

재난분야 특화 언어모델기반 재난정보 탐색 모델 개발

최선화

국립재난안전연구원

shchoi33@korea.kr

A development of Disaster Information Exploration Model based on Language Model Specialized in Disaster

Choi Seon Hwa

National Disaster Management Research Institute

요약

본 논문에서는 생성형 AI 기술을 활용해 개발한 재난정보 지능형 탐색 모델(DiGPT : Disaster information Generative Pre-trained Transformer)을 소개하고자 한다. 대량의 재난정보에서 원하는 정보를 탐색, 분석, 그리고 정리 작업은 재난관리 업무 중 상당한 인력과 시간이 소모되는 반복적 업무다. DiGPT는 이를 지능화하기 위한 모델로 초거대 언어모델(Large Language Model : LLM)과 검색증강생성(RAG) 기술을 활용해 개발하였다. 공개된 언어모델에 재난문서를 추가 학습하고 파인 티닝하여 재난분야 특화 언어모델을 구축하였다. 더불어, 언어모델이 내부 학습된 데이터만을 기반으로 답변할 경우, 발생할 수 있는 정보의 최신성 부족과 환각 현상을 보완하기 위해서 재난문서 검색을 통해 언어모델의 프롬프트를 증강하고 양질의 답변을 생성하도록 한다. 본 논문에서는 DiGPT의 아키텍처와 구축 방법을 설명하고 이를 기반한 대화형 탐색 서비스를 소개하겠다.

I. 서론

본 논문에서는 재난관리 업무 전반에서 정책 기획, 트랜드 및 이슈 분석을 위해 자료를 검색하고 분석하는 일은 상당한 인력과 시간을 소모한다. 이 같은 노동집약적 조사·분석은 데이터가 많아질수록 비용이 커지는 비효율적인 작업이다. 이와 같은 작업을 효율적으로 지원하기 위해 국립재난안전연구원에서는 재난정보 지능형 탐색 모델(DiGPT, Disaster information Generative Pre-trained Transformer)을 개발하고 있다. 모델은 재난관리 분야 특화 언어모델(LLM: Large Language Model)을 기반하는 것으로, 사용자의 질문 의도를 파악하고 적절한 답변을 제공한다. 모델에는 LLM의 부정확한 답변과 최신성 부족 등의 문제를 해결하기 위해 검색증강생성(RAG: Retrieval Augmented Generation) 기법을 적용하였다. 재난관리에 필요한 정보를 지식 그래프 DB로 구축하고 각종 재난문서는 벡터 DB로 구축하여 재난관리 지식베이스를 구성하였다. 재난관리 분야 지식베이스에서 사용자 질의와 관련된 정보를 먼저 검색한다. 검색 결과는 LLM의 프롬프트를 증강하는데 활용되어 LLM이 정확한 답변을 생성하도록 한다. 본 논문에서는 DiGPT의 아키텍처와 구축 방법을 설명하고 이를 기반한 대화형 탐색 서비스를 소개하고자 한다[1].

II. 재난정보 지능형 탐색 모델(DiGPT)

재난정보 지능형 탐색 모델(DiGPT)은 반복적이고 노동집약적인 업무를 효율적으로 지원하기 위한 모델로 초거대 언어모델과 검색증강생성(RAG) 기술을 활용하여 개발하였다. DiGPT는 재난분야 특화 언어모델과 재난분야 내부 지식베이스로 구성된다[1].

재난분야 특화 언어모델은 알리바바가 공개한 디코더 기반의 오픈소스인 Qwen2.5 32B 모델에 대량 한국어 코퍼스를 학습한 한국어 특화 모델을 기본 모델(Foundation model)로 활용하였다. 여기에 위기관리메뉴얼,

재해연보, 재난연감 등 재난분야의 문서 14종과 재난관련 법령 197종을 추가 학습하여 재난분야 특화 언어모델을 구축하였다. 이 모델이 RAG 기반의 질의응답이 가능하도록 Instruction-finetuning을 통해 재난분야 질의응답에 적합하도록 미세조정하였다.

재난분야 내부 지식베이스는 정형과 비정형 데이터를 구분하여 구축하였다. 정형 데이터는 상황 파악에 필요한 42종을 필수 데이터, 즉 기상, 시설, 상황대응, 사고피해 데이터와 연구원에서 보유한 재난사례 기록물 데이터를 지식그래프로 구축하였다. 비정형 데이터는 위기관리메뉴얼, 재해연보, 재난연감 등 재난분야 문서 14종과 재난관련 법령 197종을 대상으로 벡터 DB로 구축하였다. 또한, 최근 3년간의 재난안전 관련 뉴스 기사를 추가하여 지식베이스를 보강하였다.

RAG 파이프라인은 RAG 서비스 게이트웨이, 지식그래프 기반 질의응답 시스템(KBQA), 벡터 검색 시스템, LLM으로 구성되어 있다. 전체 워크플로우는 사용자가 입력한 질문을 기반으로 정보를 검색하고 답변을 생성하는 과정을 단계적으로 수행한다. 우선 사용자는 RAG 서비스 게이트웨이를 통해 질문을 입력한다. 이 게이트웨이는 전체 워크플로우를 관리하며 질문에 대한 처리를 시작한다. 첫 번째 단계로, 게이트웨이는 KBQA와 벡터 검색 시스템으로 전달한다. KBQA는 지식 그래프를 활용하여 질문과 관련된 구조화된 데이터를 검색하고, 이를 기반으로 초기 응답을 생성하여 게이트웨이로 반환한다. 벡터 검색 시스템은 입력된 질문을 벡터 임베딩으로 변환한 후 데이터베이스 내에서 관련성이 높은 결과를 검색하고, 이를 게이트웨이로 반환한다. 이후 게이트웨이는 KBQA에서 받은 응답과 벡터 검색 시스템의 결과를 조합하여 LLM이 이해할 수 있는 프롬프트를 생성한다. 생성된 프롬프트는 LLM으로 전달되며, LLM은 이를 기반으로 자연스러운 형태의 최종 답변을 생성한다. 마지막으로 게이트웨이는 LLM에서 생성된 응답을 사용자에게 반환함으로써 질의응답 프로세스를

마무리한다. 이러한 워크플로우는 지식 그래프와 벡터 검색, LLM의 협업을 통해 구조화 및 비구조화된 데이터의 통합적 활용을 가능하게 하며, 사용자에게 정확하고 풍부한 정보를 제공할 수 있도록 설계하였다.



그림 1. DIGPT와 대화형 탐색 서비스

DIGPT를 활용해 재난정보를 대화하면서 탐색할 수 있는 대화형 탐색 서비스를 개발하였으며, 그림 1은 DIGPT 구성과 서비스 처리 과정을 설명한 개념도이다. 재난분야 전문 질의응답이 가능한 서비스로 대량의 자료와 데이터에서 질의와 관련된 정보를 탐색하고 이를 요약분석하여 답변한다. 재난특화 언어모델이 답변을 생성할 때 활용한 자료의 출처 즉, 재난 문서의 다운로드가 가능하여, 답변의 정확성을 확인할 수 있다.

DIGPT 기반 대화형 탐색 서비스 테스트 목적으로 실질적 사용자의 수요와 기능상의 문제들을 수집하고 이를 서비스 기능 개발에 반영하기 위해 국립재난안전연구원 내부에 제한적 베타 서비스를 운영하였다. 연구원에서는 재난 정책, 이슈, 그리고 재난 동향 등의 분석보고서를 수시로 보고한다. 이는 상당한 인력과 시간을 소모하는 대표적인 노동집약적 업무이다. 금번 베타 서비스 운영을 통해 연구원들이 분석보고서를 작성하는 과정에서 가장 많은 시간과 인력을 소모하는 자료 탐색과 검토 과정을 획기적으로 단축할 수 있다는 가능성을 확인하였다.

III. 결론

본 논문에서는 생성형 AI 기술을 활용해 개발한 재난정보 지능형 탐색 모델, DIGPT와 이를 기반한 대화형 탐색 서비스를 소개하겠다. 재난 특화 언어모델을 학습하고 재난분야 내부 지식베이스를 구축하여 재난분야 전문 대화형 탐색 서비스를 개발하여 베타 서비스를 운영하고 가능성을 확인하였다. 또한, 재난분야 최초 생성형 AI 기술을 활용한 모델과 서비스를 제시하였다.

지식베이스에 재난분야 뉴스 기사를 확대하여 구축할 예정이며, 답변의 정확성과 정교함의 고도화를 지속적으로 추진하고 있다. 특히, DIGPT는 과거 재난사례에 대한 요약분석에 탁월한 성능과 강점을 보이도록 이에 집중하여 연구개발 할 계획이다.

참 고 문 헌

- [1] 최선화, “재난정보 지능형 탐색 모델의 벡터 검색 성능에 관한 연구”, 2025 한국컴퓨터종합학술대회 논문집, pp. 635-637, 2025
- [2] Vinicius G. G and Nicholas R., W., “DisasterResponseGPT: Large Language Models for Accelerated Plan of Action Development in

Disaster Response Scenarios”, arXiv:2306.17271v1, 2023

- [3] 최선화 외 3인 , “재난 상황관리 지능화를 위한 지식 그래프 설계”, 2023 한국산업정보학회 추계학술대회 논문집, pp. 51-53, 2023
- [4] 최선화 외 3인 , “재난관련 언론기사 분석에 관한 초거대 언어모델 활용성 검토 연구”, 2024 한국컴퓨터종합학술대회 논문집, pp. 1528-1530, 2024
- [5] 최선화 외 1인 , “재난 상황정보 지능형 탐색 모델에 관한 연구”, 2024 한국산업정보학회 추계학술대회 논문집, 2024