



2025 2학기 컴퓨터소프트웨어학부 CSE융합세미나



장소

ITBT관 911호

날짜 / 시간

2025.12.10 16:00~18:00

Upcoming Talk



송경우

조교수 / 연세대학교

**LLM 성능 개선을 위한 통계적
방법론**

LLM의 안정적인 학습 및 성능 개선을 위해서, layer norm, retrieval augmented generation, 그리고 여러가지 환각탐지 방법들이 널리 활용되고 있다. 본 강연에서는 이러한 기존 방법들의 문제점을 탐색하고, 더욱 개선할 수 있는 통계적인 방법에 대해서 다룬다. 첫번째로, 기존 layer norm의 문제점을 개선하여, 효율성과 안정성을 모두 갖춘 새로운 activation function인 bounded hyperbolic tangent (BHyT)를 소개한다. BHyT는 layer norm과는 다르게 variance를 계산하지 않기에 효율적이며, 체비셰프 부등식을 활용하여 안정적인 학습을 돕는다. 두번째로, 여러 embedding 모델들의 장점을 결합할 수 있는, 새로운 embedding ensemble 방법론을 소개한다. Laplace approximation을 통해서 특정 데이터에 대한 기존 임베딩들의 불확실성을 정량화 하고, 이를 기반으로 embedding을 ensemble 하여, 보다 우수한 검색 성능을 달성하게 한다. 마지막으로, 이러한 기법들에도 불구하고 발생할 수 있는 환각 문제를 다루기 위하여, conformal inference 기반의 새로운 환각 탐지 방법에 대해서 다룬다.



한양대학교
HANYANG UNIVERSITY