

2024 CNU CSE 전문가 초청세미나

일시: 2024. 1. 16. (화) 13:00~14:30

장소: 공대 5호관 607호 세미나실

대상: 컴퓨터공학과 대학원생 및 학부생 누구나

문의: 임성수 교수 (sungsu@cnu.ac.kr, 042-821-5448)



주제: **효율적인 그래프 계산:**

응집성을 갖는 서브그래프, 커뮤니티 검색, 네트워크 안정성 등

연사: **김정훈 교수 (UNIST 컴퓨터공학과)**

초록: 그래프는 개체 간의 복잡한 관계를 분석하고, 소셜 네트워크, 인터넷, 금융 거래 네트워크와 같은 다양한 대규모 네트워크를 모델링할 수 있습니다. 이 발표는 응집성을 갖는 서브그래프, 그래프 분해, 네트워크 안정성에 중점을 둔 모델들을 소개할 것이며, 광범위한 그래프 데이터 처리를 위한 효율적인 알고리즘과 업데이트 기술에 대해서도 다룰 것입니다. 조밀하게 연결된 노드들로 구성된 응집성을 갖는 서브그래프는 그래프 데이터의 기본 구성 요소로 간주됩니다. 이러한 서브그래프의 계산은 그래프를 분해하고 그 연결 구조를 추상화함으로써, 숨겨진 정보를 효율적으로 발견할 수 있도록 합니다. 또한, 그래프 분해는 다양한 네트워크의 구조적 안정성을 분석하고, 네트워크 안정성을 증진시키거나 유지하는 데 중요한 노드 및 연결을 식별하는 데 사용됩니다. 이 발표는 최근 연구 동향 및 향후 연구 방향에 대해서도 논의할 예정입니다.

13:00~13:40



주제: **효율적인 트랜스포머 기반 언어 모델 추론**

연사: **송환준 교수 (KAIST 데이터사이언스대학원)**

초록: 최근 자연어 처리(NLP) 분야에서 GPT 모델과 같은 트랜스포머 기반 거대 언어 모델이 우수한 성능을 보여주고 있습니다. 그러나 생성형 작업인 문서 요약 및 번역과 같은 과업을 수행하기 위한 트랜스포머 모델의 자기회기 디코딩 (Autoregressive Decoding)은 모든 매개변수를 반복적으로 활용하므로 추론 지연 시간이 높아지는 문제가 있습니다. 이로 인해 트랜스포머 디코딩 과정의 지연시간을 줄이는 것은 거대 모델을 폭넓게 활용하기 위한 중요한 과제 중 하나입니다. 본 세미나에서는 트랜스포머 추론의 효율성을 높이면서도 생성 텍스트의 품질을 유지하거나 향상시킬 수 있는 방법론을 소개합니다. 구체적으로는 트랜스포머 추론 과정을 근사화할 수 있는 조기 정지(Early Exiting) 방법과 거대 모델의 지식을 소형 모델로 전이시키는 지식 전이(Knowledge Distillation)를 중심으로 한 최근 연구 내용을 논의합니다.

13:50~14:30

