

6G를 위한 시맨틱 통신: 응용, 기술 및 도전과제

한경국립대학교 | 최현호

1. 서론

1세대(1G)부터 5세대(5G) 이동통신의 발전 방향은 데이터 속도와 비트 오류 확률과 같은 데이터 지향 성능 지표를 최적화하는 것이었다. 이러한 통신 속도의 비약적인 발전은 Shannon이 정보 이론을 통해 잡음 채널에서 신뢰할 수 있는 통신이 가능하다는 것을 증명함으로써 가능하였다. 그러나 5G까지의 데이터 중심의 통신 아키텍처는 사용자에게 높은 품질 경험(quality-of-experience, QoE)을 제공하는 데 장벽이 되고 있다. 이는 특히 6G에서 등장할 서비스가 인간 중심에서 더 엄격한 경험적인 성능 지표를 요구하기 때문이다 [1]. 일례로 메타버스 서비스에서는 헤드 마운트 디스플레이를 통해 가상 세계를 탐험하고, 증강 현실 안경을 통해 증강된 물리적 세계를 탐색하여, 가상과 물리적 도메인에서의 인간 행동을 서로 연결할 것이다. 이러한 응용에 수반되는 방대한 데이터와 긴 종단간 지연 문제를 극복하고 사용자 중심의 QoE 지표를 만족시키기 위해서는 고전적인 정보 이론 기반의 통신 네트워크에 대한 재고가 필요하다 [2].

6G 네트워크에서는 메타버스와 같은 서비스의 등장으로 많은 양의 데이터를 무선으로 전송할 수 있어야 한다. 또한, 협력 로봇 및 초 지능형 Internet-of-things (IoT)와 같은 많은 수의 노드를 지원하는 6G 서비스는 빠른 시스템 응답과 신뢰성을 갖고 효율적으로 정보를 주고받아야 한다. 아울러, 더 나은 QoE 보장하기 위해 정보의 실시간 업데이트와 사용자 데이터 분석에 더 많은 네트워크 컴퓨팅 자원을 제공해야 한다. 이를 위해 시맨틱 통신(semantic communication)이라는 새로운 패러다임이 새로운 기술로 등장하였다 [3]. 이는 통신 중 메시지의 의미를 식별하고 활용하는 기술로 기존의 데이터 중심 통신과는 달리, 데이터가 artificial intelligence(AI) 기술을 통해 사전 처리된

후 수신자의 의도나 작업 목표와 관련 있는 정보만을 전송하여 네트워크 부담을 경감시킨다. 즉, 시맨틱 통신은 기존의 ‘전송 후 이해’ (transmission-before-understanding) 방식의 통신 패러다임을 ‘이해 후 전송’ (understanding-before-transmission) 패러다임으로 전환한다 [4]. 이를 통해 송신단은 원본 데이터 대신 통신 의도나 목적에 부합하는 필요한 정보만 전송함으로써 대역폭 사용을 줄일 수 있다.

원래 시맨틱 관련 연구는 컴퓨터 비전에서 이미지의 동일 객체로 묶인 부분들을 군집화하는 시맨틱 분할 [5], 콘텐츠와 사용자 의도의 의미를 도출하고 일치시키는 시맨틱 컴퓨팅 [6], 연결된 데이터와 지능형 콘텐츠를 결합하여 형성된 지식 그래프를 활용하여 추천 시스템으로 널리 사용되는 시맨틱 웹 [7]과 같은 분야에서 수행되어왔다. 이들과 비교하여, 시맨틱 통신은 새로운 연구 분야로, 모든 통신 주체는 시맨틱 표현과 해석을 고도로 일치시켜야 하며, 시맨틱 정보의 전송시에는 강한 시간 민감성과 높은 정확도가 요구된다 [8]. 이는 다른 분야에서는 찾아볼 수 없는 엄격한 요구사항으로 새로운 기술과 도전을 필요로 한다.

시맨틱 관련 다양한 연구 분야 중에서 본 고에서는 6G 통신 환경 및 전송과 관련된 시맨틱 통신 기술 소개에 중점을 둔다. 2장에서는 시맨틱 통신의 기본 개념과 시스템 모델을 소개하고, 3장에서 6G에서의 시맨틱 통신의 응용 분야를 제시한다. 4장에서는 시맨틱 정보 전송 기술을 소개하고, 5장에서는 시맨틱 통신을 위한 프로토콜 구조를 제시한다. 6장에서는 6G에서 시맨틱 통신을 실현하기 위한 도전과제를 논의하고, 마지막으로 7장에서 결론을 맺는다.

2. 시맨틱 통신이란?

전통적인 통신 시스템에서는 소스 데이터의 평균 정보가 중요한 반면, 시맨틱 통신 설계에서 중요한 것은 소스 데이터의 의미 내용의 정확성이다 [9]. 그림

* 이 성과는 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1011901).

1의 두 시스템 모델을 비교하면, 시맨틱 통신 시스템의 주요 변화는 송신 전과 수신 후의 데이터 처리 단계에 있다. 전통적인 소스 인코딩은 메시지에 내재된 의미와 상관없이 많은 가능성의 소스 데이터를 짧은 코드로 변환하여 정보를 정량화하고자 하지만, 시맨틱 통신에서는 정보를 지식을 산출하는 소스로 보고, 전달하는 정보를 통해 무언가 배울 수 있는 것으로 정의한다 [10].

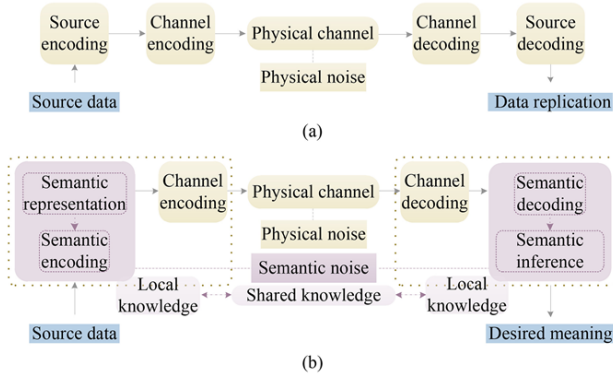


그림 1 (a) 기존 통신 시스템 모델과 (b) 시맨틱 통신 시스템 모델 [4]

이러한 관점에서 시맨틱 통신에서는 인코딩 전에 시맨틱 표현 모듈을 도입하여 소스 데이터에 내재된 핵심 정보를 포착하고 불필요한 중복 정보를 걸러낸다. 일반적으로 시맨틱 표현 기능은 시맨틱 인코딩 모듈과 통합하여 기존 통신의 소스 인코딩과 유사한 역할을 수행한다. 즉, 시맨틱 추론과 시맨틱 디코딩은 기존 소스 디코딩에 대응된다. 시맨틱 통신의 목적이 수신자가 시맨틱 정보를 성공적으로 추론할 수 있도록 하는 것이기 때문에, 일반적으로 시맨틱 디코딩은 인코딩의 역과정이 되며, 이는 트랜스포머(transformer)나 오토인코더(autoencoder)와 같은 사전 지식을 가진 강력한 AI 기술을 기반으로 한다 [11], [12].

또한, 사람들이 효과적인 대화를 위해서 서로의 언어와 문화에 대한 공통 지식이 필요한 것처럼, 효과적인 시맨틱 통신을 위해서는 통신 당사자들의 로컬 지식(local knowledge)이 실시간으로 공유되어야 이해 및 추론 과정이 모든 소스 데이터에 대해 잘 일치할 수 있다. 로컬 지식이 동기화되지 않으면 물리적 전송 오류가 없는 경우에도 의미적 잡음(semantic noise)이 발생하여 의미적 모호함이 생긴다. 따라서, 그림 1(b)와 같이 시맨틱 통신 시스템에서는 송수신단 간에 공유된 로컬 지식이 포함된다.

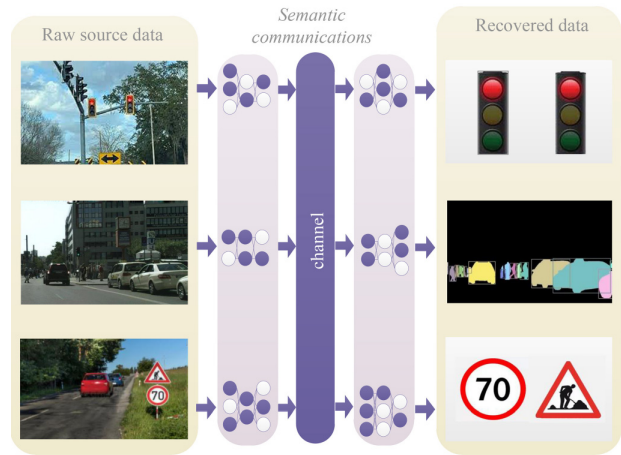


그림 2 ITS를 위한 시맨틱 통신의 예

그림 2는 intelligent transportation system (ITS)를 위한 시맨틱 통신의 예를 보여준다 [4]. 이미지 복제를 목표로 하는 기존 통신에서는 이미지의 내용과 상관없이 수신한 이미지가 원본 이미지의 모든 픽셀값을 보존해야 하지만, 시맨틱 통신에서는 시맨틱 인코딩을 통해 전송 전 적절한 이미지 처리를 수행하여 다양한 작업에 대해 관련 없는 이미지 세부 사항을 걸러낸다. 이는 교통 신호 감지, 주변 객체 탐색, 표지판 인식 등의 통신 목적에 부합하면서도 전송 트래픽을 현저히 줄일 수 있다.

3. 시맨틱 통신의 6G 응용 분야

본 장에서는 6G 환경에 적용 가능한 시맨틱 통신 기반 응용 서비스를 소개한다.

3.1 eXtended Reality (XR)

6G는 물리적 세계와 가상 세계의 동기화를 통해 XR 서비스를 제공하고자 한다. XR의 성능은 사람의 움직임과 주변 환경의 변화를 반영하는 데이터의 수집 및 처리에 크게 의존한다. 또한, 이상적인 몰입형 메타버스 QoE를 제공하기 위해서는 종단 간 지연 및 전송 속도 요구사항이 엄격히 충족되어야 한다 [13], [14]. 이러한 측면에서, 시맨틱 통신은 XR 기반 메타버스의 구현을 가능하게 하는 기술로 볼 수 있다 [15]. 그림 3에서 보듯이 시맨틱 통신을 통해, 단말 장치(예: 스마트 카메라)에서 캡처된 배경, 객체, 움직임, 표정, 음성 등으로부터 먼저 의미를 추출하고 불필요한 정보를 필터링함으로써 대역폭 및 전송 지연을 줄일 수 있다. 동시에, 압축된 시맨틱 정보를 수신한 XR 서버는 생성형 AI 기술을 통해 사용자의 의도에 맞게 가상의 3D 모델을 생성할 수 있다.

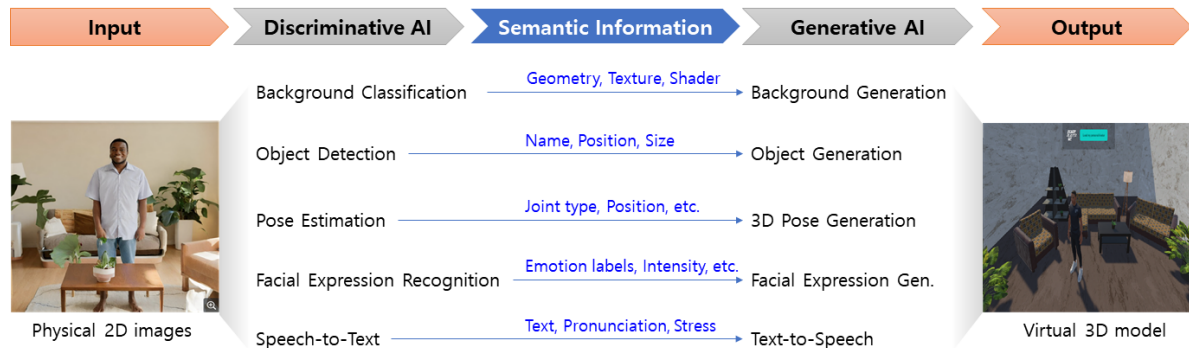


그림 3 XR 서비스를 위한 시맨틱 통신의 적용 예

3.2 Hologram Telepresence (HT)

HT는 원거리의 사람이나 사물을 실시간으로 물리적 존재에 버금가는 현실감 있는 3D 이미지로 투영해 줌으로써 가상 회의, 가상 게임, 원격 수술 등에 사용된다 [16]. 그러나 몰입형 XR과 마찬가지로, 실제와 같은 서비스를 제공하려면 HT도 엄격한 QoE가 요구된다. 이러한 통신 및 컴퓨팅 집중형 서비스의 경우, 기존 데이터 전송 방식은 대역폭과 컴퓨팅 자원의 낭비가 심하므로 시맨틱 통신 방식이 대역폭을 줄이고 수신 측의 처리 부담을 줄이는 유망한 방법이 될 수 있다. 시맨틱 인코딩은 통신 네트워크 상황과 QoE에 따라 시맨틱 정보를 적응적으로 추출함으로써 사용자에게 고품질의 HT 서비스 경험을 제공할 수 있다.

3.3 분산 학습 기반 애플리케이션

단말의 컴퓨팅 파워 증가와 사용자 프라이버시 우려로 인해 연합 학습과 같은 분산 학습이 6G에서 널리 사용될 것으로 예상된다 [17], [18]. Deep neural network (DNN)은 보통 수백만 개 이상의 가중치 파라미터를 포함하기 때문에, 분산 학습시 단말과 서버 간의 DNN 모델이나 그라디언트의 빈번한 교환이 필요하며 이는 큰 통신 오버헤드를 초래한다. 이때 시맨틱 통신은 그라디언트 희소화(sparsification) [19] 및 모델 파라미터 가지치기(pruning) [20]와 같은 방식을 사용하여 모델의 정확도와 수렴 속도에 중요하고 의미 있는 파라미터만을 추출하여 전송할 수 있다. 또한, 분산 학습시 에이전트들은 학습 모델, 관측 환경, 그들의 임무 등으로부터 시맨틱 특징을 추출하고 이 시맨틱 특징만을 교환하여 서로의 지식을 향상시킬 수 있다 [21]. 아울러, 작은 크기의 시맨틱 특징을 기반으로 협업에 의미가 있는 최적의 에이전트 집합을 찾을 수 있고 이를 통해 대용량 모델의 교환 횟수를 줄일 수 있다.

3.4 지능형 교통 시스템 (ITS)

그림 2의 예와 같이 ITS에서 시맨틱 통신의 가장 직접적인 응용은 차량의 운동 정보, 도로 상태, 교통 표지판 등의 센싱 데이터에서 필수적인 의미 정보를 추출하는 것이다 [22]. 예를 들어, 차량 앞에 사람, 자동차, 포트홀이 나타나거나 충돌이 발생하는 등의 특정 상황에 대한 시맨틱 정보를 추출함으로써, 전송 데이터양이 감소하며 전송의 시기 적절성이 향상된다. 데이터 자체의 압축 외에도, 데이터의 샘플링에도 시맨틱을 적용할 수 있다. 차량이 충분히 긴 가시 거리를 가진 경우에는 관련 정보가 자주 교환될 필요가 없는 반면, 차량의 가시 기간이 짧은 경우 정보 교환이 더 자주 되어야 한다. 이같이 언제, 어떤 정보를 얼마의 크기로 전송해야 할지를 결정하는 것이 바로 시맨틱 통신이 해야 할 일이다 [23].

3.5 협력 로봇

재난 관리, 창고 자동화, 감시와 같은 시나리오에서 협력 로봇의 사용이 증가하고 있으나 각 로봇의 제한된 컴퓨팅 능력은 집중적인 계산이 필요한 작업을 어렵게 한다 [24]. 하지만, 시맨틱 통신은 협력 로봇 간에 효율적인 데이터 교환 및 처리를 가능하게 하여 이와 같은 문제를 해결할 수 있다. 협력 로봇 간 통신 오버헤드를 줄이기 위해, L-DeepSC라는 경량 분산 시맨틱 통신 시스템이 제안되었으며, 이는 로봇 간의 상호 작용에 필요한 데이터양을 기존 방법의 2.5% 수준으로 압축할 수 있음을 보여주었다 [25].

3.6 Wireless Body Area Network (WBAN)

WBAN의 작은 센서 노드는 에너지를 효율적으로 소비해야 한다. 시맨틱 통신은 전송 비트 수를 줄임으로써 에너지를 절약하고 웨어러블 장치의 수명을 늘릴 수 있다. 시맨틱 추출 프로세싱이 추가적인 에너지를 소비하더라도 전송 duty cycle을 늘리는 것이 웨어

러블 장치의 수명 증가에 도움이 된다는 것을 보였으며, 전송 정보량이 줄어들에 따라 분류 정확도가 5% 정도 낮아지지만, 장치의 수명은 한 차원 더 늘릴 수 있음을 확인하였다 [26]. 이러한 시맨틱 통신과 WBAN과의 통합은 6G 네트워크에서 더 오래 지속되고 편리한 웨어러블 장치의 출현을 예상한다.

3.7 Unmanned Air Vehicle (UAV)

시맨틱 통신은 전송해야 할 정보량을 줄여주기 때문에, UAV 간의 효율적인 통신 프레임워크를 구현하는데 사용될 수 있다. 예를 들어, UAV가 중계기로 사용될 때 시맨틱 통신을 접목한 새로운 semantic process-and-forward 방법이 제안될 수 있다 [27]. 중계 기능 외에도, UAV는 시맨틱 인코더 및 디코더로 사용될 수 있다. 즉, 송수신기 중 하나가 메모리나 컴퓨팅 파워 부족으로 시맨틱 통신을 사용할 수 없는 경우, UAV가 인코딩이나 디코딩을 대신 수행하여 특정 링크의 데이터양을 줄이면서도 통신 성능을 유지할 수 있다. 또한, UAV는 송신단의 배경 지식에 따라 수신된 신호를 의미적으로 디코딩한 다음, 수신단의 배경 지식에 따라 신호를 다시 인코딩해줄 수 있다. 이는 송수신단 간의 배경 지식 동기화 오버헤드와 비동기화된 배경 지식으로 인한 의미적 노이즈를 크게 줄인다 [27].

3.8 Hyper-Intelligent (HI) IoT

6G에서 예상되는 트렌드 중 하나는 네트워크 노드 및 연결된 장치의 지능이 향상되는 것이다. 단순히 감지 및 전송 기능을 수행하는 IoT 장치가 복잡한 작업을 수행할 수 있는 HI를 갖게 되면 ‘더 스마트한’ 콘텐츠, 즉 시맨틱 정보를 생성할 수 있게 된다 [28]. 마찬가지로 시맨틱 통신 기술을 활용하여 유용한 정보만 추출하고 전송하면, 통신 자원을 절약할 수 있으며 수신자가 주어진 작업을 효과적으로 수행할 수 있다. HI IoT 장치에 의미적 및 효과적 수준에서 정보를 처리하고 추론하는 능력을 부여함으로써 6G 세상에서는 더 연결된 사회가 구축될 것이다.

4. 시맨틱 정보 전송

본 장에서는 그림 4에 도식된 바와 같이 무선 환경, 제한된 네트워크 자원, 이질적인 네트워크를 고려한 시맨틱 정보 전송과 관련된 기술과 이슈를 제시한다.

4.1 무선 환경

시맨틱 통신 시스템은 AI의 도움으로 무선 환경을

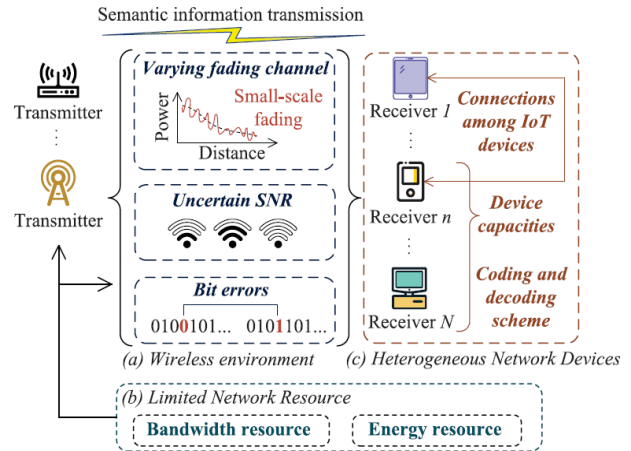


그림 4 무선 환경, 제한된 네트워크 자원, 이질적 네트워크를 고려한 시맨틱 정보 전송 개념도 [4]

고려하여 소스와 채널 코딩을 더 밀접하게 결합할 수 있다. 그러나 AI 기반 방식은 명시적으로 설명하기 어렵기 때문에, 변화하는 페이딩 채널, 불확실한 signal-to-noise ratio (SNR) 및 비트 오류가 시맨틱 통신 성능에 미치는 영향을 충분히 조사해야 한다.

4.1.1 페이딩 채널

전통적인 무선 통신에서는 Rician, Rayleigh, Gaussian 채널 모델과 같은 여러 고전적인 채널 모델이 성능 분석에 사용되며, 이는 무선 환경의 다양한 조건을 나타내는 여러 매개변수를 가지고 있다. 그러나 종단간 구조를 가진 시맨틱 통신 시스템에서 채널 레이어는 대개 고정 채널 레이어와 생성 채널 레이어 방식으로 모델링된다.

고정 채널 레이어 모델링은 채널 레이어를 훈련 과정 전체에서 사용되는 고정 페이딩 모델로 모델링한다 [29]. 예를 들어, eraser 채널은 인코더에서 나온 이진화된 비트 벡터를 입력으로 받아 드롭 확률에 따라 출력 벡터의 요소를 {-1, 0, 1}로 변환한다. 이는 과적합 문제를 방지하기 위한 드롭아웃 기법과 유사하다. 이러한 고정 채널 모델은 무선 환경의 변화에 적응하지 못하므로 채널 조건이 변경되면 모델을 이에 맞게 다시 훈련해야 하는 오버헤드가 발생한다.

생성 채널 레이어 모델링에서는 생성적 적대 신경망(generative adversarial network, GAN)을 사용하여 채널을 모델링한다 [30]. 조건부 GAN은 송신기의 파일럿 정보와 인코딩된 신호를 받아 실제 데이터와 유사한 출력 신호를 생성한다. 이 방식은 다양한 채널 상태를 동적으로 반영할 수 있지만, 훈련시 더 높은 복잡도를 요구한다.

고정과 생성 채널 모델링의 장점을 결합하여, 1단계에서는 적절한 채널 모델로 훈련하여 합리적인 정확도에서 모델 파라미터를 얻고, 2단계에서 실제 채널에서 수신기를 미세 조정하는 하이브리드 방안이 사용될 수 있다 [31]. 이처럼, 다양한 채널 모델링 방식은 시맨틱 통신 성능에 영향을 미치며, 채널 모델링을 위한 최적의 훈련 방법은 여전히 연구되어야 한다.

4.1.2 불확실한 SNR

SNR의 불확실성은 잡음과 간섭의 영향, 그리고 전송 전력의 변화에서 비롯된다. 일반적으로 시맨틱 통신 모델은 고정 SNR에서 훈련되므로 SNR의 변화가 성능에 어떤 영향을 미칠지 고려해야 한다. 특히, 낮은 SNR 영역에서 모델의 강인성을 높이기 위해 SNR 적응 메커니즘이 제안되었다 [32]. 이 방식은 수신기에서 파일럿 신호로 SNR을 추정하고, 이를 채널 출력 피쳐 맵과 합산하여 노이즈 제거 모듈의 입력으로 사용한다. 이를 통해 높은 SNR 영역과 낮은 SNR 영역 간의 peak SNR 차이를 줄여 SNR에 대한 높은 적응성을 갖는다.

SNR 값을 채널 피쳐에 추가하는 대신, [33]에서는 채널별 소프트 어텐션(attention)을 사용하여 SNR 값에 따라 채널 피쳐를 스케일링하는 방법을 제안했다. 이 방법은 SNR 값과 컨텍스트 정보를 두 개의 완전 연결 레이어에 전달하여 각 피쳐 채널의 스케일링 요소를 얻는다. 이러한 솔루션들은 특정 통신 문제 해결을 위해 설계되었기 때문에, 일반화된 시맨틱 통신 시스템의 경우 훈련된 시맨틱 모델이 가변 SNR에 적응할 수 있도록 하는 더 나은 방법이 요구된다.

4.1.3 비트 오류

시맨틱 정보의 전송 오류를 최소화하기 위해 시맨틱 통신에 맞는 오류 정정 알고리즘의 적용이 필요하다. 시맨틱 텍스트의 전송 오류를 줄이기 위해 hybrid automatic repeat request (HARQ)의 사용이 고려되어 소스-채널 코딩과 HARQ를 공동 설계할 경우 오류율을 현저하게 낮출 수 있음을 보였다 [34]. 이처럼, 기존 오류 정정 메커니즘과 시맨틱 통신을 결합한 새로운 재전송 방식의 설계는 매우 중요하다.

4.2 제한된 네트워크 자원

시맨틱 통신은 비트 흐름 뒤에 있는 정보의 의미를 더 가치 있게 여기기 때문에, 이를 고려한 시맨틱 통신 전용 자원 할당 프레임워크가 필요하다. 여기에서는 대역폭 및 에너지 측면에서 시맨틱 통신 자원의 할당 방법을 알아본다.

4.2.1 대역폭 자원

한정된 주파수 자원을 고려하여 더 많은 시맨틱 정보를 가진 에이전트에게 대역폭을 더 많이 할당해야 한다. 유망한 접근 방식 중 하나는 훈련 과정에서 대역폭 할당을 같이 수행하는 것으로 [35]에서는 여러 에이전트가 무선 네트워크를 통해 정책을 공유하고 각 작업에 대해 최적의 정책을 공동으로 학습하는 협력적 심층 강화학습 알고리즘을 제안했다. 학습을 통해 에이전트 간의 의미적 관련성이 정의되고, 이 값이 의미적 관련성이 높은 에이전트를 선택하는 데 사용된다. 이 의미적 관련성이 높은 에이전트에게 대역폭을 많이 할당함으로써 시스템 성능을 향상시킬 수 있다. 같은 맥락으로 시맨틱 통신 시스템은 더 중요한 시맨틱 정보 전송에 더 많은 대역폭을 할당하여 전송 콘텐츠의 품질을 보장할 수 있다.

4.2.2 에너지 자원

에너지 수확 기능을 가진 IoT 장치가 증가함에 따라, 더 많은 시맨틱 정보를 갖는 데이터를 전송하는데 더 많은 에너지를 할당하면 에너지를 효율적으로 사용할 수 있다. [36]에서는 IoT 장치가 수확한 에너지 가치를 도출하기 위해 문장 유사성 메트릭 기반의 평가 함수를 사용하였다. 제안 시스템에서 IoT 장치는 hybrid access point (HAP)로 부터 무선 에너지를 얻기 위해 입찰을 제출하고 HAP는 승자와 지불을 결정한다. 이 경매 메커니즘에서 IoT 장치는 기존 통신 성능 지표 대신 bilingual evaluation understudy (BLEU) 점수와 유사성 점수를 기반으로 에너지에 대한 실제 평가 점수를 구하고 이를 입찰에 활용한다. 하지만, 현재 에너지 자원 할당에 시맨틱 메트릭의 적용은 아직 초기 단계로, 에너지 수확이 가능한 시맨틱 통신 네트워크는 앞으로 연구가 필요한 분야이다.

4.3 이질적인 네트워크 장치

시맨틱 통신 네트워크는 많은 이질적인 장치들로 구성되므로 장치들의 각기 다른 능력과 사용하는 무선 링크의 이질성을 고려해야 한다.

4.3.1 장치 능력

시맨틱 통신 시스템은 송수신기에 각각 시맨틱 인코더와 디코더를 설치해야 하므로 구현 비용과 함께 이를 훈련 시키기 위한 연산 및 통신 자원이 필요하다. 또한, 성능 향상을 위해 모델을 확장하면 더 많은 모델 파라미터를 저장하기 위한 큰 메모리가 요구된다. 현실적으로, 통신 장치는 제한된 계산 능력, 통신

자원, 저장 용량을 가지고 있기 때문에, 시맨틱 통신 네트워크에서는 성능과 비용 간에 균형이 필요하다. 시맨틱 모델의 크기를 줄이기 위한 대표적인 방식으로 모델 가지치기와 양자화를 꼽을 수 있다. 주목할 만하게, [37]에서는 60%의 희소성 비율로 가지치기 되고 8비트 정수로 양자화된 압축된 시맨틱 인코더/디코더를 사용하더라도 비압축 모델과 유사한 BLEU 점수를 달성할 수 있음을 보였다. 그러나 시맨틱 모델의 추가 압축으로 인한 성능 손실은 체계적으로 조사될 필요가 있으며, 네트워크 내 다양한 장치의 능력과 성능 요구를 고려해야 한다.

4.3.2 장치 간 연결 링크

시맨틱 통신 네트워크는 다양한 종류의 무선 링크를 고려해야 한다. 한 가지 해결책은 훈련 과정에서 무선 링크를 고려하는 것으로, [38]에서는 차량 네트워크 스펙트럼 다중화 시나리오에서 vehicle-to-infra (V2I) 링크와 vehicle-to-vehicle (V2V) 링크를 고려하여 시맨틱 비디오 전송을 위한 자원 할당 알고리즘을 제안하였다. 제안 알고리즘은 다중 에이전트 deep Q-network를 사용하여 비디오 전송의 의미 이해 정확도를 최대화하는 방향으로 에이전트가 스펙트럼 자원 블록을 재사용할지 결정하고, 이후 에이전트는 V2I 평균 객체 탐지 정확도와 V2V 평균 전송 속도를 기준으로 보상을 받는 방식으로 훈련된다.

4.3.3 인코딩 및 디코딩 방식

시맨틱 통신 네트워크에서 사용자별 다양한 채널 조건에 따라 인코딩 및 디코딩 방식이 변경될 필요가 있다. 이를 위해 [39]에서는 다양한 채널 상태를 고려한 자가 지도 메커니즘을 제안했다. 여기에서 전송 메시지는 기준이 충족될 때까지 여러 번 인코딩되고 디코딩될 수 있다. 이 방식에서 각 인코딩/디코딩 사이클마다 인코딩/디코딩된 정보는 신뢰 메커니즘에 의해 의미 신뢰도가 결정되고 임계값에 도달해야 인코딩/디코딩을 해제한다.

또한, 다중 사용자가 존재하는 시나리오에서는 가용 스펙트럼 및 전송 전력과 같은 자원의 변동이 시맨틱 통신 성능에 영향을 미칠 수 있다. 이러한 동적으로 변하는 자원에 대처하기 위해 기존 가변 비디오 코딩과 같은 가변 길이 시맨틱 인코딩 기술이 필요하다.

5. 시맨틱 통신 프로토콜 구조

시맨틱 통신 아키텍처에서는 시맨틱 정보의 생성 및 전송을 지원하는 지능적이고 효율적인 프로토콜

구조가 필요하다. 본 장에서는 [40]에서 제시된 그림 5의 시맨틱 통신을 위한 프로토콜 구조를 소개한다. 시맨틱 통신 프로토콜은 semantic application-intent (S-AI) 계층, semantic network-protocol (S-NP) 계층, semantic physical-bearing (S-PB) 계층으로 구성되며, 이 세 계층과 물리적 환경이 별도의 semantic intelligent plane (S-IP)을 두고 semantic information flow (S-IF)를 통해 연동된다.

5.1 Semantic Intelligent Plane (S-IP)

전체 시맨틱 통신 네트워크의 조정자로서 S-IP는 다음 세 가지 기능을 수행한다.

- 1) 시맨틱 환경 표현: S-AI 계층에서 제공하는 시맨틱 의도 정보를 기반으로 원본 데이터와 통신 컨텍스트에서 시맨틱 정보를 추출한다. 이 시맨틱 정보는 S-IF에 삽입되어 각 계층의 인터페이스를 통해 각 계층의 기능에 전달된다.
- 2) 배경 지식 관리: 시맨틱 통신하는 양쪽이 동일한 배경 지식을 갖도록 S-AI 계층의 의도 분석의 기초를 마련하고, 송수신 노드간 동기화 및 배경 지식 저장을 수행한다.
- 3) 의사 결정 및 추론: S-AI 계층에서 제공된 의도 분석 결과에 따라 달성 가능한 성능을 기반으로 모든 계층에 대한 의사 결정을 한다.

5.2 Semantic Application-Intent (S-AI) 계층

S-AI 계층은 기존 애플리케이션 계층의 역할 외에 특정 통신에 대한 의도 분석 기능을 수행한다. 또한, 분석된 사용자 의도를 네트워크 배치, 설정 또는 제어 정책에 반영하는 기능을 지원한다. S-AI 계층의 세 가지 주요 기능은 아래와 같다.

- 1) 의도 마이닝: 사용자 또는 애플리케이션의 원래 의도를 추출, 분석, 집계, 합성하는 책임을 진다. 이미지의 경우 이러한 의도는 타겟 식별이나 다른 세부 사항 확인이 될 수 있으며, 또는 이미지 품질 메트릭으로만 결정될 수도 있다.
- 2) 의도 분해: 시맨틱 인코딩을 용이하게 하기 위해 의도 마이닝 후 얻어진 의도를 하위 의도 집합으로 분해할 수 있다. 예를 들어, ITS에서 차량이 벽으로 둘러싸인 구역을 통과하기 전에 출구의 교통 흐름 정보를 제공하는 것이 의도일 때, 하위 의도는 출구의 연속된 뷰가 될 수 있다.
- 3) 의도 표현: 얻은 하위 의도 집합을 표현한다. 이 의도 표현은 시맨틱 인코딩을 위해 S-IP에 전달된다.

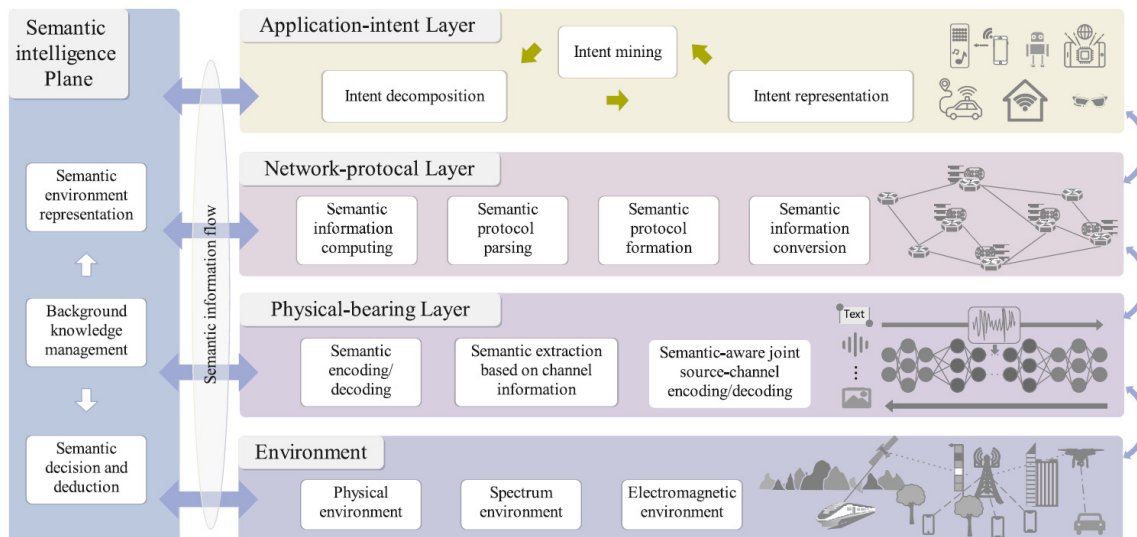


그림 5 시맨틱 통신을 위한 프로토콜 구조 [40]

5.3 Semantic Network-Protocol (S-NP) 계층

S-NP 계층은 상위 애플리케이션의 의도를 효율적으로 서비스하는 것을 목표로 다음과 같은 주요 기능을 수행한다.

- 1) 시맨틱 정보 계산: 시맨틱 정보를 전달하는 중간 노드가 S-IF로부터 의도 정보를 확인하고 관련 지식을 얻는 기능을 수행한다.
- 2) 시맨틱 프로토콜 분석: 기존 프로토콜의 사용 가능한 기능을 분석한다.
- 3) 시맨틱 프로토콜 형성: 애플리케이션의 의도에 맞추기 위해 기존 프로토콜을 최적화하거나 새로운 프로토콜을 형성한다.
- 4) 시맨틱 정보 변환: 형성된 프로토콜을 기반으로 시맨틱 정보를 캡슐화하는 역할을 한다.

물입형 AR 애플리케이션을 예로 들면, S-NP 계층은 S-IP와 상호 작용하여 사운드에 대한 attention, 이미지에 대한 객체, 그리고 촉각 피드백 등을 S-IF 내의 데이터에 매핑한다. 이후 attention이 작은 데이터는 제거하여 전송 자원을 확보하고 촉각 패킷과 같은 지연 민감 트래픽에 높은 우선순위를 부여한다 [41].

5.4 Semantic Physical-Bearing (S-PB) 계층

S-PB 계층은 상위 계층으로부터 받은 시맨틱 정보를 물리 신호로 변환하는 역할을 한다. 전통적인 소스 및 채널 인코딩/디코딩이 대신 S-PB 계층은 다음 세 가지 기능을 수행한다.

- 1) 시맨틱 인코딩/디코딩: 구축된 지식 정보를 기반으로 시맨틱 인코더 및 디코더 쌍을 운용한다.
- 2) 채널 정보 기반 시맨틱 추출: 페이딩, SNR 및 간

섭과 같은 채널 정보를 시맨틱 인코딩 및 디코딩에 통합하여 채널 변화에 적응적으로 동작시킨다.

- 3) 시맨틱 인식 통합 소스-채널 인코딩/디코딩: 채널 인코딩/디코딩을 시맨틱 인코딩/디코딩 모듈에 통합하여 수행한다.

6. 도전과제

6.1 무선 환경에서의 시맨틱 통신

시맨틱 통신에서는 종단간 시맨틱 모델 훈련 시 페이딩 채널을 반영하기 위해 채널 레이어를 사용해야 한다. 이때, 채널 레이어 모델의 훈련에 더 많은 자유도를 제공하기 위해 생성 채널 레이어의 도입이 고려되어야 한다. 또한, 시맨틱 통신의 목적은 작업과 관련된 시맨틱 정보를 정확하고 효율적으로 전송하는 것이다. 제한된 무선 자원을 고려할 때, 작업 요구사항과 가용 무선 자원의 통합 최적화가 필요하다. 아울러, 단말의 용량 및 통신 링크의 이질성을 고려하여 시맨틱 모델의 배포 및 코딩 스킴을 개발해야 한다.

6.2 시맨틱 인코딩 정확도와 통신 오버헤드 간의 균형

기존 연구들은 주로 무선 자원을 절약하고 통신 성능을 높이기 위해 정확한 시맨틱 추출에 중점을 두고 있지만, 이를 위해 필요한 통신 오버헤드는 무시하고 있다. 시맨틱 모델의 학습 및 업데이트에는 상당한 자원이 필요하며, 각 사용자의 로컬 지식 베이스는 지속적으로 업데이트되어야 한다. 특히, 지리적으로 떨어진 참가자가 많을수록 통신 오버헤드는 매우 커지기

때문에 시맨틱 추출 정확도와 통신 오버헤드 간의 균형을 맞추는 것이 중요하다.

연합 학습을 통해 지리적 영역별로 학습된 여러 시맨틱 모델을 집계하여 서버에서 일반화된 시맨틱 인코딩 모델을 얻을 수 있다. 이는 저장 자원을 효율적으로 활용하고 통신 오버헤드를 줄일 수 있게 해준다. 그러나 지리적 영역의 합리적인 분할, 서버의 전략적 배치, 집계 주기와 참가자의 선택은 여전히 해결해야 할 과제이다.

6.3 시맨틱 통신 성능과 보안 간의 균형

데이터 보안과 개인 정보 보호는 무선 통신 분야에서 항상 중요한 주제로 시맨틱 통신은 데이터의 일부만 전송하고 시맨틱 정보의 디코딩이 수신자의 배경 지식에 의존하기 때문에 보안에 유리하다 [42]. 더불어 추출된 시맨틱 정보를 암호화하여 보안을 강화할 수 있다. 그러나 이는 계산 오버헤드와 데이터 보안 간의 tradeoff를 야기한다. 물리 계층 보안 기술을 사용하여 간섭을 도입하면 비밀 통신을 보장할 수 있으며, 전송 전력을 적절히 조절하여 데이터 도청을 방지할 수 있다. 그러나 간섭 신호는 시맨틱 정보 전송에 부정적인 영향을 미칠 수 있으므로 비밀성과 신호 품질 간의 균형점이 필요하다.

6.4 시맨틱 인코딩의 해석 및 설명 가능성

시맨틱 통신 환경은 네트워크 환경 변화나 새로운 소스 정보 등의 다양한 불확실성을 내포하고 있다. 이로 인한 시맨틱 인코딩 모델의 블랙박스 특성은 예측 가능성을 낮춰 실용성을 제한하고 최적화를 어렵게 한다. 또한, 본질적으로 시맨틱 인코딩 모델은 내부 상태와 기능이 특정 예제에 어떻게 작용하는지 이해하기 어렵기 때문에 SE의 해석 가능성과 설명 가능성 문제가 발생한다 [43].

해석 가능성은 인간이 모델의 결정을 일관되게 예측할 수 있는 정도를 나타낸다. 해석 가능성이 커지면 시맨틱 인코딩 모델이 특정 결정이나 결과에 도달하는 이유와 방법을 알 수 있어 모델의 전반적인 강점과 약점을 이해하고 모델을 개선할 수 있다 [44]. 반면, 설명 가능한 AI는 DNN의 숨겨진 상태를 분석하여 블랙박스를 개방하는 데 중점을 둔다. 예를 들어, 시맨틱 디코더의 기울기 정보를 분석하여 입력되는 시맨틱 특징이 추론 정확성에 기여하는 정도를 정량화할 수 있으며, 이를 통해 중요한 시맨틱 특징에 전송 우선순위를 부여하여 전송 신뢰성과 추론 정확성을 높일 수 있다.

7. 결 론

본 고에서는 6G 이동통신 시스템과 관련된 시맨틱 통신 기술에 중점을 두고 전반적인 조사를 수행하였다. 먼저 시맨틱 통신의 기본 개념과 시스템 모델을 소개하고, 6G에서의 시맨틱 통신의 잠재적인 응용 사례를 살펴보았다. 또한, 시맨틱 정보의 전송 기술 및 프로토콜 구조를 제시하고, 6G에서 시맨틱 통신의 실현을 위한 도전과제를 논의하였다. 시맨틱 통신 연구와 6G의 개발은 상호 보완하면서 발전하여 향후 사용자에게 다양한 서비스와 경험을 제공할 수 있을 것으로 전망된다.

참고문헌

- [1] M. Xu et al., "A Full Dive Into Realizing the Edge-Enabled Metaverse: Visions, Enabling Technologies, and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 656-700, Firstquarter 2023.
- [2] H. Du, B. Ma, D. Niyato, J. Kang, Z. Xiong, and Z. Yang, "Rethinking Quality of Experience for Metaverse Services: A Consumer-Based Economics Perspective," *IEEE Netw.*, vol. 37, no. 6, pp. 255-263, Nov. 2023.
- [3] P. Zhang et al., "Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks," *Engineering*, vol. 8, pp. 60-73, Jan. 2022.
- [4] W. Yang et al., "Semantic Communications for Future Internet: Fundamentals, Applications, and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213-250, Firstquarter 2023.
- [5] M. Thoma, "A survey of semantic segmentation," 2016, arXiv:1602.06541.
- [6] P. C.-Y. Sheu, *Semantic Computing*. Boston, MA, USA: Springer, 2010, pp. 1-9.
- [7] P. Hitzler, "A review of the semantic Web field," *Commun. ACM*, vol. 64, no. 2, pp. 76-83, 2021.
- [8] C. She et al., "Deep learning for ultra-reliable and low-latency communications in 6G networks," *IEEE Netw.*, vol. 34, no. 5, pp. 219-225, Sep./Oct. 2020.
- [9] E. C. Strinati and S. Barbarossa, "6G networks: Beyond Shannon towards semantic and goal-oriented communications," *Comput. Netw.*, vol. 190, May 2021, Art. no. 107930.
- [10] F. I. Dretske, *Knowledge and The Flow of Information*. Cambridge, MA, USA: MIT Press, 1981.

-
- [11] Q. Zhou, R. Li, Z. Zhao, C. Peng, and H. Zhang, "Semantic communication with adaptive universal transformer," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 453-457, Mar. 2022.
- [12] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked Autoencoders are Scalable Vision Learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16000-16009.
- [13] H. Du et al., "Exploring Attention-Aware Network Resource Allocation for Customized Metaverse Services," *IEEE Netw.*, vol. 37, no. 6, pp. 166-175, Nov. 2023.
- [14] Y.-J. Liu, H. Du, D. Niyato, G. Feng, J. Kang, and Z. Xiong, "Slicing4Meta: An intelligent integration framework with multidimensional network resources for Metaverse-as-a-service in Web 3.0," 2022, arXiv:2208.06081.
- [15] L. Ismail, D. Niyato, S. Sun, D. I. Kim, M. Erol-Kantarci, and C. Miao, "Semantic Information Market for the Metaverse: An Auction Based Approach," in *Proc. IEEE Future Netw. World Forum (FNWF)*, Montreal, QC, Canada, 2022, pp. 628-633.
- [16] C. De Alwis et al., "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 836-886, 2021.
- [17] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Distributed federated learning for ultra-reliable low-latency vehicular communications," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1146-1159, Feb. 2020.
- [18] W. Zhang et al., "Optimizing federated learning in distributed industrial IoT: A multi-agent approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3688-3703, Dec. 2021.
- [19] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546-3557, 2020.
- [20] Y. Jiang et al., "Model Pruning Enables Efficient Federated Learning on Edge Devices," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10374-10386, Dec. 2023.
- [21] F. Lotfi, O. Semiari, and W. Saad, "Semantic-Aware Collaborative Deep Reinforcement Learning Over Wireless Cellular Networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 5256-5261.
- [22] H. Bista, I.-L. Yen, F. Bastani, M. Mueller, and D. Moore, "Semantic-based information sharing in vehicular networks," in *Proc. IEEE Int. Conf. Web Services (ICWS)*, 2018, pp. 282-289.
- [23] A. Nanda, D. Puthal, J. J. Rodrigues, and S. A. Kozlov, "Internet of autonomous vehicles communications security: Overview, issues, and directions," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 60-65, Aug. 2019.
- [24] Y. Yue, C. Zhao, Z. Wu, C. Yang, Y. Wang, and D. Wang, "Collaborative semantic understanding and mapping framework for autonomous systems," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 2, pp. 978-989, Apr. 2021.
- [25] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142-153, Jan. 2021.
- [26] P. Zalewski, L. Marchegiani, A. Elsts, R. Piechocki, I. Craddock, and X. Fafoutis, "From bits of data to bits of knowledge – An on-board classification framework for wearable sensing systems," *Sensors*, vol. 20, no. 6, p. 1655, 2020.
- [27] X. Luo, B. Yin, Z. Chen, B. Xia, and J. Wang, "Autoencoder-based Semantic Communication Systems with Relay Channels," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2022, pp. 711-716.
- [28] P. Popovski et al., "Internet of Things (IoT) connectivity in 6G: An interplay of time, space, intelligence, and value," 2021, arXiv:2111.05811.
- [29] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2018, pp. 2326-2330.
- [30] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional GANs as unknown channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3133-3143, May 2020.
- [31] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 132-143, Feb. 2018.
- [32] M. Ding, J. Li, M. Ma, and X. Fan, "SNR-adaptive deep joint source-channel coding for wireless image transmission," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2021, pp. 1555-1559.
- [33] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Trans. Circuits*
-

Syst. Video Technol., vol. 32, no. 4, pp. 2315-2328, Apr. 2022.

- [34] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep Source-Channel Coding for Sentence Semantic Transmission With HARQ," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5225-5240, Aug. 2022.
- [35] F. Lotfi, O. Semiari, and W. Saad, "Semantic-Aware Collaborative Deep Reinforcement Learning Over Wireless Cellular Networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2022, pp. 5256-5261.
- [36] Z. Q. Liew, Y. Cheng, W. Y. B. Lim, D. Niyato, C. Miao, and S. Sun, "Economics of Semantic Communication System in Wireless Powered Internet of Things," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2022, pp. 8637-8641.
- [37] H. Xie and Z. Qin, "A lite distributed semantic communication system for Internet of Things," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 142-153, Jan. 2021.
- [38] M. Zhu, C. Feng, J. Chen, C. Guo, and X. Gao, "Video semantics based resource allocation algorithm for spectrum multiplexing scenarios in vehicular networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, 2021, pp. 31-36.
- [39] K. Lu et al., "Rethinking Modern Communication from Semantic Coding to Semantic Communication," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 158-164, Feb. 2023.
- [40] P. Zhang et al., "Toward wisdom-evolutionary and primitiveconcise 6G: A new paradigm of semantic

communication networks," *Engineering*, vol. 8, pp. 60-73, Jan. 2022.

- [41] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-effectiveness filtering and control for post-5G wireless connectivity," *J. Indian Inst. Sci.*, vol. 100, no. 2, pp. 435-443, Feb. 2020.
- [42] P. Basu, J. Bao, M. Dean, and J. Hendler, "Preserving quality of information by using semantic relationships," *Pervasive Mobile Comput.*, vol. 11, pp. 188-202, Apr. 2014.
- [43] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digit. Signal Process.*, vol. 73, pp. 1-15, Feb. 2018.
- [44] Y. Dong, H. Su, J. Zhu, and B. Zhang, "Improving interpretability of deep neural networks with semantic information," in *Proc. IEEE Comput. Soc. Conf. (CVPR)*, 2017, pp. 4306-4314.

약 력



최 현 호

2001.2 KAIST 전기 및 전자공학과 졸업
2003.2 KAIST 전기 및 전자공학과 석사
2007.2 KAIST 전기 및 전자공학과 박사
2007.3~2011.2 삼성종합기술원 전문연구원
2011.3~현재 한경국립대학교 ICT로봇기계공학부 교수
관심분야 : 이동통신시스템, 시맨틱 통신, 머신러닝 및 빅데이터, 분산 네트워크, 저전력 통신, 무선전력전송, 생체모방 알고리즘
Email : hhchoi@hknu.ac.kr