

대형 언어 모델을 활용한 시맨틱 통신 연구 동향

이서빈, 김용준*
포항공과대학교

{lsb0918, yongjune}@postech.ac.kr

A Survey on Semantic Communications with Large Language Models

Seobin Lee, Yongjune Kim*

Pohang University of Science and Technology (POSTECH)

요약

시맨틱 통신은 전송되는 데이터의 의미를 고려하여 효율성과 정확성을 함께 확보하려는 새로운 통신 패러다임이다. 최근에는 대형 언어 모델(large language models, LLM)의 문맥 이해 및 조건부 확률 추론 능력을 활용하여, 시맨틱 구조에 기반한 부호화 및 복호화를 구현하려는 연구들이 제시되고 있다. 이러한 시스템들은 시맨틱 오류 정정, 중요도 기반 자원 할당, 의미 복원 품질 향상 등 다양한 측면에서 진전을 보여주고 있다. 본 논문은 LLM 기반 시맨틱 통신 연구들을 종합적으로 소개하고 향후 연구 방향을 제안하고자 한다.

I. 서론

시맨틱 통신은 단순한 비트 정확도를 넘어, 전송된 정보가 수신자에게 의미적으로 올바르게 해석될 수 있도록 하는 것을 목표로 한다. 이를 위해서는 통신 시스템이 메시지의 의미를 이해하고, 그 의미를 중심으로 소스 코딩 방법과 채널 코딩 방법을 조정할 수 있어야 한다. 그러나 시맨틱 구조는 일반적으로 고차원적이며 복잡한 문맥을 포함하기 때문에, 기존의 기계학습 기반 통신 시스템에서는 데이터의 의미를 온전하게 추출하는데 큰 어려움이 존재했다.

이러한 배경 속에서, LLM의 등장은 시맨틱 통신 분야에 새로운 전환점을 제시하고 있다. LLM은 대규모 텍스트 데이터를 통해 학습된 사전 지식과 강력한 문맥 추론 능력을 바탕으로, 의미 기반 정보 표현과 예측을 가능하게 한다. 특히, 조건부 확률 분포 추정, 문장 수준의 시맨틱 유사도 계산, 시맨틱 오류 복원, 다중 모달 변환 등은 기존 모델로는 어려웠던 기능들로, LLM을 통해 보다 정교하고 실용적인 시맨틱 통신 시스템을 설계할 수 있는 기반이 마련되었다. 본 논문에서는 이러한 문제의식을 바탕으로, LLM을 중심으로 구성된 최근 시맨틱 통신 연구들을 정리하고, 미래 연구 방향을 논의하고자 한다.

II. 본론

[1]에서는 LLM이 문맥을 기반으로 추론한 조건부 확률 분포를 사용하여 텍스트 데이터에 산술 부호화(arithmetic coding)를 적용하였다. 채널 부호로는 저밀도 패리티 체크(low-density parity-check, LDPC) 부호를 사용하였으며, 복호 성능을 높이기 위해

트랜스포머 기반 복호기인 트랜스포머 기반 오류 정정 부호기(error correction code transformer, ECCT)를 활용하여 오류 패턴을 복원하였다.

[1]은 소스와 채널을 분리해서 최적화하는 것이 기존에 제안되었던 소스-채널 부호화(joint source-channel coding, JSCC) 기법들보다 나을 수 있다는 가능성을 보여주었다. 하지만 [1]의 방법론은 텍스트 데이터만 다루고 있다는 한계가 존재한다.

[2]에서는 이미지 데이터를 LLM 친화적인 데이터로 전환하기 위해, 송신단에 이미지 캡션 모델(image caption neural network, ICNN)을 도입하고 수신단에 텍스트 기반 이미지 생성 모델을 도입하였다. 해당 논문에서 제시한 구체적인 전송 프레임워크는 다음과 같다.

송신단에서는 입력 이미지를 ICNN을 통해 자연어 문장으로 변환한다. 이 문장은 L 개의 프레임으로 나뉘고 각각 이진 매핑 및 채널 부호화를 거치게 되며, 이렇게 생성된 이진 시퀀스 $\{b_1, b_2, \dots, b_L\}$ 가 최종적으로 전송된다.

전송 과정에서는 무선 채널의 특성상 일부 프레임에 오류가 발생할 수 있다. [2]는 이러한 오류를 식별하기 위해 순서 번호(sequence number) 기반의 오류 검출 기법을 사용할 것을 제안한다. 이 기법은 각 프레임에 고유한 순서 번호를 부여함으로써, 수신 측이 프레임의 순서를 추적하고 누락 여부를 확인할 수 있도록 한다. 누락된 프레임이 있는 것으로 확인되면, 수신단은 오류 정정 모드로 전환되어 각 프레임에 포함된 체크섬(checksum)을 이용해 프레임의 무결성을 검증한다. 만약 체크섬 검증에 실패하는 경우 해당 프레임 또한 손실된 것으로 간주한다.

수신단에서는 이러한 손실을 보완하기 위해 누락된 프레임의 위치를 마스크 처리한 문장을 언어 모델인 BERT 모델에 입력하여 누락된 단어의 원래 모습을 추론한다. 이 과정은 프레임 단위로 이루어지기 때문에 시맨틱 중요도는 프레임 손실에 따른 시맨틱 손실 형태로 정의된다. [2]는 원본 문장 \mathbf{m}_0 와 특정 프레임 l 이 제거된 상태에서 복원된 문장 \mathbf{m}_l 간의 의미적 유사도를 BERT 모델 기반 코사인 유사도 $\phi(\mathbf{m}_0, \mathbf{m}_l)$ 를 통해 계산하고, 이를 기반으로 중요도를 다음과 같이 정의하였다.

$$F_l = 1 - \phi(\mathbf{m}_0, \mathbf{m}_l)$$

이때, $\phi(\mathbf{m}_0, \mathbf{m}_l) = \frac{B_\psi(\mathbf{m}_0)^T B_\psi(\mathbf{m}_l)}{\|B_\psi(\mathbf{m}_0)\| \|B_\psi(\mathbf{m}_l)\|}$ 이고 $B_\psi(\cdot)$ 은 BERT 모델이다.

더 낮은 오더의 변조 방식이나 더 강한 채널 코딩 방식을 사용하면 프레임이 손실될 확률이 줄어드나 그만큼 시스템의 지연이 커진다. 따라서 시맨틱 손실을 최소화하면서도 시스템이 전송 지연 한계 내에서 운영될 수 있도록, 다음과 같은 최적화 문제가 제시되었다.

$$\min \sum_{l=1}^L F_l \cdot P_{Loss,l}^{frame} \quad \text{s.t.} \quad \sum_{l=1}^L T_l \leq T_{th}, \mathcal{M}_l \in \mathcal{M}_{all}, \mathcal{C}_l \in \mathcal{C}_{all}$$

여기서 $P_{Loss,l}^{frame}$ 은 l 번째 프레임의 손실 확률, T_l 은 l 번째 프레임의 전송 지연, T_{th} 은 총 지연 허용 한계이다. \mathcal{M}_l , \mathcal{C}_l , \mathcal{M}_{all} , \mathcal{C}_{all} 은 각각 l 번째 프레임에서 사용된 변조 방식, 채널 코딩 방식, 그리고 선택 가능한 변조 방식 집합, 채널 코딩 방식 집합을 뜻한다. [2]는 위 문제를 풀기 위해 그리디 탐색 알고리즘(greedy search)을 선택하였다.

[3]은 [2]와 동일하게 목표 이미지를 직접 전송하는 대신, 그로부터 추출된 캡션을 추출하여 전송하는 프레임워크를 제안하고 있다. 다만 차이점은 [2]와 달리 [3]은 특별히 훈련 데이터 분포를 벗어난 입력(out-of-distribution, OOD)에 대한 문제를 다루고 있고, 이를 해결하기 위해 여러 종류의 모델을 종합적으로 활용하는 방식을 제안한다는 것이다.

이 시스템의 송신단은 YOLO 모델과 BLIP 모델을 함께 사용하는 구조를 채택하고 있다. YOLO 모델은 객체 감지 전문 모델로서 높은 재현율(recall)을 보장하는 반면, BLIP 모델과 같은 다중 모달 대형 언어 모델은 정밀도(precision) 측면에서 강점이 있다. 따라서 이 두 모델은 서로의 단점을 보완하는 역할을 할 수 있다.

YOLO 모델이 감지한 객체 중에서 분류에 대한 확신이 낮은 경우가 있을 수 있는데, 이러한 객체에 대해서는 해당 이미지 영역만을 BLIP 모델에 입력하여 “이 그림에는 무엇이 있습니까?”와 같은 질문을 던지고 그 답을 받는다.

하지만 BLIP 모델은 분류 작업에 직접 필요하지 않은 불필요한 단어들을 포함한 문장을 생성할 수 있다. 이러한 비효율을 해결하기 위해 논문에서는 CET(cross-entropy transformation) 기법을 사용한다. CET는 사전에 정의된 분류 레이블의 수에 따라 BLIP 모델의 출력 확률 분포를 조정하여, 분류 작업과 관련이 없는 표현들이 출력되는 것을 방지하고 분류 정확도를 높인다.

이렇게 구성된 텍스트 형태의 시맨틱 표현은 JSCC 부호기를 거친 후 통신 채널로 전송된다. 수신 측에서는 JSCC 복호기로 전송받은 시맨틱 정보를 복원한 뒤, DALL-E 모델과 같은 이미지 생성 모델을 사용하여

이미지를 재구성한다. 이어서 강력한 다중 모달 LLM 중 하나인 GPT-4V 모델이 생성된 이미지와 설명 간의 의미적 일치성을 평가하며, 만약 일치하지 않는다고 판단되면 원문 프롬프트를 수정하여 DALL-E 모델에 재입력하는 방식으로 품질을 보정한다.

III. 결론

본 논문에서는 LLM을 활용한 시맨틱 통신의 다양한 구현 사례를 살펴보았다. 이들은 시맨틱 기반 오류 복원, 중요도 기반 자원 할당, 의미 복원 품질 향상 등 다양한 측면에서 진전을 이루었지만, 여전히 LLM의 다중 모달 처리 능력을 충분히 활용하지 못하고 있는 한계가 존재한다. 예컨대, 이미지 전체를 하나의 캡션으로 대체하여 전송하는 방식은 지나치게 단순화된 접근으로, 원본 데이터의 정교한 시각적 정보를 지나치게 해칠 가능성이 크다. 이는 LLM을 단순히 언어 생성 도구로만 활용하는 데 그친 결과로 볼 수 있다.

그러나 [4]에서 이미 LLM이 이미지나 오디오처럼 상대적으로 익숙하지 않은 모달리티에 관해서도 강력한 조건부 확률 예측 능력을 발휘함이 입증된 바 있다. 이는 LLM이 단지 언어 이해에 그치지 않고, 다양한 종류의 데이터에 대해서도 의미 기반 압축을 수행할 수 있는 유망한 도구임을 보여주는 사례다.

따라서 추후의 시맨틱 통신 연구들은 다중 모달 데이터를 LLM이 다룰 수 있는 텍스트 형태로 단순히 변환하기보다는, LLM의 다중 모달 확장 가능성을 보다 직접적으로 활용하는 방향으로 나아가야 할 것으로 보인다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00212103).

참고 문헌

- [1] T. Ren, R. Li, M.-m. Zhao, X. Chen, G. Liu, Y. Yang, Z. Zhao, and H. Zhang, “Separate source channel coding is still what you need: An LLM-based rethinking,” arXiv preprint arXiv:2501.04285, 2025.
- [2] S. Guo, Y. Wang, J. Ye, A. Zhang, P. Zhang, and K. Xu, “Semantic importance-aware communications with semantic correction using large language models,” IEEE Transactions on Machine Learning in Communications and Networking, 2025.
- [3] F. Zhang, Y. Du, K. Chen, Y. Shao, and S. C. Liew, “Addressing out-of-distribution challenges in image semantic communication systems with multi-modal large language models,” in 2024 22nd International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt). IEEE, 2024, pp. 7–14.
- [4] G. Delétang, A. Ruoss, P.-A. Duquenne, E. Catt, T. Genewein, C. Mattern, J. Grau-Moya, L. K. Wenliang, M. Aitchison, L. Orseau et al., “Language modeling is compression,” arXiv preprint arXiv:2309.10668, 2023.