

MoE 기반 다중 사용자 목적 지향적 통신 시스템 타당성 분석

안세영, 고승우
인하대학교, 인하대학교

syahn020@inha.edu, swko@inha.ac.kr

Feasibility Study on MoE-Driven Multi-user Goal-Oriented Communications

Se-Yeong Ahn, Seung-Woo Ko
Inha Univ., Inha Univ.

요약

본 연구에서는 Large Language Model (LLM) 및 artificial general intelligence (AGI)의 핵심 동작 원리인 Mixture-of-Expert (MoE) 구조를 활용하여 여러 유저들이 혼재한 goal-oriented 통신 시스템을 설계하는 것을 목적으로 한다. 특히 본 연구에서는 다른 유저의 신호로 인한 간섭이 존재할 때, MoE 를 통해 효율적으로 간섭을 제어하여 미션을 수행할 수 있는지를 확인한다. 두명의 유저가 공존하는 binary classification 문제를 MoE 의 효율적 간섭 제어를 통해 synthetic 데이터 상황에서 테스트 정확도가 100% 달성함을 보였다.

I. 서론

Large Language Model (LLM) 및 artificial general intelligence (AGI)의 등장함에 따라, 핵심 동작 원리인 Mixture-of-Expert (MoE)에 대한 관심이 높아지고 있다. MoE 는 다수의 문제가 복합적으로 있는 복잡한 task 를 효율적으로 여러 개의 expert 에 분산시켜 해결하는 방법론이다. 본 연구에서는 MoE 구조를 활용하여 여러 유저들이 혼재한 goal-oriented 통신 시스템을 설계하는 것을 목적으로 한다. 특히 본 연구에서는 다른 유저의 신호로 인한 간섭이 존재할 때, MoE 를 통해 효율적으로 간섭을 제어하여 미션을 수행할 수 있는지를 확인한다.

II. 본론

본 논문에서는 하나의 기지국과 두 명의 업링크 유저로 MAC 상황을 고려한다. 두 유저는 모두 binary classification 미션을 가지고 있다. 유저 단말의 연산 능력 부족으로 인해 on-device 학습이 불가능하다고 가정한다. 대신에 두 유저는 데이터를 업링크 전송으로 기지국에 오프로딩하여 기지국이 가지고 있는 연산 서버에서 대신 처리하는 edge learning 구조를 고려한다. 각 단말이 가지고 있는 typical 데이터 샘플을 x_1 과 x_2 로 정의한다. 전송 시 서로 간의 간섭을 줄이기 위해

데이터에 semi-orthogonal 한 시퀀스를 곱해서 보낸다. 이때 코딩된 결과들을 \hat{x}_1 와 \hat{x}_2 로 정의한다. 본 데이터는 h_1 과 h_2 채널을 통해 기지국으로 전송되고 기지국에서 수신된 신호는 다음과 같다.

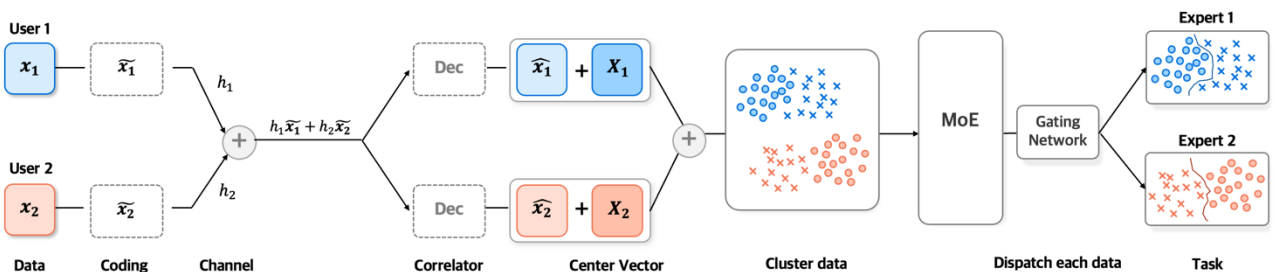
$$y = h_1 \hat{x}_1 + h_2 \hat{x}_2$$

위 수신된 신호는 semi-orthogonal 한 시퀀스를 곱하여 각각 추출한다. 이때 시퀀스의 non-orthogonality 로 인해 각 단말의 수신 신호는 다른 단말의 신호에 간섭을 받게 된다. 즉 수신된 데이터는 다음과 같이 표현된다.

$$\hat{x}_1 = x_1 + \alpha \frac{h_2}{h_1^*} x_2$$
$$\hat{x}_2 = x_2 + \alpha \frac{h_1}{h_2^*} x_1$$

1) MoE

Mixture-of-Experts (MoE) 구조는 gating (routing) 네트워크와 여러 expert 로 구성된다. gating 네트워크는 주어진 입력에 가장 적합한 expert 를 선택한다. 각 expert 는 동일한 구조로 설계되어 하위 작업을 수행한다. 데이터가 클러스터 구조를 가질 때 MoE 는 단일 모델보다 높은 성능을 보인다. 각각의 클러스터를 가장



적합한 하나의 Expert 에 dispatch 한다. MoE 의 이러한 특성을 이용해 각 유저가 동시에 데이터를 전송했을 때, 모델이 자체적으로 데이터를 구분해 처리할 수 있도록 한다.[1],[2]

2) 데이터 처리

MoE 의 input 을 클러스터 형태로 만들기 위해 다음과 같이 데이터를 처리한다.

유저가 전송할 데이터 $x^{(1)}$ 은 feature signal 로 label signal vector v_k 에 label 을 곱해 생성한다.

$x^{(1)}$ 에 $x^{(2)}$ 를 더해 클러스터를 형성하고 이는 MoE 의 input 데이터이다. $x^{(2)}$ 는 cluster-center signal with noise 로 cluster-center signal vector c_k 와 random 가우시안 노이즈, $v_{k'}$ ($k' \neq k$)를 더해 생성한다. $v_{k'}$ 는 다른 유저에 의해 발생하는 간섭으로 간주할 수 있다.[1]

3) 실험 결과

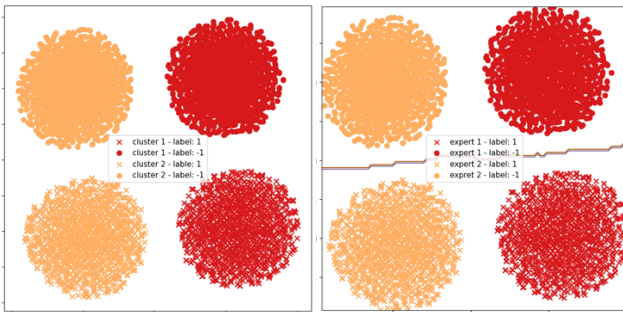


Figure 2. $\alpha = 0$

	User 1	User 2
Expert 1	3190	0
Expert 2	0	3210

Table 1. $\alpha = 0$

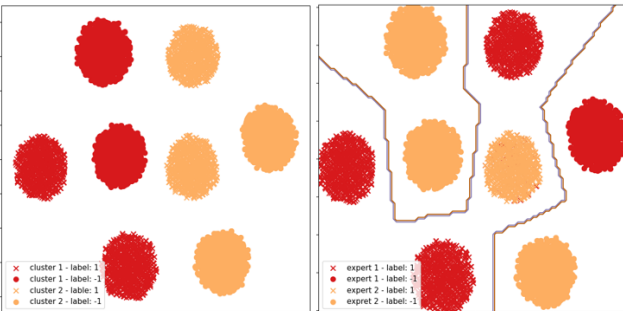


Figure 3. $\alpha = 1$

	User 1	User 2
Expert 1	1632	1611
Expert 2	1623	1534

Table 2. $\alpha = 1$

Figure 2는 간섭을 조절하는 scaling factor α 가 0일 때 결과를 t-SNE 를 사용하여 시각화한 것이다. 이 경우, 각 유저가 전송한 데이터 간의 간섭이 없다. Figure 3 는 α 가 1 일 때의 결과를 동일한 방법으로 시각화한 것으로, 이 때는 각 유저의 신호 크기가 동일하다. 두 figure 모두 왼쪽은 데이터 분포를, 오른쪽은 두 개의 expert 를 사용하는 MoE 의 결과를 보여준다.

데이터의 색은 해당 데이터가 어느 expert 에 dispatch 되었는지를 나타낸다. 주황색은 expert 1, 빨간색은 expert 2 로 배정된 데이터이다. 사용된 마커의 형태는 레이블을 나타내며, 'o'는 레이블 0 을, 'x'는 레이블 1 을 의미한다.

Table 1 과 Table 2 는 각 유저의 데이터가 어떤 Expert 로 dispatch 되었는지 나타낸다. 모든 α 값에서 classification test accuracy 는 100%를 달성하였다. α 가 0.8 이하일 때는 유저 간 간섭이 없어 각 유저의 데이터가 적절하게 단일 expert 에 dispatch 되었다. 그러나 α 가 0.9 이상일 때는 각 유저 데이터를 제대로 구분하지 못하고 여러 expert 에 dispatch 되었다.

III. 결론

본 논문에서는 MoE 기반 다중 유저 Goal-Oriented 통신 시스템에 대한 가능성에 대해 알아보았다. 두명의 유저가 공존하는 binary classification 문제를 MoE 의 효율적 간섭 제어를 통해 synthetic 데이터 상황에서 테스트 정확도가 100% 달성함을 보였다. 추후 무선 신호의 노이즈 및 페이딩 환경을 고려한 실험으로 확장할 예정이다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00347, 6G 통신을 위한 Post Mac).

참 고 문 헌

- [1] Z. Chen, Y. Deng, Y. Wu, Q. Gu, and Y. Li, Towards understanding mixture of experts in deep learning. arXiv preprint arXiv:2208.02813,s 2022.
- [2] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, J. Dean, Outrageously large neural networks: The sparsely-gated Mixture-of-Experts layer. arXiv preprint arXiv:1701.06538, 2017.