

검색 증강 생성 기반의 학술 논문 자동 리뷰 시스템

홍건우, 김용강*

국립공주대학교

redgil77@smail.kongju.ac.kr, *ygkim@kongju.ac.kr

Automated Academic Paper Review System Based on Retrieval Augmented Generation

Geonwoo Hong, Yonggang Kim*

Kongju National University

요약

본 논문은 검색 증강 생성 기반의 학술 논문 자동 리뷰 시스템을 제안한다. 제안하는 시스템은 공개된 학회의 실제 리뷰 데이터를 지식 베이스로 활용하여 전문적인 리뷰를 자동 생성한다. 이를 통해 연구자는 논문 제출 이전에 연구의 타당성과 완성도를 사전에 점검할 수 있다.

I. 서론

최근 인공지능 기술의 비약적인 발전과 연구 지원 도구의 보급으로 인해 전 세계적으로 학술 논문의 생산 및 투고 건수는 매년 급격한 증가세를 보이고 있다. 그럼 1에서 확인할 수 있듯이, OpenAlex 데이터를 기반으로 분석한 연도별 논문 출판 수는 주요 출판사를 중심으로 지속적인 증가 추세를 보이며 급격히 확대되고 있음을 확인할 수 있다 [1]. 이러한 학술 논문의 양적 성장은 지식의 확산을 가속화한다는 긍정적인 측면이 있으나, 동시에 기존의 인간 중심 리뷰 시스템에 심각한 과부하를 초래하고 있다. 일반적으로 논문 심사 과정은 수개월에서 길게는 1년 이상 소요되는 경우가 빈번하며, 이러한 심사 지연은 연구 결과의 시의성을 저해하는 주요 원인이 된다. 따라서 연구자가 논문 제출 이전에 신속하고 전문적인 피드백을 받아 논문의 완성도를 높일 수 있는 자동화된 리뷰 시스템의 필요성이 대두되고 있다.

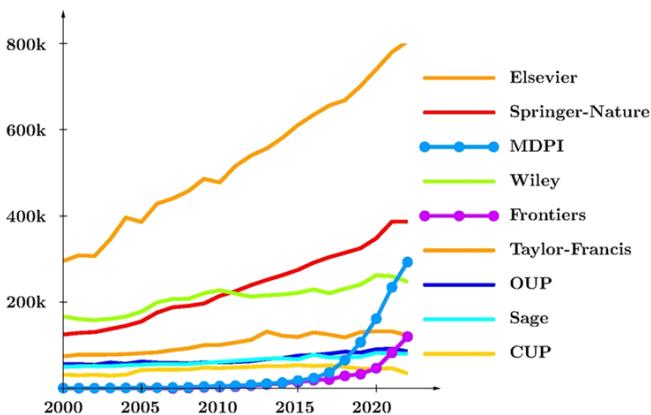


그림 1 연도별 및 출판사별 논문 출판 수 증가 추이

기존에도 대규모 언어 모델 (LLM)을 활용하여 논문 초안을 검토하거나 요약하려는 시도가 있었으나, 이는 주로 문법적 오류를 수정하거나 문장을 다듬는 표면적인 수정에 그치는 한계가 존재하였다. 학술 논문 리뷰는 연구의 독창성, 실험의 타당성, 관련 연구와의 비교 등 깊이 있는 전문 지식을 요구하지만, 범용 LLM은 특정 학술 도메인의 심사 기준을 정확히 반영하지 못하거나 사실이 아닌 내용을 생성하는 환각 현상을 보일 위험

이 존재한다. 단순한 생성 모델만으로는 실제 학회 심사위원 수준의 구체적이고 비판적인 피드백을 제공하기 어렵다는 점이 기술적 난제로 지적되어 왔다.

이에 본 연구에서는 단순한 텍스트 생성을 넘어, 실제 학술적 평가 기준에 부합하는 피드백을 제공하기 위해 검색 증강 생성 (RAG) 기반의 논문 자동 리뷰 시스템을 제안한다 [2]. 본 시스템은 LLM의 사전 학습 지식에만 의존하지 않고, 신뢰할 수 있는 외부 지식 베이스를 참조하여 답변을 생성하는 방식을 선택하였다. 이를 위해 권위 있는 인공지능 학회의 오픈 리뷰 데이터 플랫폼인 OpenReview에서 수집한 실제 리뷰 데이터를 벡터 데이터베이스로 구축하여 활용한다.

제안하는 시스템은 사용자가 논문을 업로드하면 구축된 지식 베이스에서 해당 논문의 문맥과 유사한 실제 리뷰 사례를 검색하고, 이를 LLM에 프롬프트로 함께 제공하여 맞춤형 리뷰를 생성한다. 이러한 방식은 기존 생성 모델의 한계를 보완하여 논문의 논리적 구조와 실험적 검증의 미비점을 구체적으로 지적할 수 있게 한다. 본 연구를 통해 구현된 시스템은 연구자에게 사전 심사 환경을 제공함으로써, 제출 전 논문의 품질을 스스로 개선하고 연구 역량을 강화하는 데 기여할 것으로 기대된다.

II. 데이터셋 구축 및 지식 베이스 설계

본 시스템은 범용 LLM이 가진 학술적 전문성 부족 문제를 해결하고 신뢰할 수 있는 피드백을 제공하기 위해, 권위 있는 인공지능 학회의 실제 리뷰 데이터를 기반으로 지식 베이스를 구축하였다. 고품질의 학술적 피드백을 참조하기 위한 데이터 소스로는 세계적인 인공지능 학회인 ICM, ICLR, NeurIPS 등의 오픈 리뷰 플랫폼 OpenReview를 선정하였다. 다음으로, OpenReview API를 활용하여 각 학회에 투고된 논문 메타 데이터와 해당 논문에 대해 작성된 리뷰 텍스트를 수집하였다. 이때 단순한 리뷰 텍스트뿐만 아니라, 논문의 최종 게재 여부, 리뷰 점수, 메타 리뷰 등 평가 정보를 함께 수집하여, 향후 모델이 긍정적 평가와 부정적 평가의 패턴을 구분하고 심사 결과에 따른 피드백의 차이를 학습할 수 있도록 메타데이터를 구조화하였다.

수집된 논문과 리뷰 데이터는 원문에는 불필요한 노이즈가 다수 포함되어 있으므로, RAG 시스템의 검색 정확도를 높이기 위해 정제 과정을 거쳤다. 우선 텍스트 내에 포함된 LaTeX 수식 코드, Markdown 형식의 불필요한 태그, 특수 기호 등을 제거하여 순수 텍스트 위주로 정제하는 작업을 수행하였다. 이후 논문의 문맥 유지를 위해 의미 기반의 청킹 기법을 적용하였다. 논문의 전체 텍스트를 문단 단위로 분할하되, 섹션 헤더 정보를 메타데이터로 보존하여, 향후 검색 시 단순 키워드 매칭을 넘어 특정 섹션의 문맥에 특화된 리뷰를 찾을 수 있도록 데이터를 가공하였다.

전처리가 완료된 텍스트 데이터는 검색 가능한 형태인 벡터로 변환되어 저장된다. 본 연구에서는 학술 및 기술 문서의 문맥 이해도가 높은 Qwen 임베딩 모델을 사용하여 텍스트를 고차원 벡터로 변환하였다 [3]. 생성된 벡터는 ChromaDB에 저장하여 관리하며, 이때 논문의 제목, 초록, 학회명 등의 메타데이터를 벡터와 함께 매핑하여 저장하였다. 이를 통해 시스템은 단순한 유사도 기반 검색 뿐만 아니라, 특정 학회나 연도, 접수대별 필터링을 결합한 하이브리드 검색이 가능한 견고한 지식 베이스를 갖추게 되었다.

III. RAG 기반 리뷰 생성 시스템의 설계 및 구현

본 연구에서 제안하는 시스템은 사용자가 업로드한 논문을 분석하고, 구축된 지식 베이스에서 관련 정보를 검색하여 전문적인 리뷰를 생성하는 검색 증강 생성(RAG) 파이프라인을 기반으로 동작한다. 전체 시스템은 크게 검색 모듈과 생성 모듈로 구성되며, FastAPI 기반의 REST API 환경에서 유기적으로 연동된다.

우선 검색 단계에서는 사용자가 입력한 논문의 텍스트가 쿼리로 활용된다. 시스템은 입력된 논문을 문단 단위로 분석하고, 데이터 구축 단계에서 사용된 것과 동일한 Qwen 임베딩 모델을 사용하여 텍스트를 벡터 공간에 투영한다. 이후 벡터 데이터베이스인 ChromaDB 내에서 코사인 유사도를 기반으로 입력 논문의 내용과 가장 문맥적으로 유사한 기준의 합격 및 불합격 논문의 리뷰 데이터를 검색한다. 이때 단순히 텍스트 유사도만이 아니라, 논문의 분야 및 섹션 정보를 필터링 조건으로 활용하여 검색의 정확도를 높인다. 이때 입력 논문과 유사한 리뷰 사례들의 검색 결과 예시는 그림 2에 제시하였다.

검색된 유사 리뷰 데이터와 입력 논문의 원문은 생성 단계로 전달되어, 최종 리뷰를 작성하기 위한 핵심 문맥으로 활용된다. 본 시스템의 생성 모델로는 긴 문맥 처리 능력과 추론 성능이 우수한 Gemini 2.5 Flash Lite 모델을 선택하였다. 생성 모듈은 검색된 실제 리뷰 사례들을 프롬프트에 포함시키는 문맥 주입 기법을 통해, LLM이 범용적인 조언이 아닌 학술적 어조와 형식을 갖춘 피드백을 생성하도록 유도한다. RAG 기반으로 자동 생성된 논문 리뷰 결과 화면은 그림 3과 같다.

최종적으로 본 시스템은 연구자의 접근성을 높이기 위해 웹 기반 플랫폼으로 구현되었다. 사용자는 별도의 설치 없이 논문 파일을 업로드하는 것 만으로 서비스를 이용할 수 있으며, 연구 단계에 맞춰 두 가지 리뷰 모드를 선택할 수 있다. Fast 모드는 신속한 피드백을 목표로 하여 핵심적인 리뷰 요소를 빠르게 제공하며, Deep 모드는 상대적으로 긴 처리 시간이 소요되지만 논문의 구조적 완성도, 연구 방법의 타당성, 실험 결과의 신뢰성 등을 심층적으로 분석하여 실제 투고 시 받아볼 수 있는 상세한 피드백을 제공한다.

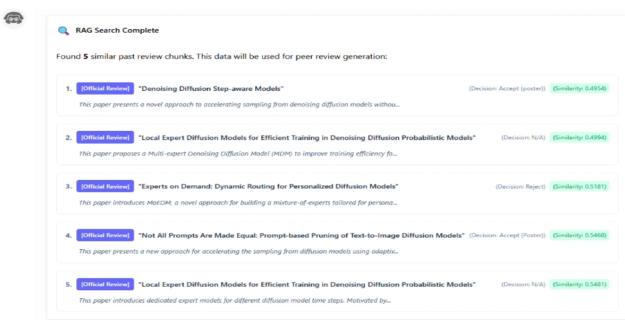


그림 2 논문 유사도 검색 결과 화면

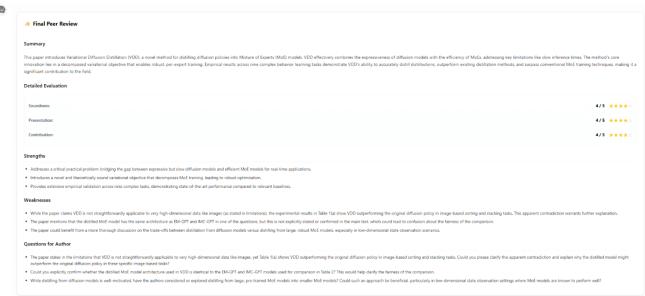


그림 3 자동 생성된 논문 리뷰 결과 화면

IV. 결 론

본 연구에서는 급증하는 학술 논문 투고 수와 이에 따른 심사 지연 문제를 해결하고, 특히 초심 연구자들에게 양질의 피드백을 제공하기 위해 RAG 기반의 학술 논문 자동 리뷰 시스템을 제안하고 구현하였다. 기존의 범용 언어 모델이 가진 학술적 전문성 부족과 환각 현상의 한계를 극복하기 위해, 권위 있는 인공지능 학회의 실제 리뷰 데이터를 수집하여 지식 베이스를 구축하였다.

구현된 시스템은 논문 텍스트를 문맥 단위로 분석하고, 유사한 평가 사례를 검색하여 생성 모델에 주입함으로써, 단순한 문장 교정을 넘어 연구의 타당성과 논리적 완결성을 평가하는 전문적인 리뷰를 생성할 수 있음을 확인하였다. 또한 웹 기반 서비스로 구현하여 접근성을 확보하고, Fast 모드와 Deep 모드를 통해 사용자 상황에 맞춘 유연한 서비스를 제공하였다. 본 시스템은 연구자가 논문 제출 전에 자신의 연구를 객관적으로 점검하고 보완할 수 있는 사전 심사 도구로서, 연구 역량을 강화하고 학술 생태계의 효율성을 높이는 데 기여할 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임 (IITP-2026-RS-2024-00438430).

참 고 문 헌

- Pier B. and Romary L., "Excessive growth in the number of scientific publications," *Ouvrir la Science*, Oct. 21, 2024.
- Lewis P. et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459 - 9474, 2020.
- Yang A. et al., "Qwen2.5 Technical Report," *arXiv preprint*, arXiv:2412.15115, 2024.