

LLM을 활용한 뉴스 요약 및 추천 시스템 연구

안영민 김기훈 강민지 지은주

Abstract

최근 대형 언어 모델(Large Language Models, LLM)은 자연어 처리뿐만 아니라 다양한 정보 요약 및 응용 분야에서 그 활용도가 급격히 증가하고 있다. 뉴스 기사 생산 속도가 증가함에 따라 수많은 정보를 신속하고 정확하게 파악하는 것은 사용자에게 큰 부담이 되고 있다. 이에 본 연구에서는 LLM을 활용하여 뉴스 기사를 효율적으로 요약하고, 사용자가 원하는 키워드나 질문에 따라 관련성 높은 최신 기사를 제공하는 시스템을 제안한다. RAG(Recursively Augmented Generation) 기반 접근 방식을 통해 뉴스 기사를 벡터화하고 유사도 계산을 통해 관련 뉴스를 검색하여 요약하는 방법을 구현하였다. 이를 통해 최신 뉴스 정보를 효과적으로 요약하여 제공할 수 있는 시스템의 가능성을 제시한다.

Keywords: LLM, 뉴스 요약, RAG, 벡터화, ChromaDB, 자연어 처리, 정보 검색

1. 서론

연구 배경 및 목적

대형 언어 모델(Large Language Models, LLM)은 자연어 처리 분야에서 혁신적인 변화를 이끌어오고 있으며, 최근에는 자연어 요약, 정보 검색, 대화형 인공지능 등 다양한 영역으로 활용 범위가 확장되고 있다. 이러한 변화 속에서 뉴스 기사는 시시각각 새로운 정보로 업데이트되며, 사용자가 원하는 정보를 신속히 찾기 어려운 문제를 야기하고 있다. 특히, 수많은 뉴스 데이터 중에서 관련성 높은 최신 정보를 찾아내고 요약하는 작업은 매우 복잡하고 시간이 많이 소요된다.

본 연구의 목적은 LLM을 활용하여 뉴스 기사를 효과적으로 요약하고, 사용자가 특정 키워드나 질문을 입력하면 관련성 높은 최신 뉴스를 빠르게 제공하는 시스템을 제안하는 데 있다. 이를 위해 RAG 기반 접근 방식을 사용하여 뉴스 기사를 벡터화하고, 유사도 계산을 통해 최신 뉴스를 검색하고 요약하는 방법을 구현하였다.

2. 본론

2.1 시스템 아키텍처 및 구성 요소

본 연구에서 제안한 뉴스 요약 시스템은 크게 세 가지 모듈로 구성된다: 데이터 수집 모듈, 벡터화 및 검색 모듈, 요약 및 응답 생성 모듈. 첫째, 데이터 수집 모듈은 네이버 뉴스 API와 공공데이터 포털을 이용하여 최신 뉴스를 수집하며, 수집된 기사는 전처리 과정을 거쳐 JSON 형태로 저장된다.

둘째, 벡터화 및 검색 모듈에서는 HuggingFace 임베딩 모델을 사용하여 뉴스 제목과 본문을 벡터화하고, 이를 크로마DB에 저장하여 검색 속도를 높인다. 검색 시에는 사용자가 입력한 키워드를 벡터화하여 데이터베이스 내 문서들과의 코사인 유사도를 계산하여 관련성이 높은 뉴스를 추출한다.

셋째, 요약 및 응답 생성 모듈은 추출된 뉴스 기사를 LLM을 이용하여 간결하게 요약하며, 사용자가 입력한 질문과 검색된 뉴스 요약을 바탕으로 자연스러운 대화형 답변을 생성한다. 이를 통해 단순한 정보 나열이 아닌, 사용자에게 실질적인 정보를 제공할 수 있다.

2.2 RAG를 활용한 뉴스 기사 검색 및 요약

RAG 기반 접근 방식은 사용자의 질문과 관련된 뉴스 기사를 빠르게 찾아내고, 가장 관련성 높은 기사들을 추려내어 요약하는 데 효과적이다. LLM이 질문의 의미를 파악하여 키워드를 벡터화하고, 크로마DB 내 벡터와 비교하여 높은 유사도의 뉴스를 선정한다. 이후 LLM을 통해 요약을 수행하여 자연스러운 답변을 생성함으로써 정보 탐색의 효율성을 크게 향상시킨다.

3. 결론

본 연구에서는 LLM 기반 뉴스 요약 및 검색 시스템을 제안하고 구현하여, 대용량 뉴스 데이터 속에서 사용자가 원하는 정보를 신속하고 정확하게 제공할 수 있는 가능성을 확인하였다. 제안한 시스템은 뉴스의 실시간성 확보와 관련성 높은 정보 제공이라는 두 가지 문제를 효과적으로 해결하며, 정보 탐색 시간 단축과 요약 효율성 향상에 기여하였다.

향후 연구로는 다양한 뉴스 주제에 대응하기 위해 주제 분류를 강화하고, 다중 모달 데이터(이미지, 영상 등)를 활용하여 복합적인 정보 제공이 가능한 통합 뉴스 요약 시스템을 구축할 예정이다. 이를 통해 보다 직관적이고 풍부한 정보를 사용자에게 제공할 수 있을 것이다.

References

1. Mirchandani, S., Xia, F., Florence, P., Ichter, B., Driess, D., Gonzalez Arenas, M., Rao, K., Sadigh, D., Zeng, A. (2023). *Large Language Models as General Pattern Machines*. arXiv preprint arXiv:2307.04721.