

Transformer 기반 채널 예측 및 CSI report 기법

문지훈, 심병효
서울대학교

jhmoon@islab.snu.ac.kr, bshim@snu.ac.kr

Transformer-based Channel Prediction and CSI Report Technique

Jihoon Moon, Byonghyo Shim
Seoul National University

요약

본 논문에서는 인공지능(artificial intelligence; AI) 기술 중 Transformer를 사용하여 채널 예측을 수행하고, 그에 따라 채널 상태 정보(channel state information; CSI) report를 효율적으로 수행하는 기법을 제안하고자 한다. 5G NR에서는 높은 전송률을 달성하기 위해 고주파 신호 대역의 광대역 주파수를 활용하는데, 고주파 대역의 경우 전파의 경로 손실이 높다는 문제가 있다. 이를 보완하기 위해 송신단 및 수신단에서 많은 수의 안테나를 사용한 빔포밍 기법을 필수적으로 사용하는데, 빔포밍을 올바르게 수행하기 위해서는 CSI를 정확히 얻어야 한다. 기존 5G NR에서는 단말이 기지국에 CSI report를 수행할 때에는 어느 순간 수신한 파일럿 신호로부터 채널을 추정하고 그에 따라 CSI를 정해 기지국에 피드백하지만, 채널이 끊임없이 변화함에 따라 채널 추정 시의 채널과 데이터 전송 시의 채널이 달라질 경우 report한 CSI가 실제 데이터 전송에 맞지 않을 수 있는 문제가 있다. 따라서, 본 논문에서는 Transformer를 활용하여 채널을 미리 예측하고, 데이터 전송 시에 가장 효율적인 CSI를 생성하여 피드백하는 기법을 제안한다.

I. 서론

5G NR에서는 높은 전송률을 달성하기 위해 고주파 신호 대역의 광대역 주파수를 활용하는데, 고주파 대역의 경우 경로 손실이 높다는 문제가 있다. 이를 보완하기 위한 빔포밍 기법을 실현하기 위해, 정확한 채널 정보가 필요하다. 5G NR에서는 채널 정보를 얻기 위해 CSI report를 수행한다. CSI report는 기지국에서 단말에 channel state information-reference signal (CSI-RS)을 전송해 단말이 채널 추정을 수행하고, 이를 토대로 precoding matrix indicator (PMI), rank indicator (RI), channel quality indicator (CQI)이라는 세 가지의 CSI 정보를 생성해 기지국에 피드백한다. 이 중, PMI를 토대로 기지국은 빔을 형성하고 이를 통해 단말에 데이터 전송을 수행한다.

이러한 CSI report 과정의 단점은, 정확하고 최신인 CSI를 얻어내는 것이 굉장히 어렵다는 것이다. 구체적으로, CSI report 수행 주기가 길 경우 시간이 지나면서 채널 환경은 계속 변화하기 때문에, 채널이 끊임없이 변화하고, 채널 추정 시 생성한 CSI를 토대로 결정된 빔을 통해 데이터 전송을 수행할 때 실제 채널과 맞지 않아 데이터 전송률 손실이 크게 발생할 수 있다[1]. 반대로, CSI report 수행 주기가 짧을 경우에는 오버헤드가 많이 발생하거나 많은 피드백 비트 수를 사용해 CSI를 정확히 전달할 수 없다. 따라서, 데이터 전송 시의 채널을 효과적으로 얻기 위해서는 현재 및 과거에 수신한 CSI-RS 정보나 채널 추정 결과로부터 미래 데이터 전송 시의 채널을 예측할 필요가 있다.

본 논문에서는 인공지능 기술 중 Transformer를 활용해 CSI report에 채널 예측을 적용하고자 한다. 단말에서 채널을 미리 예측하고 그에 따라 CSI를 생성할 수 있다면, 빔포밍 이득을 최대화하여 데이터 전송률을 최대화할 수 있다. Transformer는 attention mechanism을 활용해 채널 시퀀스의 요소들 간 관계를 얻어 내어 미래 채널을 정확히 예측할 수 있도록 한다.

II. 본론

1) 시스템 모델

본 논문에서는 5G NR FR1 주파수 대역(3.5GHz) 신호를 사용한다. OFDM 시스템을 사용하고 부반송파 간격은 $60\text{kHz}(\mu=2)$ 로 설정하여 하나의 10ms frame 내에 40개의 slot이 포함된다. 본 논문에서는 computing power의 한계로 인해 한 단말이 10개의 resource block (RB)를 사용한다고 가정한다. 또한, 채널 추정은 8개 frame(=80ms)마다 한 번씩 수행한다.

기지국에서는 8X8 디지털 안테나 포트를 사용하는 것으로 설정하였으며, 단말은 편의상 단일 안테나를 사용한다.

2) Transformer 기반 채널 예측 네트워크 개요

본 논문에서는 N 개의 CSI-RS 수신 신호에서 추정된 채널을 입력으로 사용해 각 채널 추정 시간 이후 예측된 채널을 내보낸다. 채널 예측 네트워크의 함수는 다음과 같다.

$$f(\hat{\mathbf{H}}(-(N-1)\tau), \hat{\mathbf{H}}(-(N-2)\tau), \dots, \hat{\mathbf{H}}(0)) \\ = \left(\hat{\mathbf{H}}(0), \hat{\mathbf{H}}\left(\frac{1}{4}\tau\right), \hat{\mathbf{H}}\left(\frac{2}{4}\tau\right), \hat{\mathbf{H}}\left(\frac{3}{4}\tau\right) \right)$$

여기서 τ 는 채널 추정 간격이다. N 개의 과거 및 현재 추정 채널로부터 다음 채널 추정 전까지의 채널을 예측한다. $\frac{1}{4}\tau = 20\text{ms}$ 간격의 채널을 추정하기 때문에, 기존 추정 채널들보다 세밀하게 채널의 변화를 따라갈 수 있다.

3) Transformer 기반 채널 예측 네트워크 구조 설명

채널 예측 네트워크는 기존 Transformer의 decoder를 하나의 Transformer block으로 보고 이를 여러 층으로 쌓은 형태이다. 하나의 Transformer block에서 가장 핵심적인 부분은 self-attention으로, 입력을 queue (Q), key (K), value (V)로 변환해 입력 sequence 간의 상관관계를 파악해 attention score를 계산한다. 구체적으로, 입력 element를 X 라 하면,

$$\mathbf{Q} = \mathbf{X} \times \mathbf{W}_Q, \mathbf{K} = \mathbf{X} \times \mathbf{W}_K, \mathbf{V} = \mathbf{X} \times \mathbf{W}_V$$

의 연산을 통해 Q, K, V를 구하고 attention score를 다음과 같이 구한다.

$$\text{Attention score}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}.$$

Attention score는 어떤 한 element의 query와 다른 element의 key 사이의 중요도를 얻은 뒤 중요도에 따라 weight를 다르게 value의 weighted sum을 얻은 것을 의미한다. 실제 self-attention은 시퀀스에 있는 서로 다른 특성을 효과적으로 얻어내기 위해 여러 개의 attention을 동시에 수행하며, 이를 multi-head attention이라 한다[2].

실제 채널 예측 네트워크에서, 입력의 시간 간격은 τ 인데 반해 예측 채널의 $\frac{1}{4}\tau$ 이기 때문에, 두 가지의 시간 간격의 attention score를 학습하기 위해 두 가지의 시간 간격으로 Transformer block을 쌓고, 중간에 하나의 fully-connected layer를 넣어 시퀀스 길이를 4배로 늘려준다.

4) 학습 방식

추정한 채널들을 입력으로 하고, 출력되는 채널의 실제 값을 통해 supervised learning을 수행해 채널 예측 네트워크를 학습시킨다. 이때, loss function은 다음의 normalized MSE를 사용한다.

$$L = \frac{\|\mathbf{H}_t - \hat{\mathbf{H}}_t\|^2}{\|\mathbf{H}_t\|^2}.$$

5) 예측한 채널 기반 PMI 선택

예측한 채널을 기반으로 빔포밍에 활용할 PMI를 선택한다. 이때, 예측한 채널들에 평균적으로 효율적인 PMI를 선택하도록 한다.

$$\mathbf{W}_T = \arg \max_{\mathbf{W} \in \mathcal{C}} \sum_{t=t_s}^{t_e} \sum_{f=f_s}^{f_e} \log_2 \left(1 + \frac{\|\mathbf{W}^T \mathbf{H}(f, t)\|_2^2}{\sigma^2} \right)$$

기존 채널 추정에 의해 선택한 PMI에 비해 데이터 전송 시 채널들을 예측하고, 이를 종합적으로 고려하여 PMI를

선택하기 때문에 미래에 데이터 전송 시 채널이 추정 시 채널과 달라도 데이터 전송률 손실을 줄일 수 있다.

III. 결론

1) 실험 결과

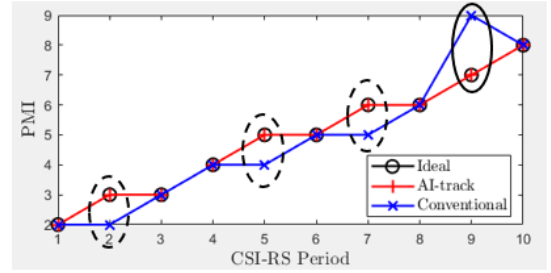


그림 1. 채널 예측에 기반한 PMI 선택 결과

그림 1에서 볼 수 있듯이, 채널 예측을 기반으로 PMI를 선택할 경우 기존 추정 채널 기반으로 PMI를 선택했을 때와 다른 PMI를 선택하는 경우가 자주 있고, 이것이 이상적인 PMI와 잘 일치하는 것을 볼 수 있다.

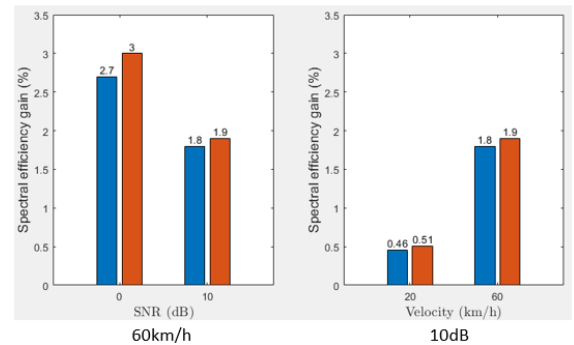


그림 2. 채널 예측 기반 PMI 선택의 데이터 전송률 증가

그림 2는 채널 예측을 기반으로 PMI를 선택할 경우 얻을 수 있는 데이터 전송률 이득을 보여준다. 채널의 변화가 빨라 PMI의 변화가 빈번할 때, 또한 SNR이 낮아 채널 예측 기법이 채널 추정의 오류를 보정할 수 있을 때 스펙트럼 효율 증가가 큰 것을 볼 수 있다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00208985)

참고 문헌

- [1] Jiang, H., Cui, M., Ng, D. W. K., and Dai, L., "Accurate channel prediction based on transformer: Making mobility negligible," IEEE Journal on Selected Areas in Communications, vol. 40, no. 9, pp. 2717-2732, 2022.
- [2] Vaswani, A., et al., "Attention is all you need," Advances in neural information processing systems, 30, 2017.