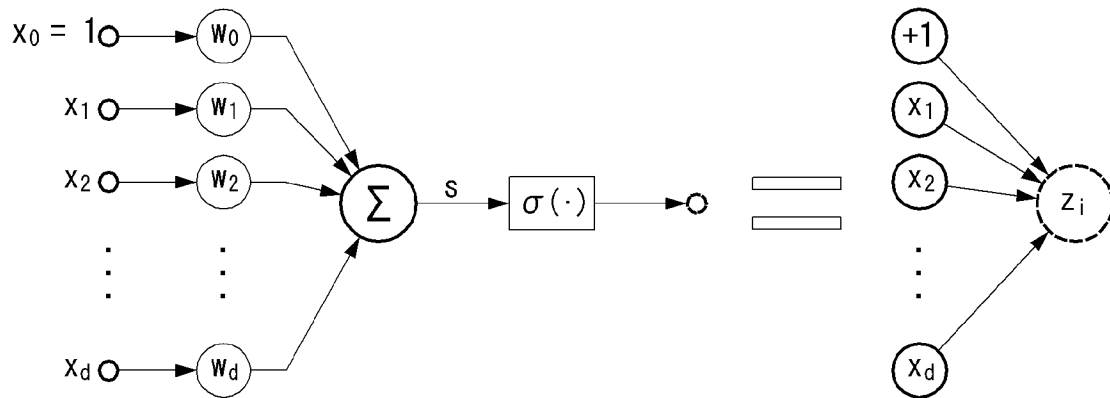


[도6]

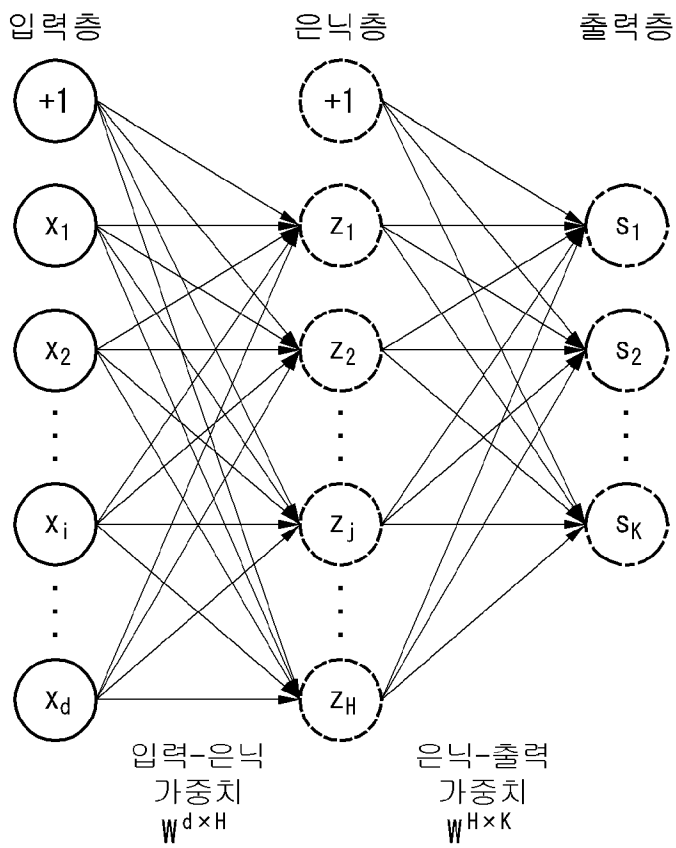


- DL/UL ACK/NACK
- UE CQI/PMI RI REPORT
- USING PUSCH AND PUCCH

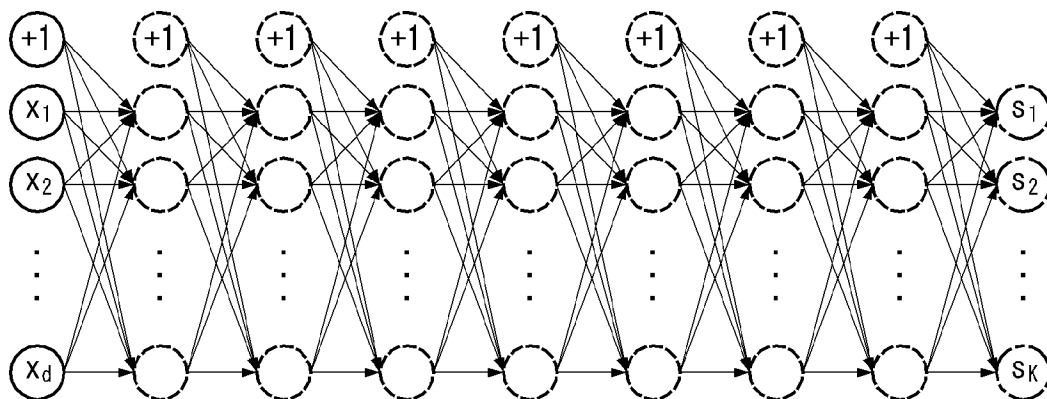
[도7]



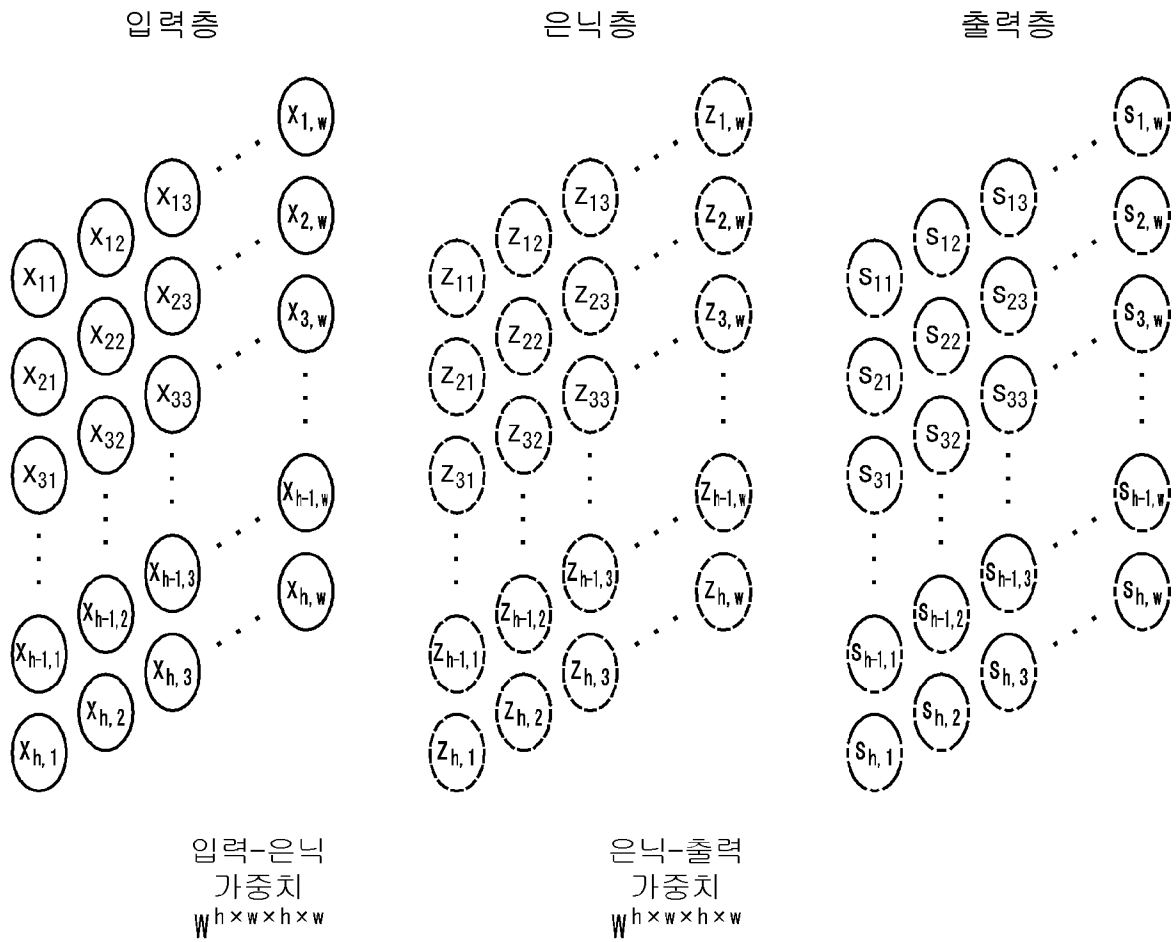
[도8]



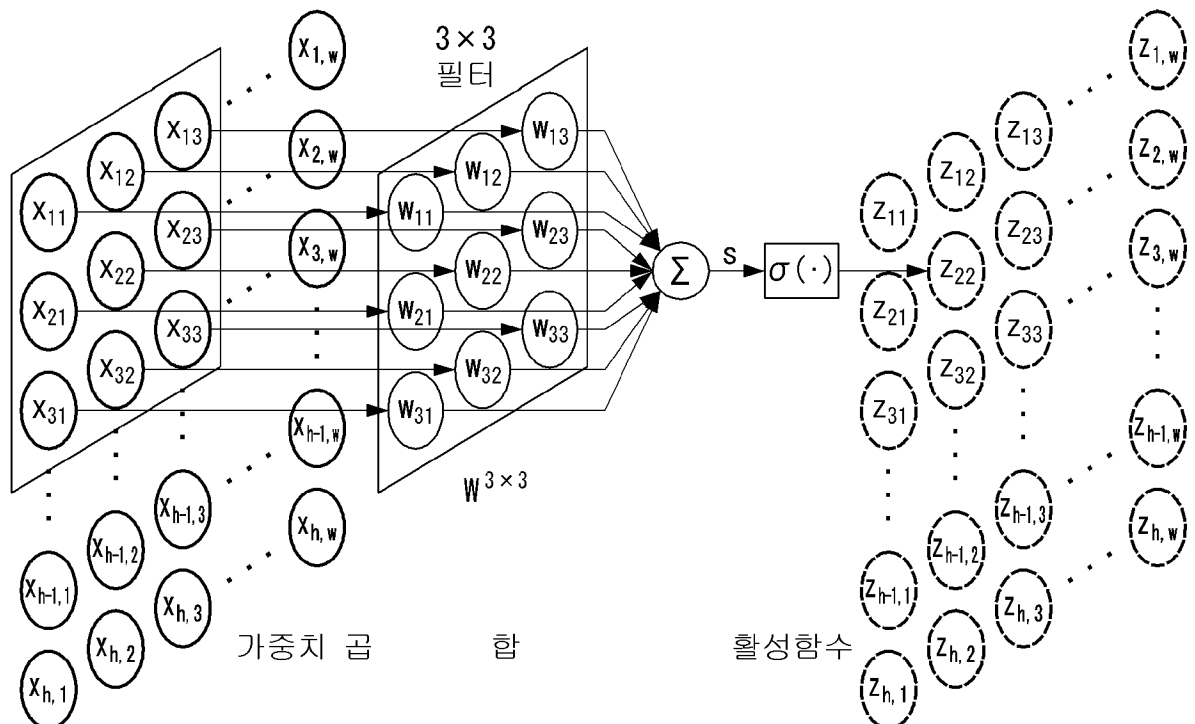
[도9]



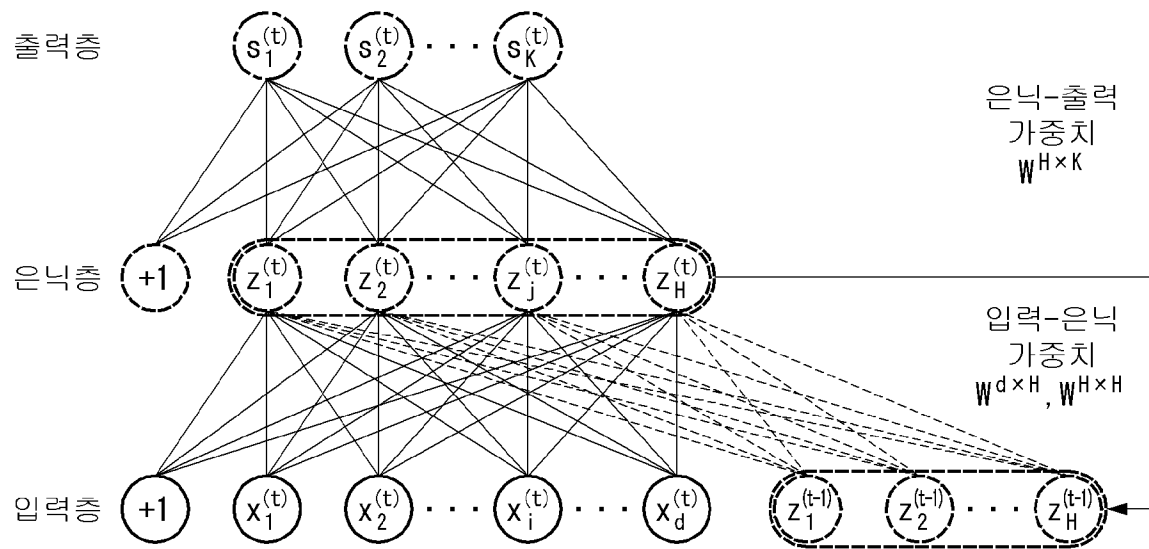
[도10]



[도11]



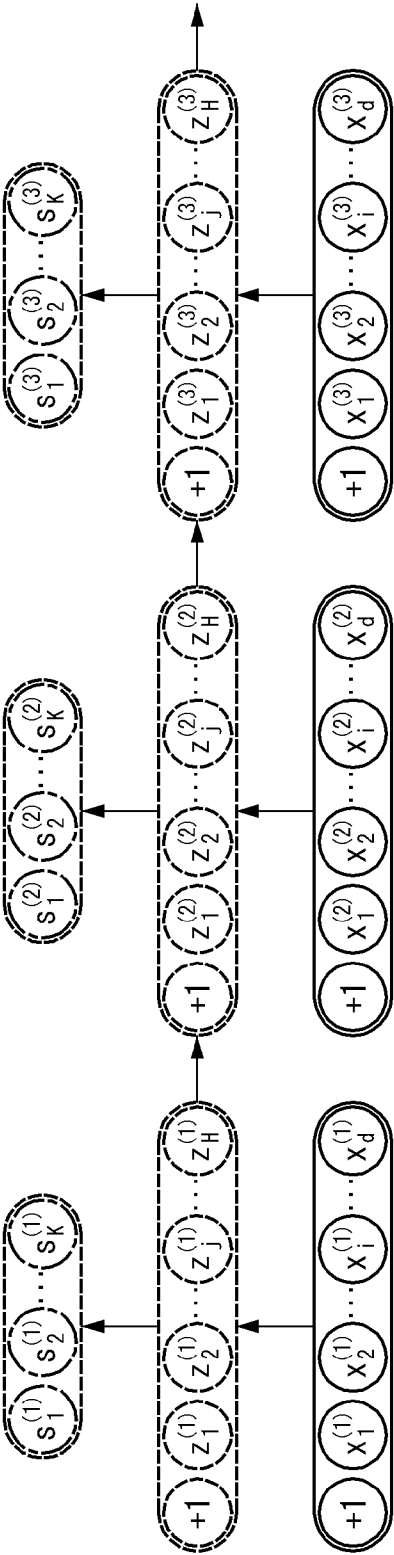
[도12]



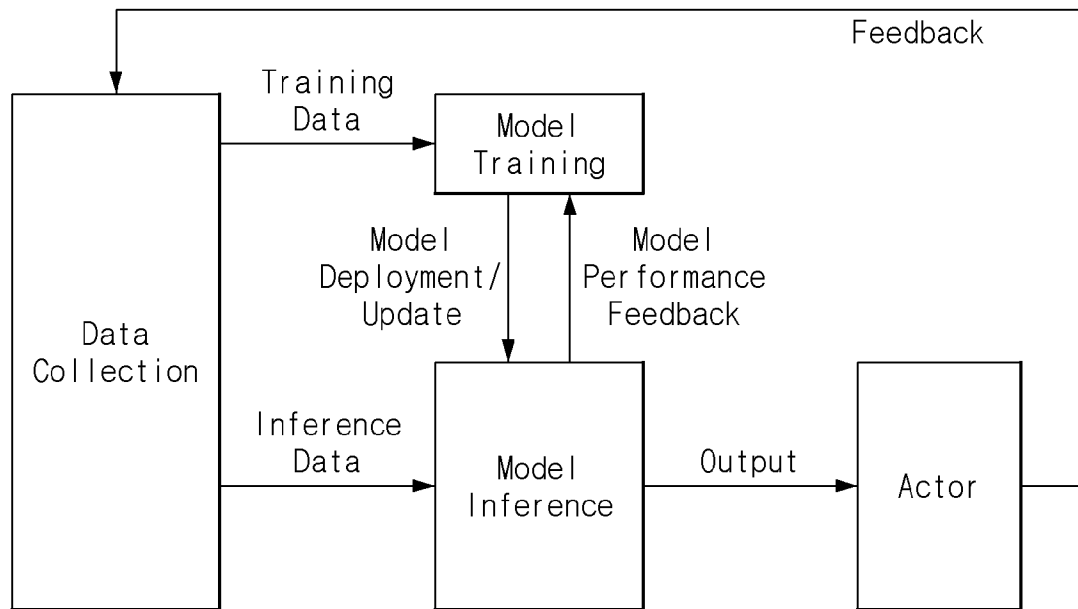
[도13]

시점 입력 시퀀스

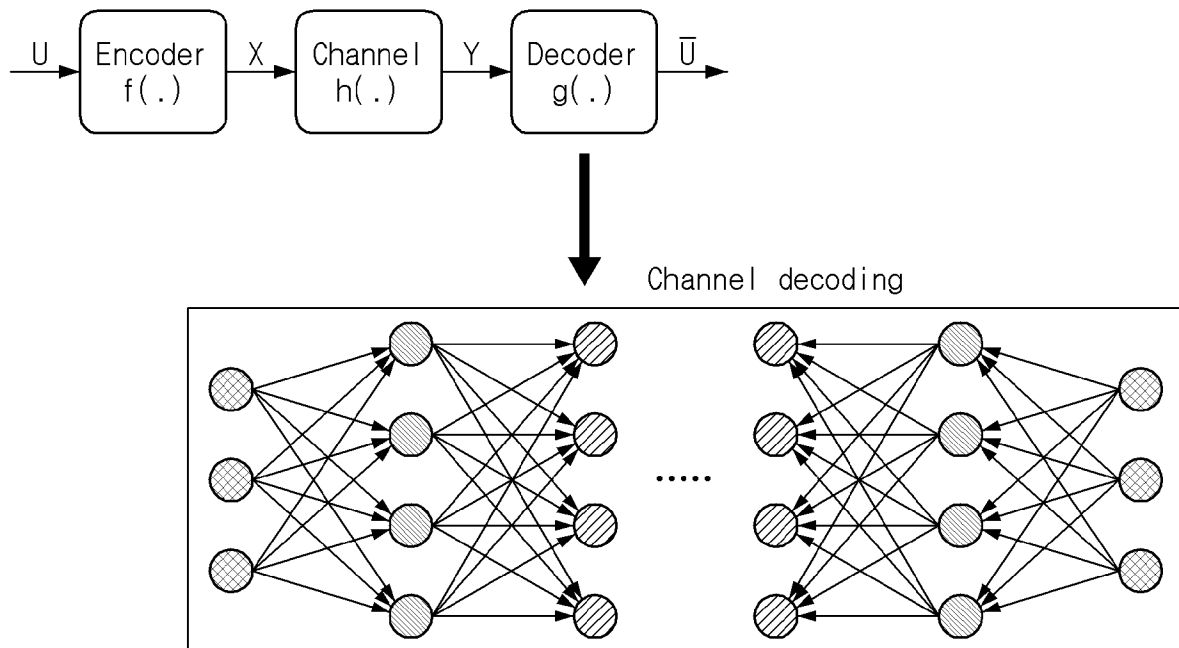
1	$(x_1^{(1)}, x_2^{(1)}, \dots, x_d^{(1)})$
2	$(x_1^{(2)}, x_2^{(2)}, \dots, x_d^{(2)})$
3	$(x_1^{(3)}, x_2^{(3)}, \dots, x_d^{(3)})$
	$\vdots$
t	$(x_1^{(t)}, x_2^{(t)}, \dots, x_d^{(t)})$
	$\vdots$
T	$(x_1^{(T)}, x_2^{(T)}, \dots, x_d^{(T)})$



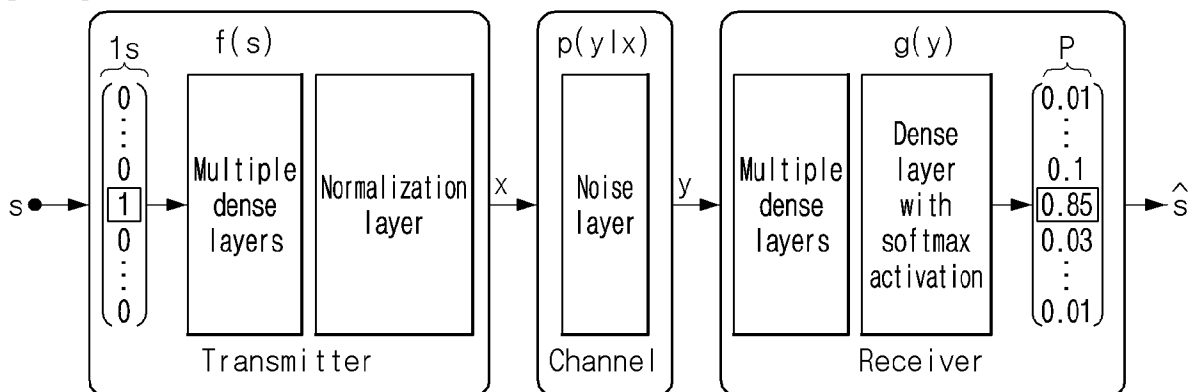
[도14]



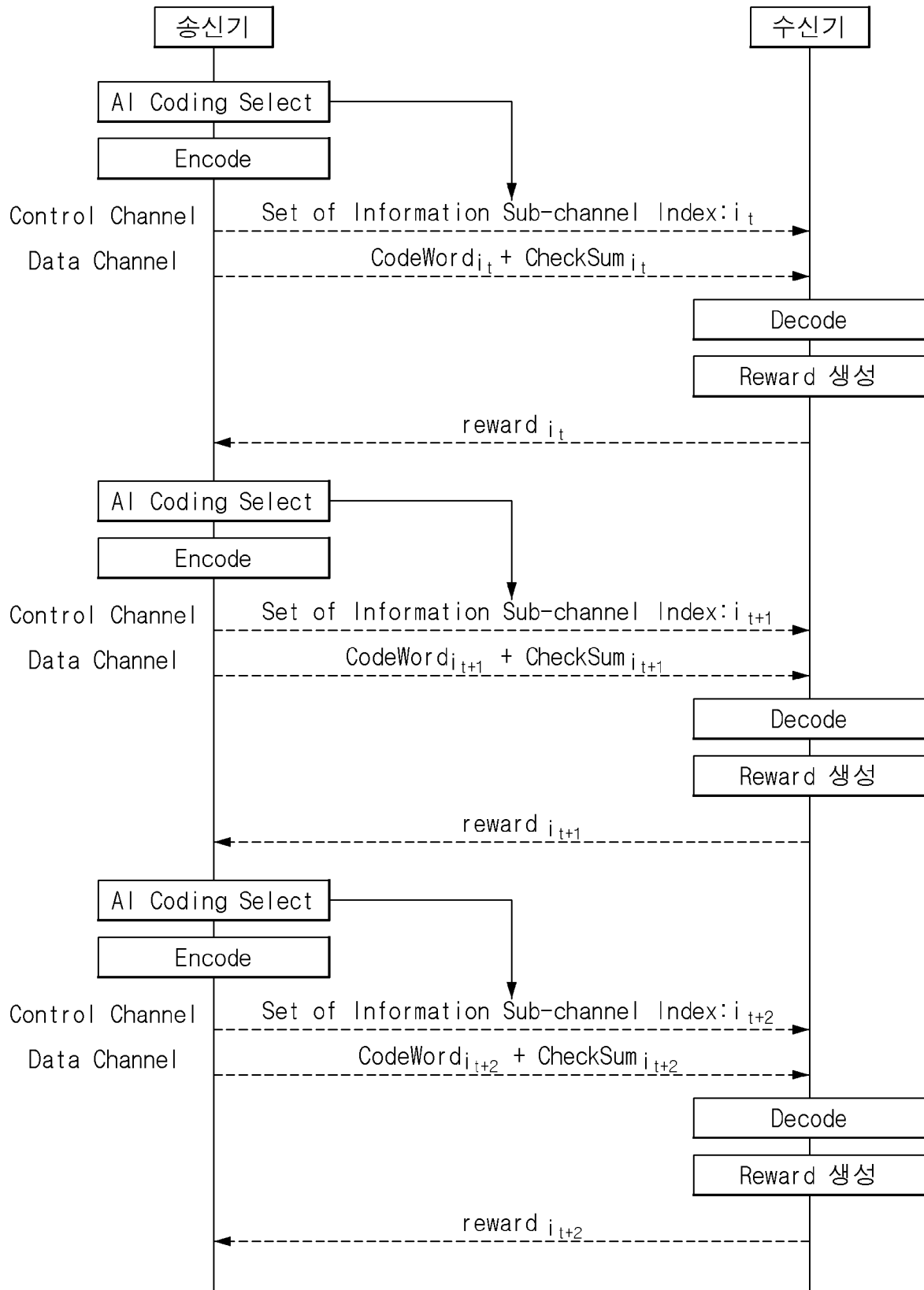
[도15]



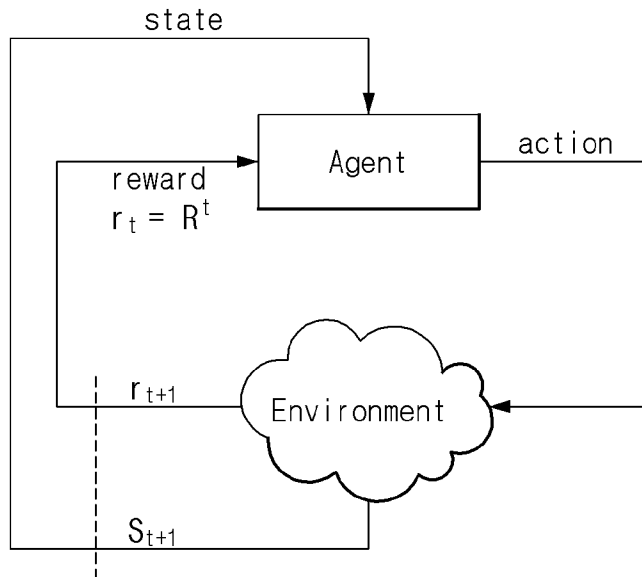
[도16]



[도17]

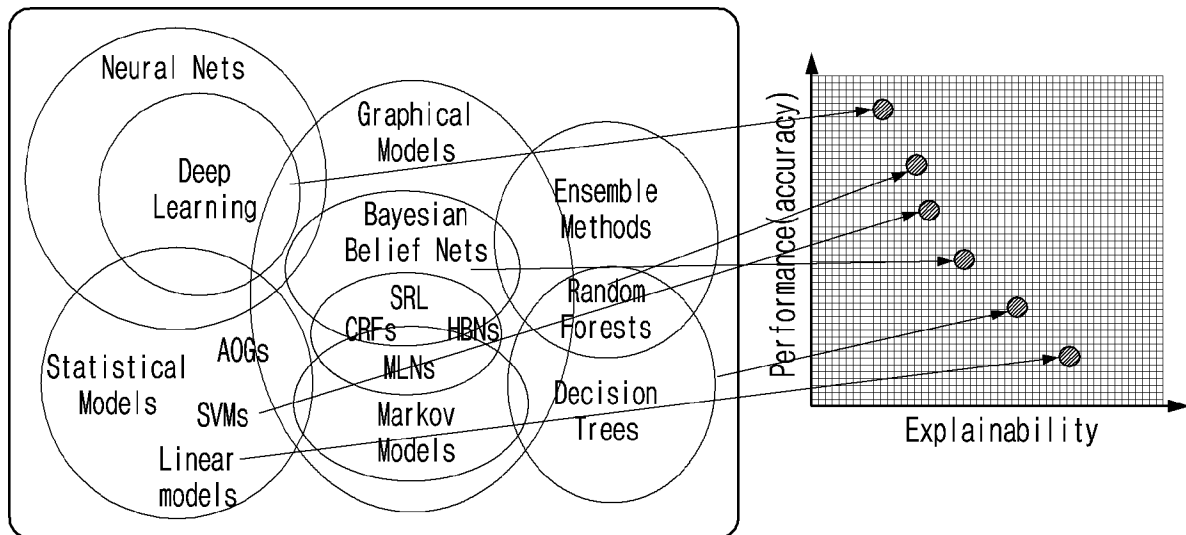


[도18]

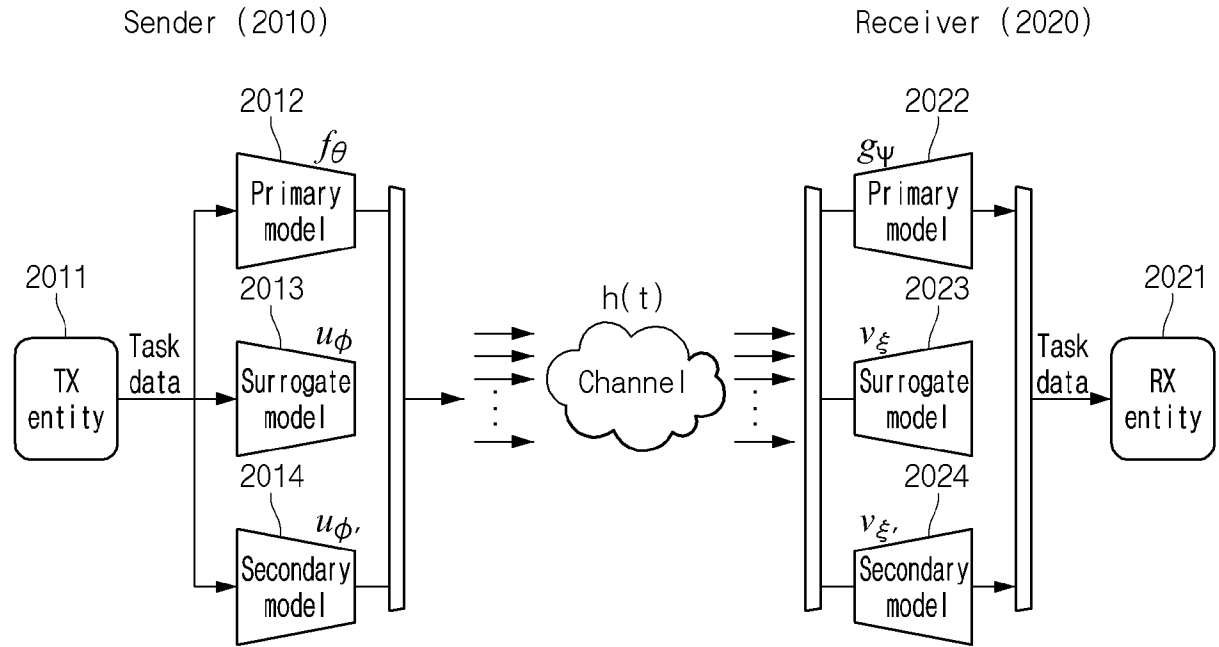


[도19]

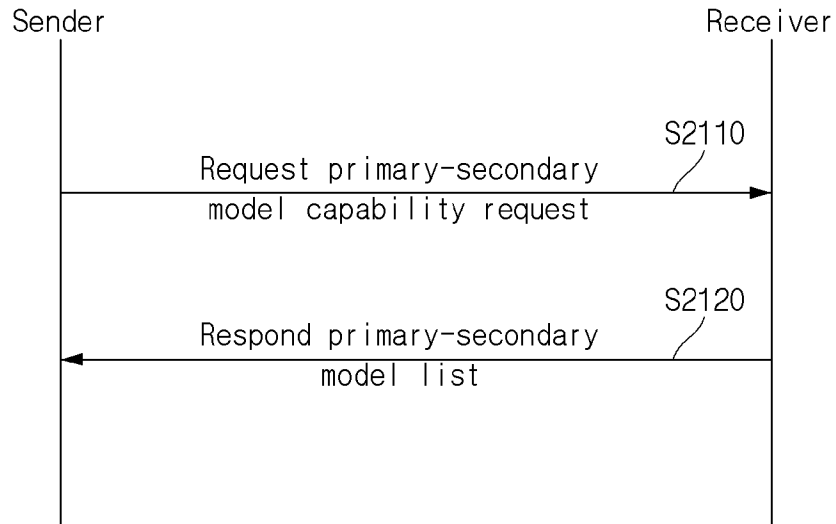
Learning Techniques



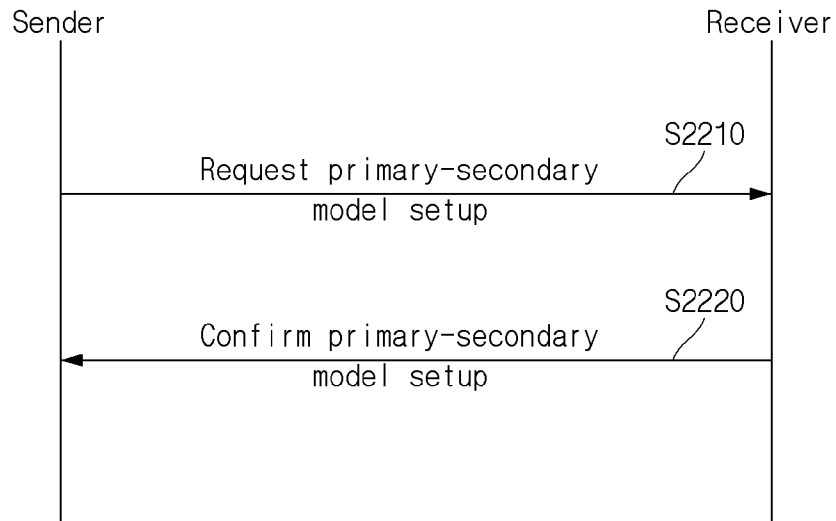
[도20]

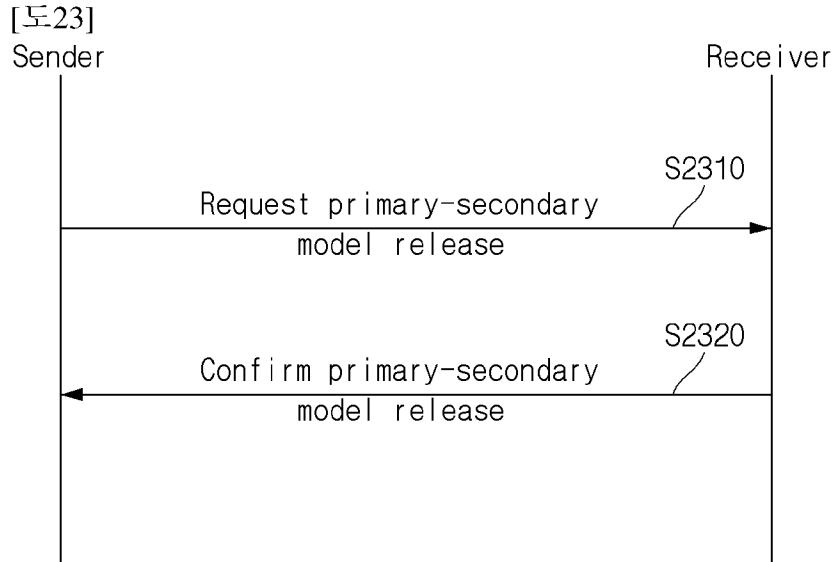


[도21]

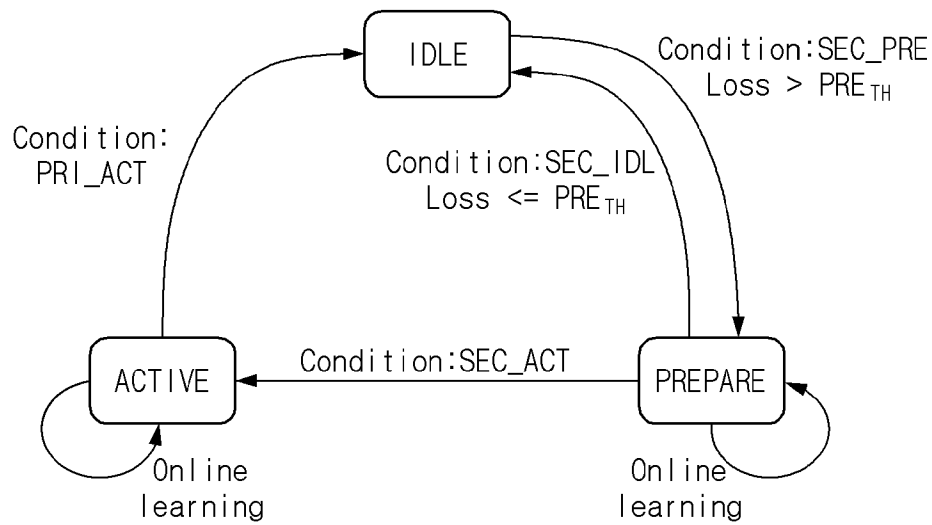


[도22]

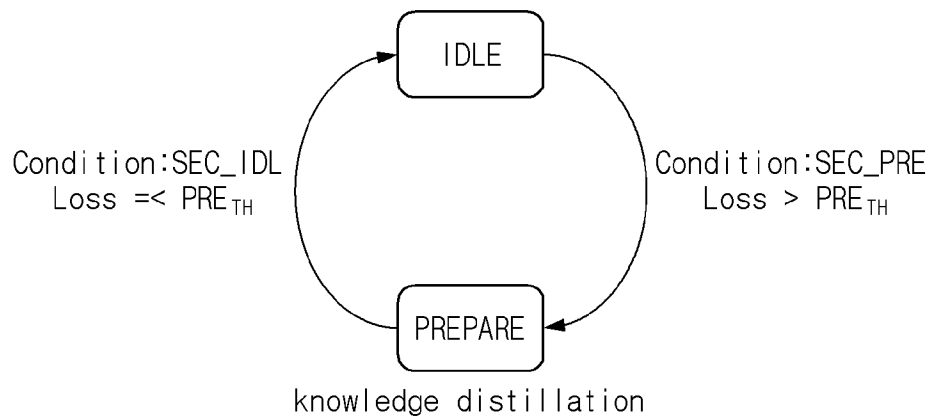




[도24]
State transition of secondary model

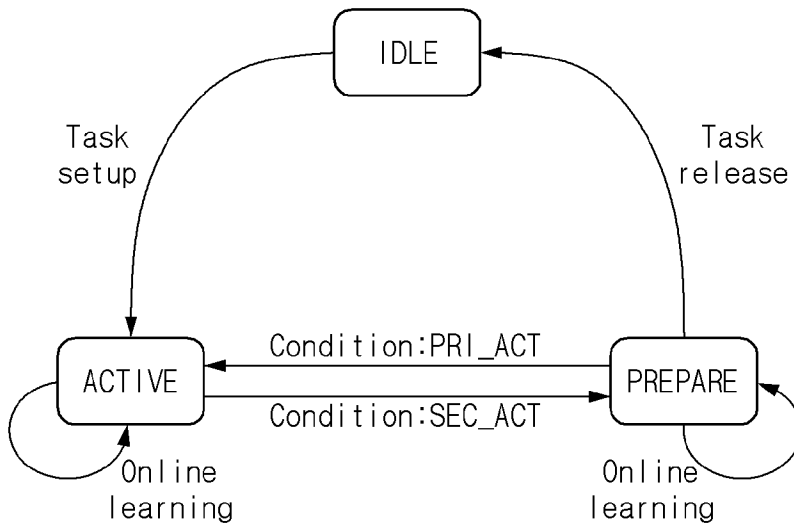


[도25]
State transition of surrogate model

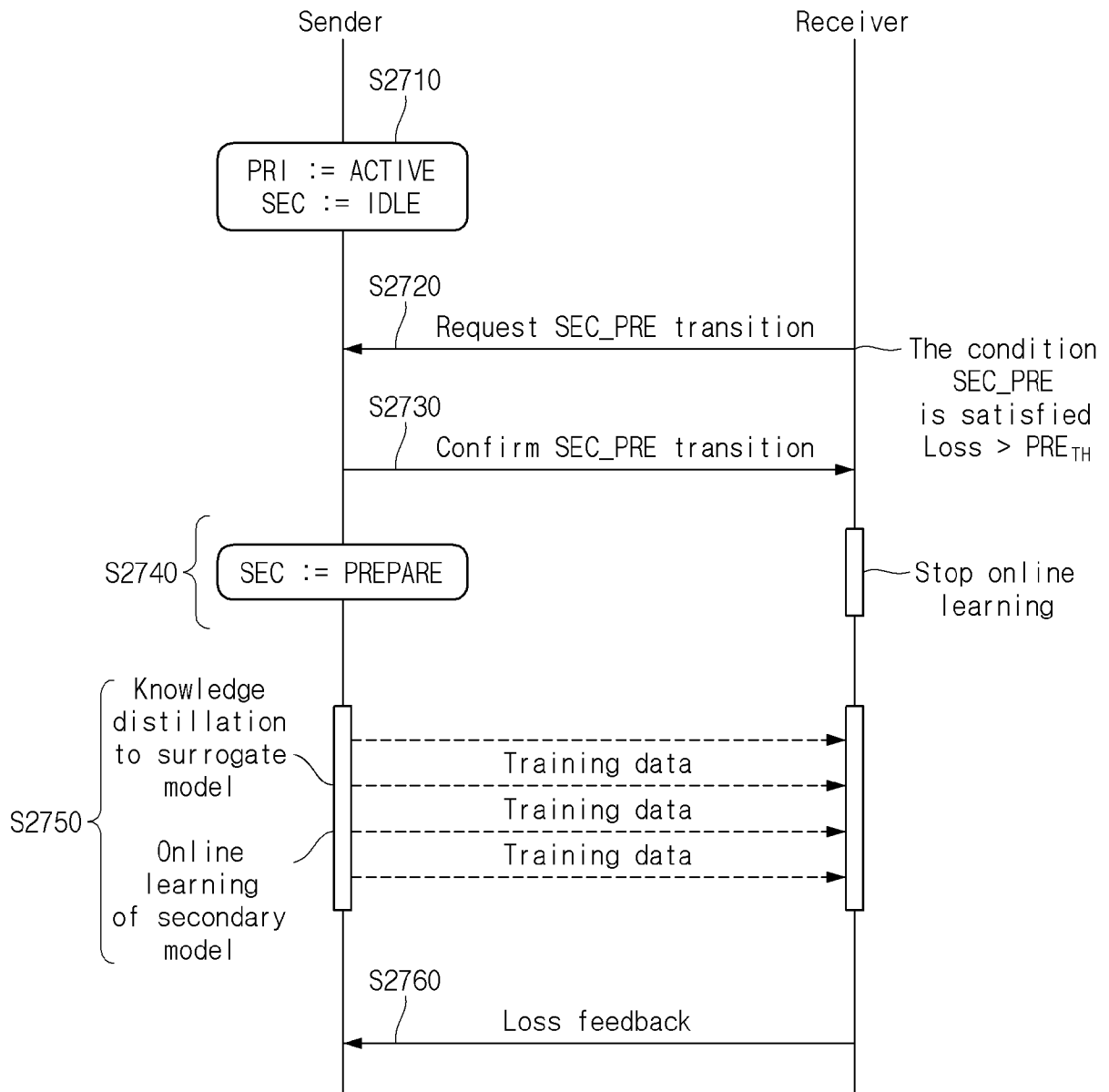


[도26]

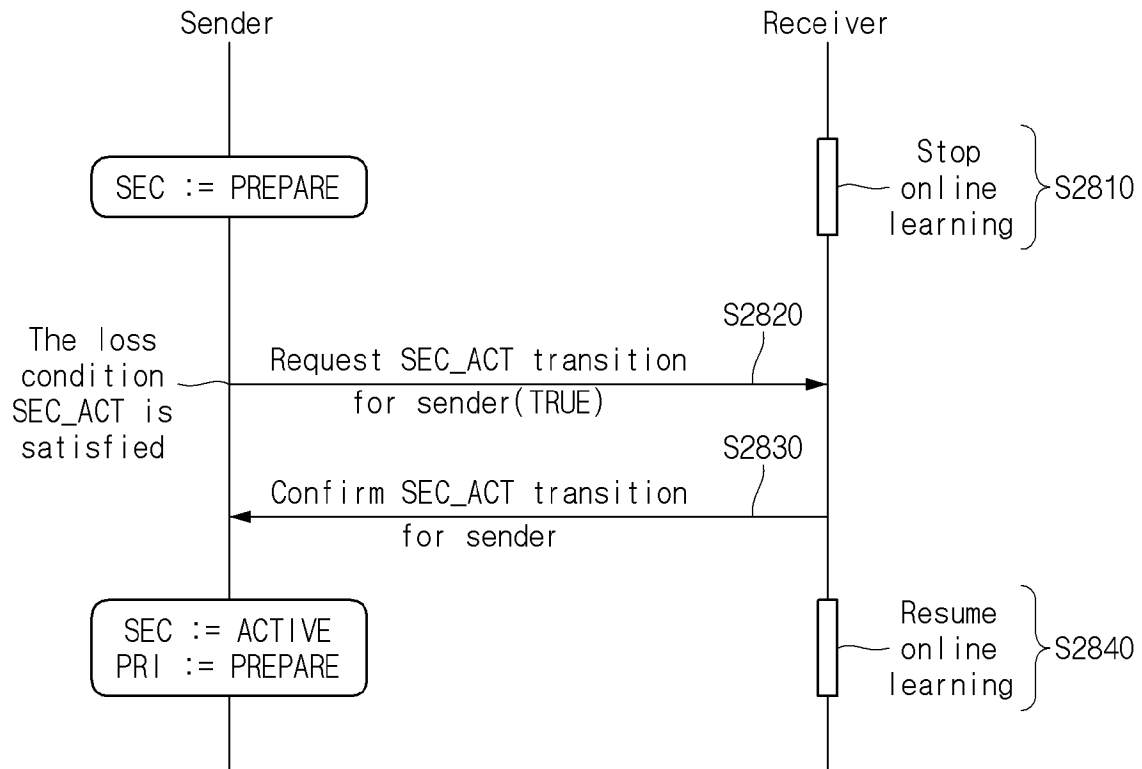
State transition of primary model



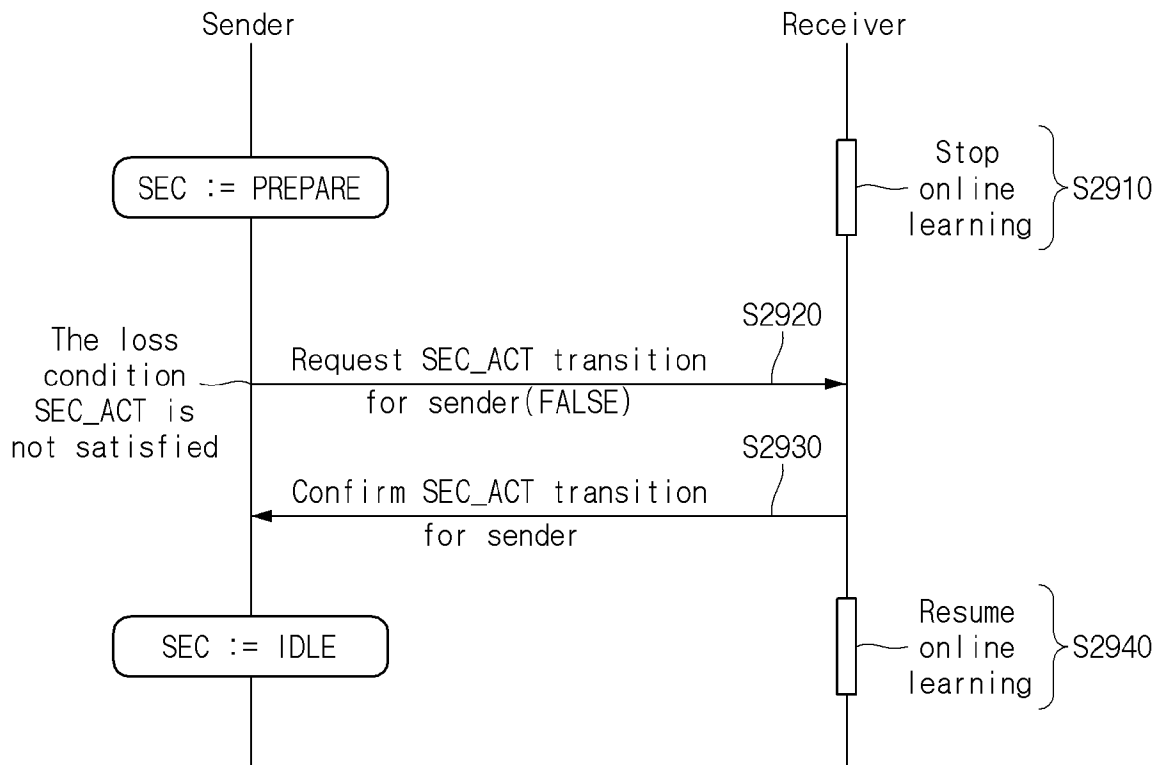
[도27]



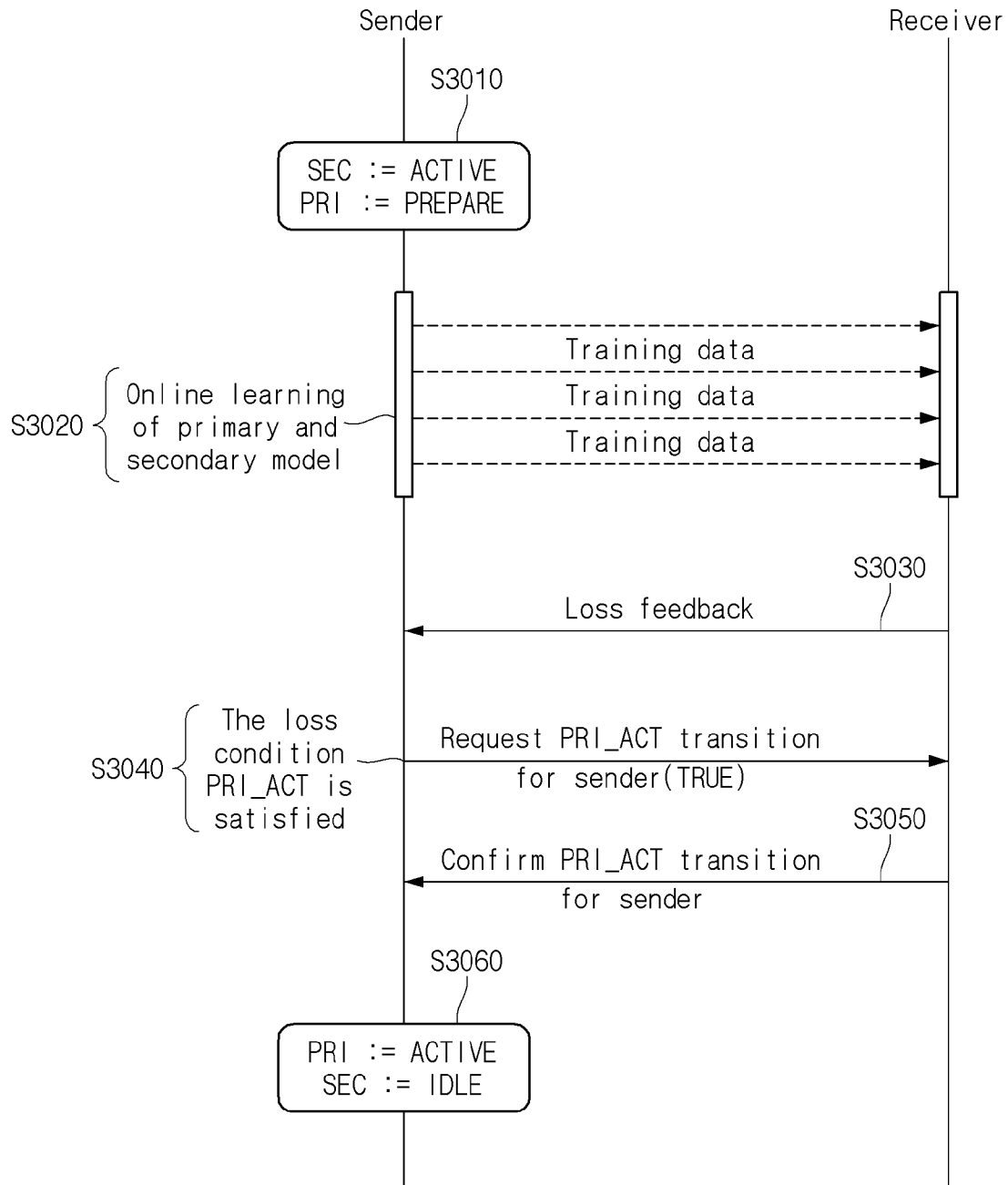
[도28]



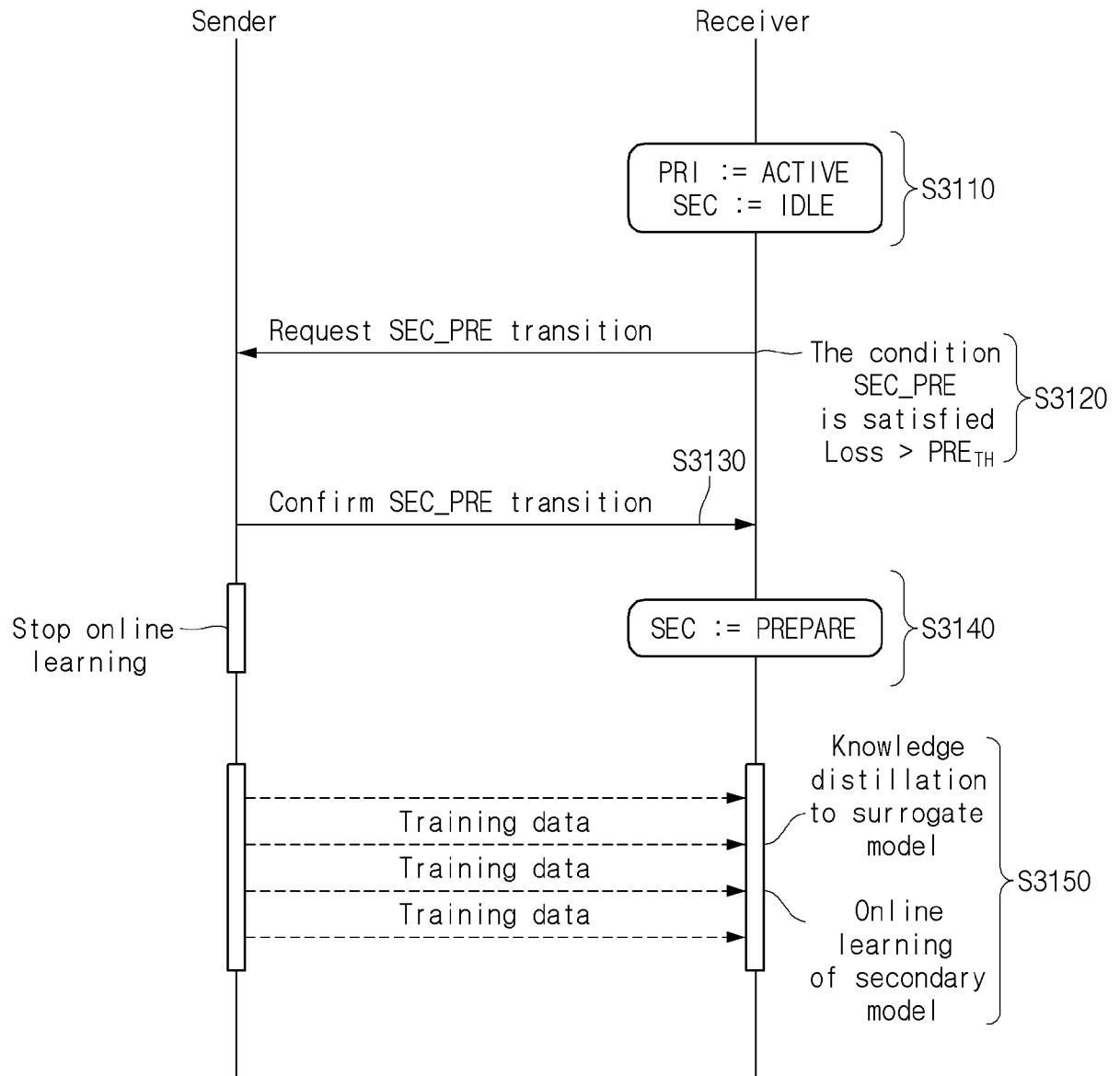
[도29]



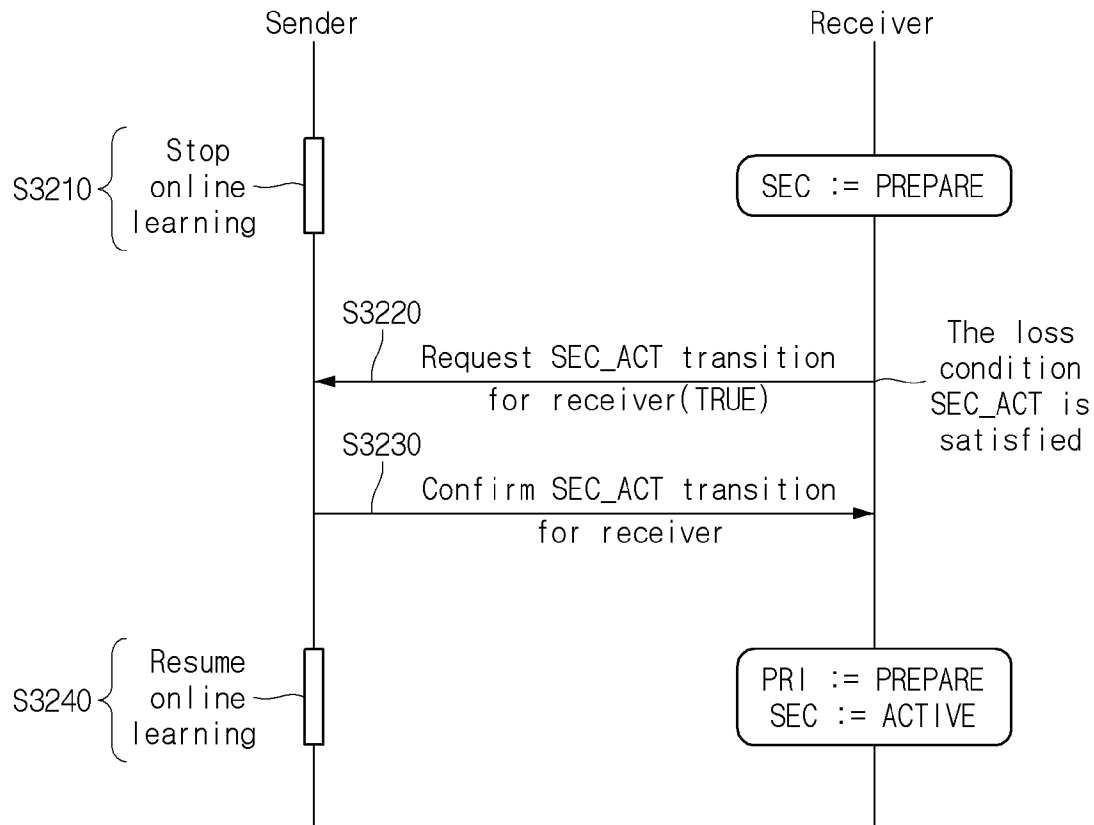
[도30]



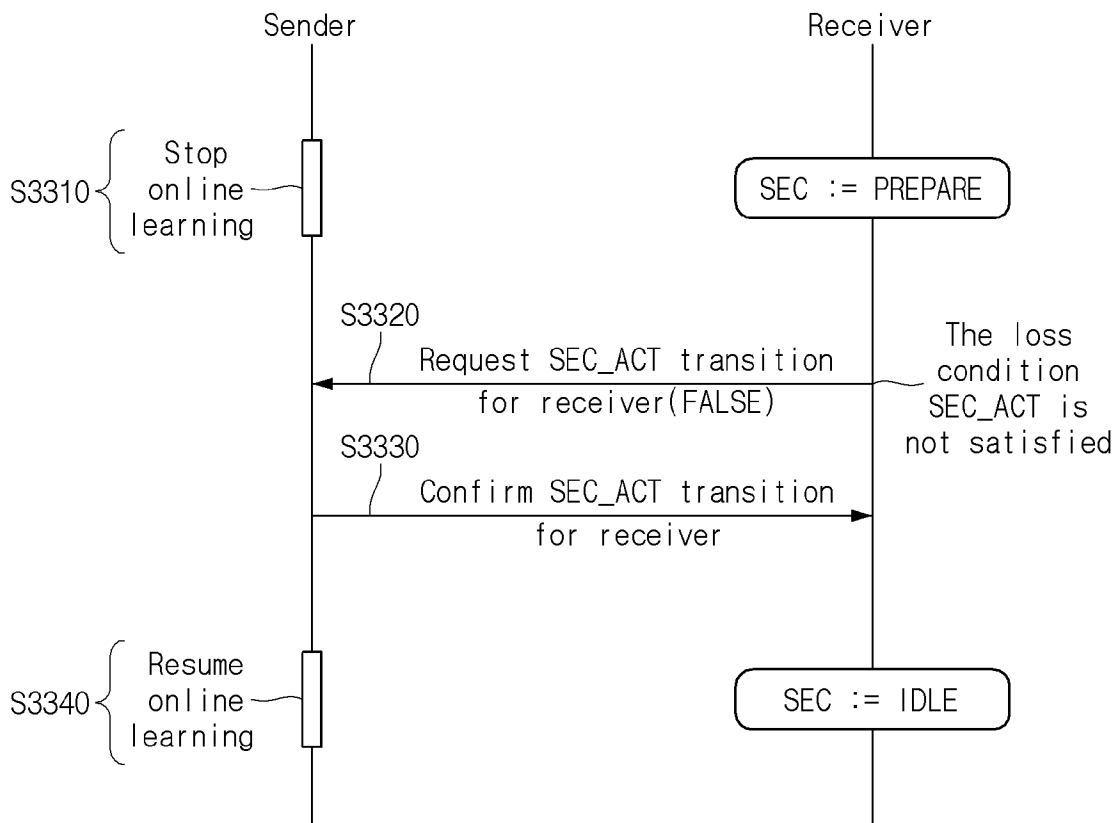
[도31]



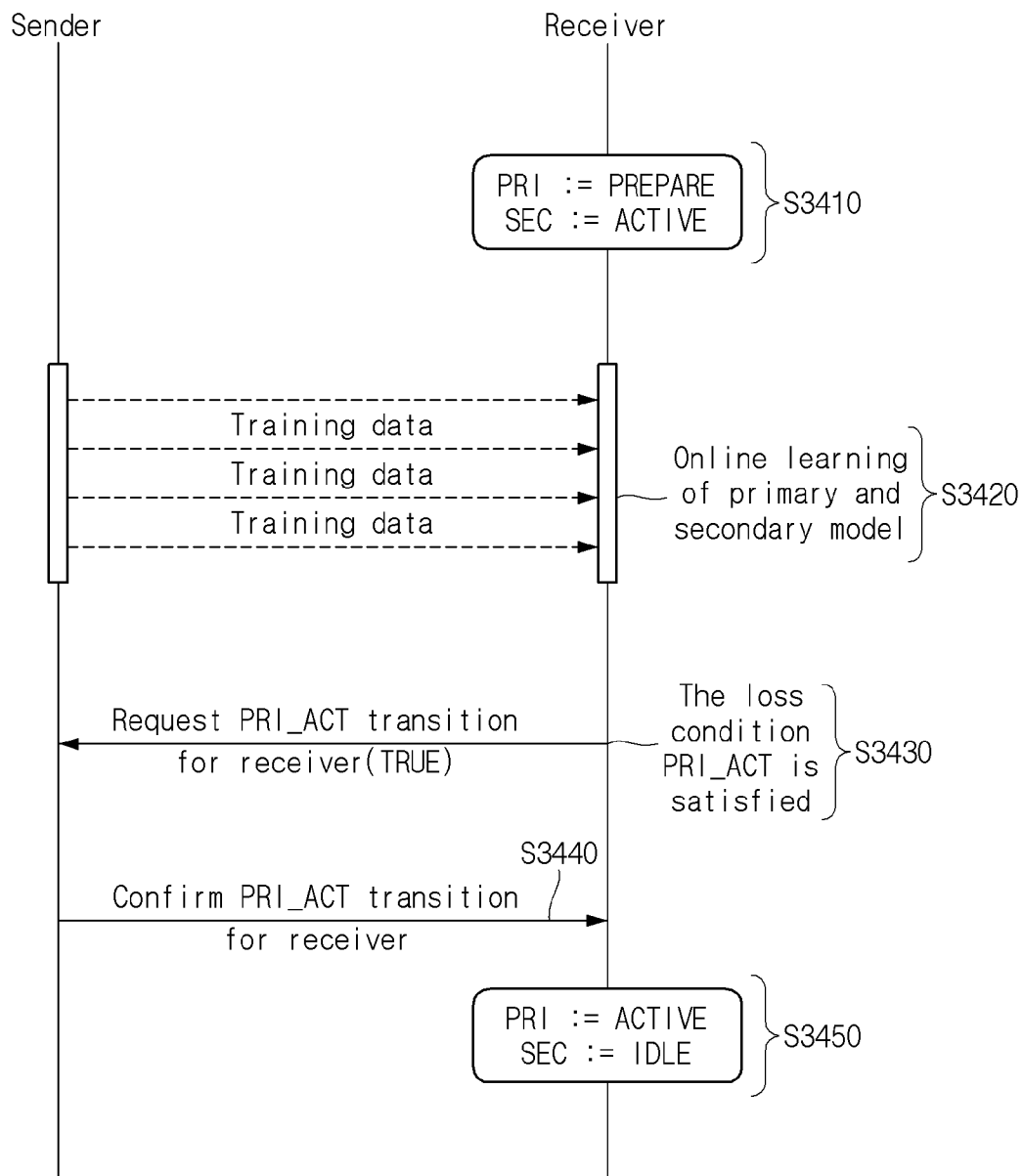
[도32]



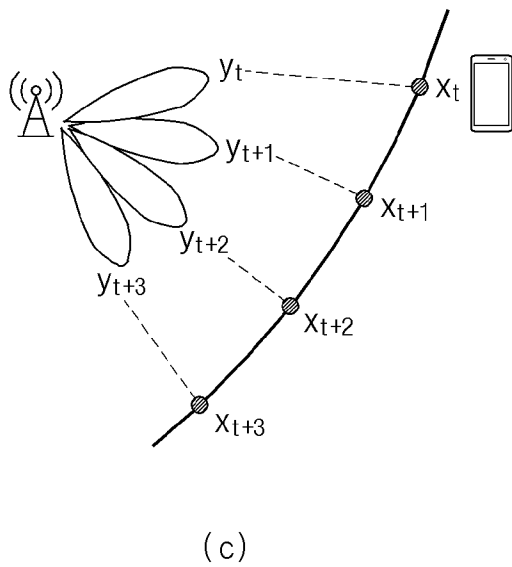
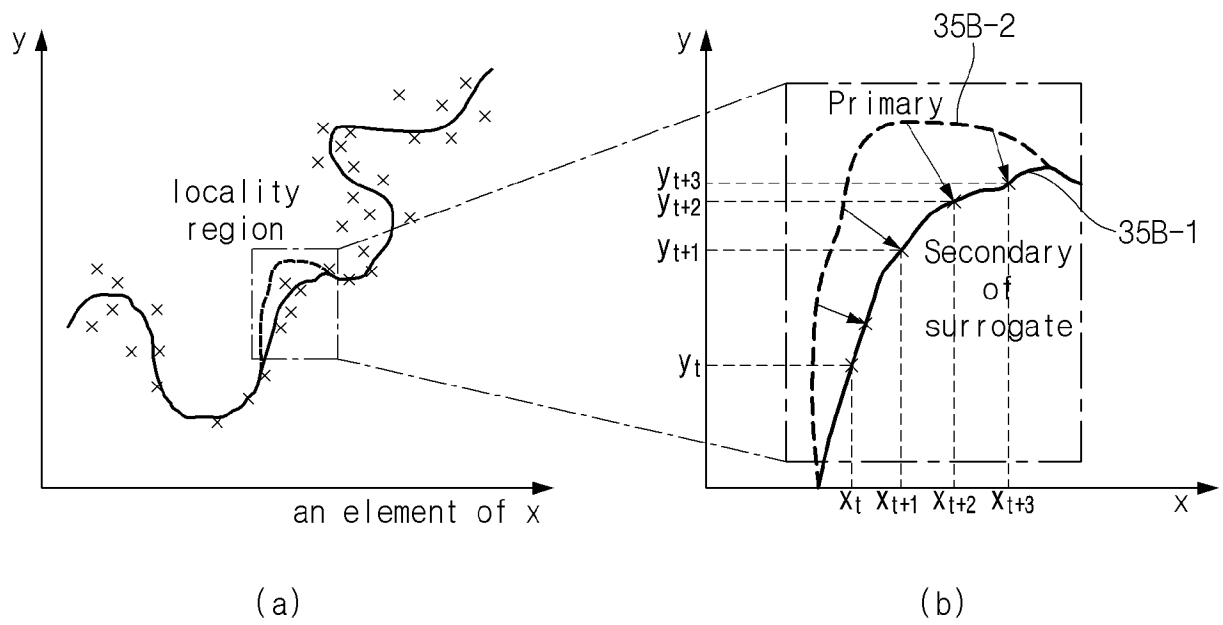
[도33]



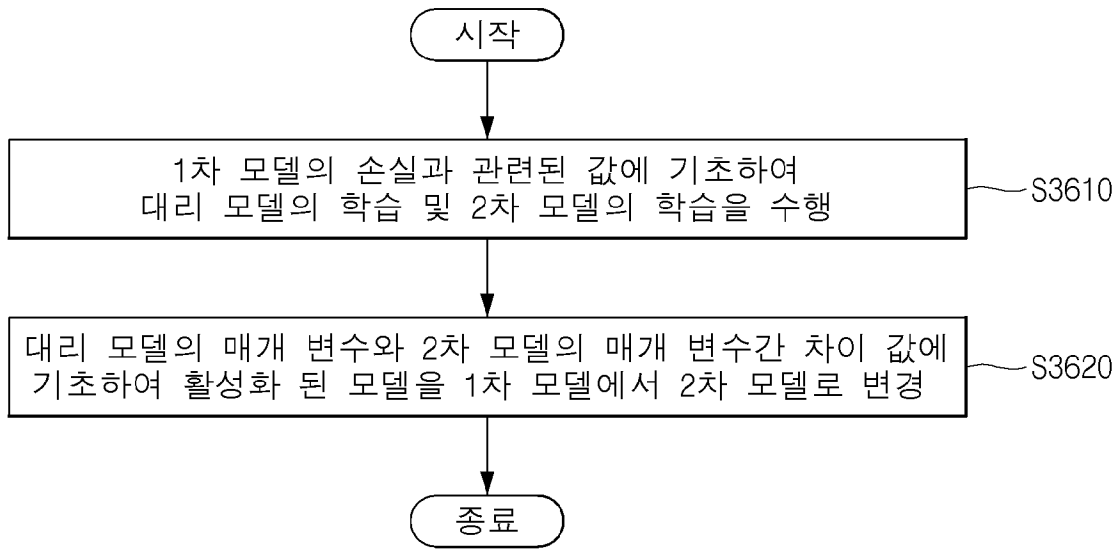
[도34]



[도35]



[도36]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2023/015353

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/096(2023.01)i; G06N 3/047(2023.01)i; G06N 3/042(2023.01)i; G06N 5/04(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N 3/096(2023.01); G06K 9/62(2006.01); G06N 20/20(2019.01); G06N 3/04(2006.01); G06N 3/08(2006.01);
G06N 3/12(2006.01); G06N 99/00(2010.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 신경망(neural network), 모델(model), 손실(loss), 2차(second), 전이 학습(transfer learning), 가중치(weight), 차이(difference), 훈련(training), 액티브 상태(active state)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KR 10-2023-0085842 A (ENERZAI INC.) 14 June 2023 (2023-06-14) See paragraphs [0042], [0045], [0060], [0090] and [0129]; and claims 1 and 4.	1-21
Y	KR 10-2023-0059435 A (SAMSUNG ELECTRONICS CO., LTD. et al.) 03 May 2023 (2023-05-03) See paragraphs [0002] and [0005]; and claim 1.	1-21
Y	KR 10-1897962 B1 (KOREA INTERNET & SECURITY AGENCY) 31 October 2018 (2018-10-31) See claim 1.	8-10,17
A	US 2022-0180185 A1 (MICRON TECHNOLOGY, INC.) 09 June 2022 (2022-06-09) See paragraphs [0052]-[0065]; and figures 3-5.	1-21
A	US 2021-0287141 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 16 September 2021 (2021-09-16) See claims 1-10.	1-21



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

20 June 2024

Date of mailing of the international search report

21 June 2024

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2023/015353

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
KR	10-2023-0085842	A	14 June 2023	KR	10-2465672	B1	14 November 2022
KR	10-2023-0059435	A	03 May 2023	US	2023-0130779	A1	27 April 2023
KR	10-1897962	B1	31 October 2018	None			
US	2022-0180185	A1	09 June 2022	WO	2022-125321	A1	16 June 2022
US	2021-0287141	A1	16 September 2021	US	11783025	B2	10 October 2023

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06N 3/096(2023.01)i; G06N 3/047(2023.01)i; G06N 3/042(2023.01)i; G06N 5/04(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06N 3/096(2023.01); G06K 9/62(2006.01); G06N 20/20(2019.01); G06N 3/04(2006.01); G06N 3/08(2006.01);
G06N 3/12(2006.01); G06N 99/00(2010.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 신경망(neural network), 모델(model), 손실(loss), 2차(second), 전이 학습
(transfer learning), 가중치(weight), 차이(difference), 훈련(training), 액티브 상태(active state)

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	KR 10-2023-0085842 A (주식회사 에너지(ENERZAI)) 2023.06.14 단락 [0042], [0045], [0060], [0090], [0129]; 및 청구항 1, 4	1-21
Y	KR 10-2023-0059435 A (삼성전자주식회사 등) 2023.05.03 단락 [0002], [0005]; 및 청구항 1	1-21
Y	KR 10-1897962 B1 (한국인터넷진흥원) 2018.10.31 청구항 1	8-10,17
A	US 2022-0180185 A1 (MICRON TECHNOLOGY, INC.) 2022.06.09 단락 [0052]-[0065]; 및 도면 3-5	1-21
A	US 2021-0287141 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2021.09.16 청구항 1-10	1-21

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을
정의한 문헌

“D” 본 국제출원에서 출원인이 인용한 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일
이후에 공개된 선출원 또는 특허 문헌“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의
공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여
인용된 문헌

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과
상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기
위해 인용된 문헌“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의
신규성 또는 진보성이 없는 것으로 본다.“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른
문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우
청구된 발명은 진보성이 없는 것으로 본다.

“&” 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일

2024년06월20일(20.06.2024)

국제조사보고서 발송일

2024년06월21일(21.06.2024)

ISA/KR의 명칭 및 우편주소

대한민국 특허청
(35208) 대전광역시 서구 청사로 189, 4동 (둔산동,
정부대전청사)

팩스 번호 +82-42-481-8578

심사관

양정록

전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2023-0085842 A	2023/06/14	KR 10-2465672 B1	2022/11/14
KR 10-2023-0059435 A	2023/05/03	US 2023-0130779 A1	2023/04/27
KR 10-1897962 B1	2018/10/31	없음	
US 2022-0180185 A1	2022/06/09	WO 2022-125321 A1	2022/06/16
US 2021-0287141 A1	2021/09/16	US 11783025 B2	2023/10/10