

강인한 토큰 통신을 위한 변조 심볼 매핑 기법

김민우, 김용준*
포항공과대학교

{minwoo.kim, yongjune}@postech.ac.kr

Modulation Symbol Mapping for Robust Token Communications

Minwoo Kim, Yongjune Kim*

Pohang University of Science and Technology (POSTECH)

요약

기존의 디지털 통신 시스템은 비트 오류율 최소화에 최적화되어 있어, 채널 상태가 악화할 경우 의미 정보가 급격히 손실되는 결벽 효과가 발생한다. 본 논문에서는 트랜스포머 기반의 토큰 통신을 위해, 토큰 임베딩의 의미적 구조를 물리 계층의 성상도 상에서 보존하는 변조 심볼 매핑 기법을 제안한다. 이를 위해 고차원의 토큰 코드북을 다차원 척도법을 이용하여 2 차원으로 축소하고, 헝가리안 알고리즘을 통해 성상도 심볼과의 유클리드 거리 차이를 최소화하는 최적의 매핑을 수행한다. 제안한 기법을 이용한 이미지 분류 실험 결과, 잡음이 많은 환경에서도 기존 방식 대비 높은 정확도를 유지할 수 있었다.

I. 서론

최근 트랜스포머(Transformer)가 다양한 분야에서 핵심적인 역할을 수행함에 따라 모델의 정보 처리 단위인 토큰은 문맥과 의미를 내포하는 핵심 요소로 자리잡았다. 특히 협력 추론, 분할 컴퓨팅 환경에서는 토큰을 효과적으로 통신하는 것이 중요하며, 이에 따라 비트 단위의 무결성보다 의미 전달에 중점을 둔 시맨틱 통신(semantic communications)의 한 흐름으로 토큰 통신(token communications)이 등장하였다. 이러한 흐름 속에서, 최근 연구 [1]은 토큰을 코드북 기반으로 벡터를 양자화한 후 벡터 대신 코드북 인덱스만을 전송함으로써 통신량을 크게 줄이는 토큰 통신 개념을 제시한 바 있다. 그러나 코드북 인덱스를 기존의 비트 오류율 최소화 최적화된 디지털 통신 시스템으로 전송할 경우, 작은 비트 오류가 전체 문맥을 파괴하는 결벽 효과(cliff effect)가 발생하는 문제가 있다.

이에 본 논문에서는 의미상으로 유사한 토큰들이 고차원 공간상에서 가깝게 위치한다는 점에 착안하여, 이러한 기하학적 구조를 유클리드 거리를 기반으로 성상도(constellation) 상에서도 보존하여 비트 오류의 영향을 최소화하는 변조 심볼 매핑 프레임워크를 제안한다. 제안하는 기법은 고차원의 의미 공간을 저차원의 변조 공간과 정렬함으로써 기존 디지털 통신 표준과의 호환성을 유지하면서도 채널 잡음에 강인한 토큰 전송을 가능하게 한다. 또한, 제안한 기법이 잡음 환경에서 기존 기법 대비 우수한 분류 정확도와 점진적 성능 저하(graceful degradation) 특성을 보임을 검증한다.

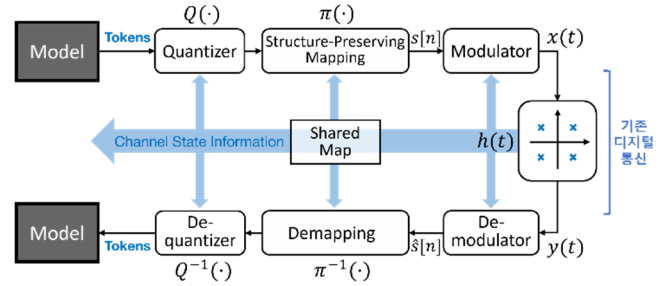


그림 1. 제안하는 토큰 통신 프레임워크.

II. 본론

그림 1은 본 논문에서 제안하는 전체적인 프레임워크를 나타낸다. 송신 단의 모델 인코더를 통과한 토큰들은 양자화기를 거쳐 토큰 코드북의 인덱스로 변환된다. 이때, 각 코드북 인덱스는 제안하는 구조 보존 심볼 매핑 알고리즘을 통해 물리 계층의 성상도 심볼과 일대일로 사전 매핑(mapping)되어 있다. 변조된 심볼은 잡음이 존재하는 채널을 통과하며 수신기에서는 이를 복조 및 역매핑(demapping)하여 수신 모델의 입력으로 사용한다.

제안하는 매핑 기법의 핵심은 고차원 의미 공간의 거리 구조를 2 차원 심볼 공간에 보존하는 것이다. 이는 수학적으로 이차 할당 문제(quadratic assignment problem, QAP)로 공식화할 수 있다. 그러나 QAP는 NP-hard 문제이므로 코드북의 크기가 커질수록 최적해를 구하는 것이 불가능해진다. 따라서 본 논문에서는 분할 정복(divide-and-conquer) 전략을 사용하여 해당 문제를 차원 축소와 선형 할당 문제(linear assignment problem, LAP)의 두 단계로 나누어 접근한다.

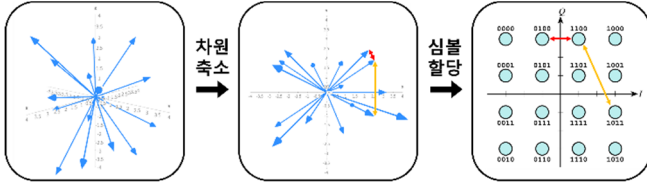


그림 2. 2 단계 구조 보존 변조 심볼 매핑의 시각화.

단계 1 [차원 축소]: 먼저 고차원의 토큰 코드북 벡터를 2 차원 평면으로 축소한다. 이때 데이터의 분산을 최대화하는 주성분 분석(principal component analysis, PCA)보다, 벡터 간의 유클리드 거리 관계를 저차원에서도 최대한 보존하는 다차원 척도법(multidimensional scaling, MDS)이 본 연구의 목적에 더욱 적합하다[2].

단계 2 [선형 할당 문제 해결]: 차원 축소된 2 차원 코드북 벡터들을 실제 quadrature amplitude modulation(QAM)의 성상도 위에 할당하는 문제는 LAP 을 통해 아래와 같이 공식화할 수 있다.

$$\min_{\pi \in S_M} \sum_{i=1}^M \|c'_i - s_{\pi(i)}\|_2$$

여기서 c'_i 은 차원 축소된 코드북 벡터이며, $s_j \in S_M$ 는 j 번째 M -QAM 심볼을 의미한다. 또한, $\pi(i)$ 는 코드북 인덱스 i 를 어느 심볼로 할당할지 결정하는 함수이다.

LAP 은 헝가리안 알고리즘을 이용해 $O(n^3)$ 의 다항 시간 복잡도로 최적해를 찾을 수 있음이 잘 알려져 있다[3]. 결과적으로, 헝가리안 알고리즘을 통해 의미적 유사성이 높은 토큰들이 성상도 상에서도 물리적으로 인접한 심볼에 할당되도록 정렬을 수행한다. 그림 2 는 이러한 두 단계의 매핑 과정을 시각화하여 보여준다.

III. 실험 결과

제안 기법의 유효성과 성능을 검증하기 위해, DeiT-Tiny 모델[4]의 [CLS] 토큰을 전송 대상으로 선정하였다. [CLS] 토큰은 고차원의 의미를 고밀도로 포함하므로 제안 기법의 유효성 확인 척도로 적합하다. 데이터셋으로는 CIFAR-10 을 사용하였다. 토큰 코드북은 k-평균 군집화(k-means clustering)를 수행하여 구성하였다.

그림 3 은 MDS 로 차원을 축소한 후 헝가리안 알고리즘을 적용하여 코드북 토큰이 1024-QAM 심볼에 할당된 결과이다. 그림에서 각 색상은 이미지의 클래스를 나타낸다. 동일한 클래스의 토큰들은 MDS 를 진행한 후 군집을 형성하며, 헝가리안 알고리즘을 통해 심볼 할당을 진행한 후에도 군집을 형성하고 있음을 확인할 수 있다. 이는 제안 기법이 의미적 유사성과 물리적 거리 간 정렬을 잘 수행했음을 의미한다.

그림 4 는 AWGN 채널 환경에서 SNR에 따른 수신 단의 이미지 분류 결과를 나타낸 것이다. 대조군으로는 코드북 인덱스를 이진 비트스트림(bit stream)으로 변환 후 그레이 코딩(Gray coding)을 적용한 방식을 사용하였다. 실험 결과, 대조군 대비 낮은 SNR 영역에서도 높은 정확도를 유지하며, 채널 상태가 악화되어도 절벽 효과 없이 점진적인 성능 저하가 일어남을 확인하였다.

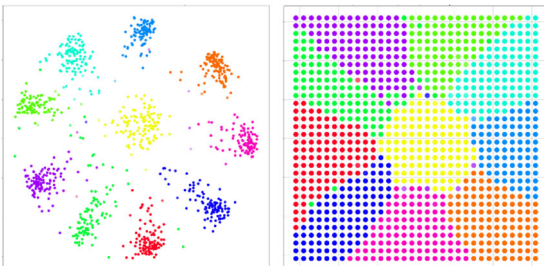


그림 3. (좌) MDS 차원 축소 결과 (우) 심볼 매핑 결과.

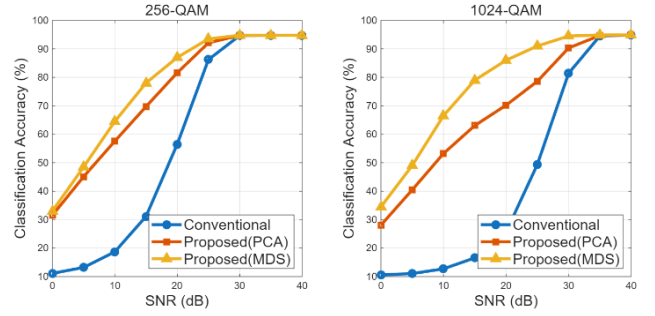


그림 4. 서로 다른 M -QAM 에서의 SNR-정확도 관계.

IV. 결론

본 논문에서는 강인한 토큰 통신 시스템을 위해, 토큰 임베딩의 의미론적 구조를 물리 계층의 성상도 상에서 보존하는 변조 심볼 매핑 기법을 제안하였다. 실험 결과, MDS 기법이 PCA 기법에 비해 고차원 토큰들 사이의 기하학적 관계를 효과적으로 보존하였다. 또한 헝가리안 알고리즘을 이용해 이를 심볼에 매핑함으로써 의미적 유사성과 물리적 거리 간 정렬을 잘 수행함을 확인하였다. 성능 평가 결과, 제안한 방식은 열악한 SNR 환경에서도 높은 분류 정확도를 유지하며 절벽 효과 없이 점진적 성능 저하를 달성하였다. 이는 기존 디지털 통신 시스템의 물리 계층을 수정하지 않고도 의미 정보를 포함한 토큰을 효율적이고 강인하게 전송할 가능성을 시사한다. 특히, 높은 차수의 변조를 사용하는 유선 통신이나 무선 백홀(backhaul) 등을 통해 토큰을 전송할 때 채널 상태가 나빠지더라도 성능 하락을 최소화하며 고차 변조를 유지할 가능성을 열어준다. 향후 심볼 간의 오류 확률 불균형을 고려한 오류 가중치 매핑, 코드북 분할 전송을 통한 적응형 통신 프로토콜, 오류정정부호 적용, 다양한 모달리티 간 확장 등의 추가적인 연구가 기대된다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (RS-2024-00398449, 네트워크 전문연구실(NRC): 통신 세대 진화를 위한 채널 부호 부복호 및 채널 추정 기술)

참 고 문 헌

- [1] L. Qiao et al., "Token Communications: A Large Model-Driven Framework for Cross-Modal Context-Aware Semantic Communications," *IEEE Wireless Communications*, vol. 32, no. 5, pp. 80-88, Oct. 2025
- [2] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1-27, Mar. 1964.
- [3] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83-97, Mar. 1955.
- [4] H. Touvron et al., "Training data-efficient image transformers & distillation through attention," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021.