KUBIG DL Basic Study

23기 정준호

▶ week 01

1. AI 패턴: Conversation & Human Interaction

2. 관심 분야 및 패턴 적용: 모호한 질문에 되물어보는 능동형 대화 에이전트

현대인들은 LLM을 검색 엔진 대신 사용하며, 여행 계획을 세우기도 하는 등 일상적인 질문을 LLM에게 던집니다. 그러나 사용자들은 프롬프트 사용이 익숙치 않아, 추상적인 질문을 던지는 경향이 있습니다. 이러한 사용자를 위한 에이전트 개발에 관심이 있습니다. 사용자의 질문이 명확하지 않거나 맥락이 결여된 상황에서는, 불확실성을 감지하여 역질문을 던지는 능동형 대화 에이전트를 생각했습니다. 불확실성이 높을 경우, 모델은 답변 생성을 멈추고 질문 생성 모드로 즉각 전환합니다. 기존 RAG 모델에 증거 기반 딥러닝 기술 및 강화학습을 추가하여 성능을 높입니다.

3. 모델의 입력 및 출력

입력 : 사용자의 질문
ex) 맛집을 추천해줄래?, 데이트 장소 추천해줄래?

출력: 불확실성 평가 이후 사용자에게 답변 (반복과정)
사용자 질문을 고차원 벡터로 변환하여 잠재 공간 내에서 연관된 정보를 추출합니다. 추출된 정보들로 불확실성 점수를 계산합니다. 정보 간의 의미적 분산이 크거나 학습 데이터 분포와 상이한 형태를 보일 경우, 높은 불확실성 값을 도출합니다. 불확실성 값이 임계값을 초과하면 역질문을 생성하는 시나리오로 이어갑니다. 불확실성 엔트로피를 효과적으로 낮출 수 있는 최적의 질문을 생성하여 사용자에게 답변합니다. 질문과 답변의 과정을 반복하며 신뢰도를 높입니다.

4. 참고자료 및 논문

- D. Bang, H. Choe, and J. Kang, "Enhancing the Reliability of LLM through GraphRAG," in Proceedings of the KICS 2024 Korea AI Conference, 2024, pp. 382-383.

일상적인 세부 질의는 RAG로 즉시 처리하되, 맥락 파악이 필요한 고난이도 질의에 대해서만 선택적으로 GraphRAG를 활성화하거나 사용자에게 구체화를 요구함으로써 정확도를 높입니다.

- S. Kang and S. Kim, "Enhancing RAG-LLM Response Performance Using Ensemble Diversity," Journal of the Korea Institute of Information and Communication Engineering, vol. 28, no. 8, pp. 916-926, Aug. 2024.

앙상블 모델들이 내놓은 답변들이 서로 비슷하다면 불확실성이 낮은 것이고, 서로 다르다면 불확실성이 높은 것으로 간주하는 식으로, 불확실성 점수를 매기면 좋을 것 같습니다.