

대규모 언어 모델 기반 mmWave 빔 추적 기법

한동민, 이훈
울산과학기술원

fkukdjd23@unist.ac.kr, hoonlee@unist.ac.kr

Large Language Model-Based mmWave Beam Tracking Methods

Dong Min Han and Hoon Lee
Ulsan National Institute of Science and Technology

요약

본 논문에서는 mmWave 다중안테나 시스템에서 빔 추적 문제를 해결하기 위한 large language model (LLM) 활용 방안을 제안한다. 단말의 이동성에 대한 수학적 모델이 존재하는 경우, 빔 추적 문제는 전통적인 extended Kalman filter (EKF) 방법으로 해결할 수 있다. 그러나 이러한 사전 정보가 주어지지 않는 현실적인 상황에서는 EKF 적용이 불가능하다. 제안하는 LLM 방법은 in-context learning (ICL) 훈련 방식을 사용하여 수학적 모델에 의존하지 않으면서도 정교한 빔 추적이 가능하다. 모의실험을 통해 제안하는 LLM 기반 빔 추적 방식의 우수성을 입증한다.

I. 서론

mmWave 대역 통신 시스템에서 시변 채널에 동적으로 대응하기 위해서는 실시간으로 빔 방향을 예측하는 정교한 빔 추적 기술이 필요하다 [1]. 시변 채널의 상태천이 모델이 사전에 주어지는 경우 extended Kalman filter (EKF)와 같은 전통적인 추정 알고리즘을 적용할 수 있다. 그러나 실제 무선 환경에서는 동적 모델을 알 수 없어 적용이 어렵다.

모델 기반 빔 추적 방식의 한계를 극복하기 위해 인공지능 기반의 데이터 주도적 빔 추적 기법이 제안되었다[2]. Long short-term memory (LSTM) 등의 순환 신경망 모델을 활용하면 수학 모델 없이 정교한 빔 추적이 가능하다. 그러나 과적합 문제로 인해 훈련 데이터와 테스트 환경이 상이한 경우 성능이 저하된다.

한편, 최근 large language model (LLM)의 높은 추론 능력이 다양한 modality의 시계열 예측 문제에서 검증되고 있다[3]. 소수의 예시들을 입력으로 사용하는 in-context learning (ICL) 기능을 활용하면 LLM의 일반화 능력을 개선할 수 있다.

본 논문에서는 mmWave 다중안테나 시스템의 빔추적 문제를 ICL 구조로 재정의하여 LLM을 학습하는 새로운 빔 추적 기법을 제시한다. 제안하는 방식은 명시적 상태천이식이나 채널 모델을 사용하지 않고, 과거 수신 신호와 빔 방향 정보를 ICL 예시로 활용하여 최적 빔 방향을 예측한다. 모의실험 결과 제안하는 방법이 종래의 EKF 방식 대비 높은 정확도를 보임을 검증한다.

II. 시스템 모델

N_t 개의 송신 안테나가 배치된 기지국이 N_r 개의 안테나를 갖는 이동 단말에 데이터를 전송하는 mmWave 대역 통신 시스템을 고려한다. 기지국과 단말은 모두 uniform linear array (ULA) 구조를 사용하며, 안테나

간격은 반파장으로 가정한다. t 번째 time slot에서 단말의 수신 신호 $\mathbf{y}_t \in \mathbb{C}^{N_r \times 1}$ 는 다음과 같이 표현된다.

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t \quad (1)$$

여기서 $\mathbf{x}_t \in \mathbb{C}^{N_t \times 1}$ 는 송신 신호, $\mathbf{H}_t \in \mathbb{C}^{N_r \times N_t}$ 는 mmWave 채널 행렬, $\mathbf{n}_t \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ 는 복소 가우시안 잡음이다. mmWave 대역의 희소성을 고려하여 채널 행렬은 다음과 같은 line-of-sight 형태로 모델링한다.

$$\mathbf{H}_t = \alpha_t \mathbf{a}_R(\phi_{A,t}) \mathbf{a}_T(\phi_{D,t})^H \quad (1)$$

여기서 α_t 는 복소 채널 이득, $\mathbf{a}_R(\cdot)$ 및 $\mathbf{a}_T(\cdot)$ 는 각각 단말 및 기지국 ULA의 array response 벡터, 그리고 $\phi_{A,t}$ 및 $\phi_{D,t}$ 는 각각 angle-of-arrival (AoA) 및 angle-of-departure (AoD)를 의미한다.

빔 추적을 위해 상태 벡터를 다음과 같이 정의한다.

$$\mathbf{x}_t = [\Re\{\alpha_t\}, \Im\{\alpha_t\}, \phi_{t,A}, \phi_{t,D}]^T \quad (2)$$

상태 벡터의 시계열은 n 차 autoregressive (AR) 시퀀스로 다음과 같이 모델링한다.

$$\mathbf{x}_t = \sum_{i=1}^n \mathbf{A}_i \mathbf{x}_{t-i} + \mathbf{u}_t \quad (3)$$

이때 행렬 \mathbf{A}_i 는 다음과 같이 정의한다.

$$\mathbf{A}_i = \begin{cases} \text{diag}([\rho, \rho, 1, 1]), & i = 1 \\ \left(\frac{1}{n}\right) \mathbf{I}, & \text{otherwise} \end{cases}$$

ρ 는 자기회귀 계수이며, $\mathbf{u}_t \sim \mathcal{N}(0, \Sigma_u)$ 의 잡음이다. 잡음의 공분산 행렬은 다음과 같이 정의된다.

$$\Sigma_u = \text{diag}([(1 - \rho^2), (1 - \rho^2), \sigma_A^2, \sigma_D^2]) \quad (6)$$

III. 제안하는 LLM 기반 빔 추적 방법

GPT2를 backbone 모델로 사용하는 LLM 기반 빔 추적 방식을 제안한다. ICL 기능을 활용하기 위해 과거 L 개의 time slot에서 취득한 수신 신호 벡터와 추정한 AoA/AoD $\{(y_{t-L}, \phi_{t-L}), \dots, (y_{t-1}, \phi_{t-1})\}$ 를 few-shot 예시로 사용한다. 이러한 문맥 정보를 기반으로, LLM은

제시된 현재 시점의 수신 신호 y_t 를 입력받아 최적의 빔 방향을 출력하도록 설계한다. Time slot k 의 입력 토큰 z_k 는 다음과 같이 정의된다.

$$z_k = \text{concat}(\Re\{y_k\}, \Im\{y_k\}, \phi_{A,k}, \phi_{D,k}) \in \mathbb{R}^{2N_r+2N_r^2} \quad (7)$$

마지막 time slot 의 경우, 아직 알 수 없는 빔 방향은 0으로 채운다.

Time slot t 의 ICL 예시 $S = (z_{t-L}, \dots, z_t)$ 는 위치 임베딩을 통해 입력 벡터 E 로 변환되고, 이를 GPT2 모델이 처리하여 출력 시퀀스 H 를 아래와 같이 반환한다.

$$H = (h_{t-L}, \dots, h_t) = \text{GPT2}(E) \quad (8)$$

여기서 GPT2(\cdot)은 n_{head} 개의 head 및 n_{layer} 개 transformer 블록으로 구성한다. 최종적으로, 전체 문맥 정보를 포함하는 마지막 은닉 벡터 h_t 가 linear layer 를 거쳐 빔 방향 예측값을 다음과 같이 생성한다.

$$\hat{\phi}_t = [\hat{\phi}_{A,t}, \hat{\phi}_{D,t}] = \text{Linear}(h_t) \quad (9)$$

LLM 은 mean-squared-error (MSE)를 최소화하도록 다음의 loss function 으로 훈련한다.

$$L = \frac{1}{T} \sum_{t=1}^T \|\hat{\phi}_t - \phi_t\|^2 \quad (10)$$

제안하는 ICL 학습 메커니즘은 LLM 이 명시적 상태 개선 없이도 문맥 내 수신 신호의 패턴을 학습하여 EKF 의 예측-개선 과정을 데이터 기반으로 내재적으로 수행하도록 한다.

IV. 실험 결과

본 절에서는 제안하는 빔 추적 기법의 우수성을 검증한다. GPT2 구조는 $n_{\text{layer}} = 2$ 개의 transformer 블록과 $n_{\text{head}} = 4$ 개의 head 로 이루어져 있으며, 임베딩 차원은 64 로 설정한다. LLM 의 훈련 데이터는 AR(3) 동적 모델로 생성하고, 문맥 길이는 $L = 3$ 으로 고정한다. 훈련에는 Adam 알고리즘을 사용하고, learning rate 는 10^{-4} , batch size 는 64 로 고정한다.

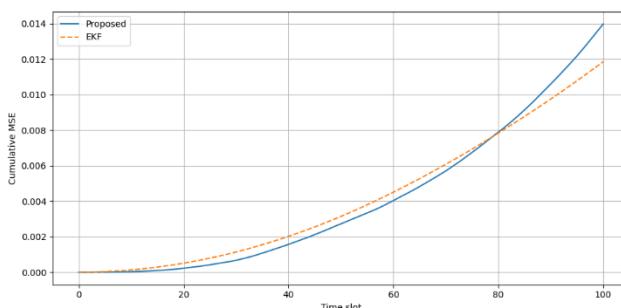


그림 1. AR(1) 환경에서 MSE 성능 비교

그림 1 은 제안하는 LLM 및 EKF 기반의 빔 추적 기법의 누적 MSE 성능을 time slot 에 대해 나타낸다. LLM 방식의 일반화 능력을 검증하기 위해 테스트는 훈련 환경과 상이한 AR(1) 동적 모델 데이터로 수행한다. 훈련과 테스트 모두 signal-to-noise ratio (SNR)은 10 dB 로 고정한다. 실험 결과, 제안하는 LLM 기반 빔 추적 방식의 성능이 종래의 EKF 방식과 유사함을 확인할 수

있다. LLM 의 모델 비의존성과 훈련/테스트 환경의 상이함을 고려하면, 모델 기반 EKF 방식과 유사한 빔 추적 성능을 보이는 것은 매우 고무적인 결과이다.

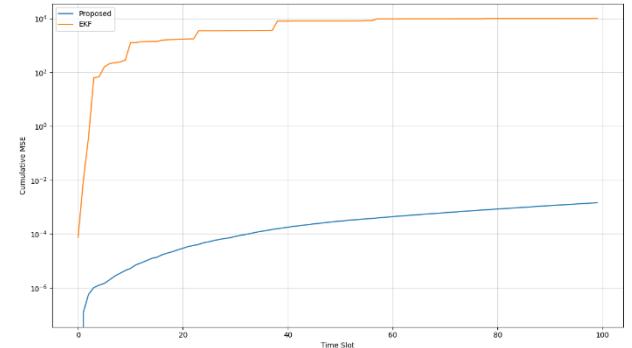


그림 2. AR(3) 환경에서 MSE 성능 비교

그림 2 는 AR(3) 테스트 환경에서의 MSE 성능 비교 결과이다. SNR 10dB 인 상황에 대해서 제안하는 LLM 방식의 누적 MSE 성능이 time slot 에 대해 완만한 증가폭을 보인다. 이에 비해 EKF 방식은 AR(1) 형태의 데이터를 추적하도록 설계되어 AR(3) 데이터에서는 모델 불일치 문제로 누적 MSE 가 빠르게 증가함을 알 수 있다. 이러한 결과를 통해 제안하는 LLM 기반 빔 추적 방식이 문맥 내 패턴을 스스로 파악하여 즉석에서 일반화를 수행할 수 있다고 결론지을 수 있다.

V. 결론

본 연구는 명시적 상태천이 및 잡음 통계 가정 없이 관측 시퀀스만으로 각도를 추정하는 ICL 기반 LLM 빔 추적 기법을 제안한다. 모의실험을 통해 제안하는 LLM 방식이 종래의 모델 기반 빔 추적 알고리즘의 성능을 상회함을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원 (2021-0-00467, 지능형 6G 무선 액세스 시스템)의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] V. Va, H. Vikalo, and R. W. Heath, "Beam tracking for mobile millimeter wave communication systems," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, pp. 743-747, 2016.
- [2] J. Park, S. Hwang, H. Lee, and I. Lee, "Deep recurrent Q-network methods for mmWave beam tracking systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13429-13434, Dec. 2022
- [3] M. Zecchin, K. Yu, and O. Simeone, "In-context learning for MIMO equalization using transformer-based sequence models," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1573-1578, 2024.