언어모델을 활용한 기계학습 기반의 학술지 추천 모형 개발*

- 국내 학술출판 생태계를 중심으로 -

A Study on Developing Machine Learning-based Journal Recommendation System Using Language Models: Focusing on the Scholarly Publishing Ecosystem of South Korea

정 재 민 (Jaemin Chung)**
남 은 경 (Eunkyung Nam)***
김 완 종 (Wan Jong Kim)****

초 록

학문 분야의 융합을 통한 학제 간 연구가 확대되고 접근 가능한 전자저널의 수가 증가함에 따라, 연구자들이 논문을 투고할 적합한 학술지를 선택하는 것이 더욱 어려워지고 있다. 특히 학술출판 시스템별로 서비스하는 학술지가 상이하고 국내 학술단체에서 발행하는 국제 학술지들이 증가하는 현 학술 생태계를 반영한 학술지 추천 연구는 부족한 실정이다. 본 연구는 언어모델을 활용한 기계학습 기반의 학술지 추천 아키텍처를 제안한다. 제안된 아키텍처는 타겟 데이터에 대하여 추가 학습된 BERT 계열 언어모델을 통해 논문의 제목과 초록을 임베딩하고, 이 임베딩 벡터를 XGBoost 모델에 입력하여 학술지를 추천한다. 분석 결과, BERT 계열 언어모델 중 RoBERTa 모델이 가장 우수한 성능을 보였으며, 특히 RoBERTa 기반 추천 시스템의 정확도는 전통적인 자연어 처리 기법 기반 시스템보다 약 13% 이상 높게 나타났다. 학습된 모델을 활용하여 서비스 대상 학술지의 범위를 벗어난 논문과 국문으로 작성된 논문에 대한 추천의 유효성을 확인하였다. 본 연구는 국내 학술논문 데이터를 통해 언어모델과 기계학습을 결합한 학술지 추천 시스템의 활용 가능성을 보였다는 점에서 학술적으로, 실제 국내 학술출판 환경을 고려한 서비스 아키텍처를 제시했다는 점에서 실무적으로 의의가 있다.

ABSTRACT

As interdisciplinary research expands through the convergence of academic fields and the number of accessible electronic journals increases, researchers face growing challenges in selecting appropriate journals for manuscript submission. There is a lack of research on journal recommendation systems that reflect the Korean academic ecosystem, in which academic services offer different sets of journals and international journals published by Korean academic societies are increasing. This study proposes a machine learning-based journal recommendation architecture that leverages language models. The proposed architecture embeds paper titles and abstracts using BERT-based language models further trained on target data, and these embedded vectors are then input into an XGBoost classifier to recommend appropriate journals. Analysis results showed that among BERT-based models, RoBERTa demonstrated the best performance, with its recommendation system outperforming approximately 13% higher compared to systems based on traditional natural language processing techniques. Furthermore, it was found that recommendations for papers outside the scope of service journals and papers written in Korean were feasible. This study contributes both academically and practically by presenting an academic journal recommendation architecture that leverages language models and machine learning by considering the actual Korean academic publishing environment.

키워드: 학술지 추천, 언어모델, 문서 임베딩, 기계학습, 다중 분류
Journal Recommendation, Language Model, Document Embedding, Machine Learning, Multi-Class Classification

* 본 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구임. (과제번호: (KISTI)K24L1M1C2, (NTIS)2710007937)
** 한국과학기술정보연구원 데이터큐레이션센터 연구원(jmchung@kisti.re.kr) (제1저자)
*** 한국과학기술정보연구원 데이터서비스센터 선임연구원(eknam@kisti.re.kr) (공동저자)
**** 한국과학기술정보연구원 데이터서비스센터 책임연구원(wjkim@kisti.re.kr) (교신저자)
논문접수일자 : 2025년 2월 14일 논문심사일자 : 2025년 2월 14일 게재확정일자 : 2025년 2월 28일
한국비블리아학회지, 36(1): 109-126, 2025. http://dx.doi.org/10.14699/kbiblia.2025.36.1.109

1. 서 론

인터넷상에 공개되는 정보의 양이 많아짐에 따라 이용자의 선호를 바탕으로 관련된 정보를 추출 및 제안해주는 추천 시스템의 필요성이 높아지고 있다. 학술 연구 분야에서도 협업 파트너(Li et al., 2024; Liu et al., 2023), 인용문헌(Da et al., 2022; Dinh et al., 2025; Lu et al., 2023), 심사위원(Liao et al., 2024; Tan et al., 2021; Zhang et al., 2020) 등을 추천하는 다양한 추천 시스템이 연구 및 서비스되고 있다. 이러한 학술 추천 시스템은 연구자들의 연구 생산성을 높일 수 있는 지능적인 도구이다. 그 가운데 연구 성과의 공개 및 확산에 가장 중요한 과제는 학술지 추천이다. 연구의 결과는 논문으로 공개가 되는 경우가 많으며, 연구자들은 영향력 있는 관련 학술지에 본인의 논문을 게재하고자 한다(Pradhan et al., 2020). 학술지 선정은 학술지의 품질, 심사 기간, 게재료 등의 요인들에 영향을 많이 받으며, 최근에는 분야별 학술지의 수가 점차 늘어나고 다양한 주제의 학술논문을 출판하는 다학제 학술지가 등장함에 따라 연구자들이 논문을 투고할 학술지를 선택하는 일은 점차 어려워지고 있다(김용우 외, 2023).

학술지 선택에 관한 연구자들의 의사결정을 지원하기 위하여 공개된 학술 데이터를 활용하여 적절한 학술지를 추천해주는 연구는 꾸준히 수행되어 오고 있다(Zhang et al., 2023). 하지만 국내 서비스를 고려한 학술지 추천 시스템에는 여전히 다음과 같은 한계가 존재한다. 첫째, 학술 서비스는 저마다 서비스 범위가 달라 추천할 수 있는 학술지도 서비스마다 상이하다. 다시 말해 연구자가 입력한 논문이 서비스 중인 학술지 범위에서 벗어나는 논문일 경우 추천할 적절한 학술지가 없다는 결과를 제공할 수 있어야 한다. 하지만 선행연구들은 분석 대상 학술지만으로 연구를 수행해 예외의 경우에 관한 결과를 제공하지 못하는 한계가 있다. 둘째, 국내 서비스에서도 영문 논문을 대상으로 한 학술지 추천 시스템을 구축할 필요가 있다. 최근 국내 과학기술 분야가 발전하고 국내 학술단체에서 발행하는 국제 학술지의 종 수도 증가하고 있다. 대다수의 국내 학술지는 국문 초록과 영문 초록을 함께 출판하거나 영문 초록만 출판하는 등 영문 초록을 필수적으로 포함한다. 하지만 선행연구는 국문 데이터만을 활용하였기 때문에 영문 초록만 사용하는 학술지를 서비스 대상에 포함할 수 없다는 한계점을 가진다. 마지막으로 학술지 추천 시스템을 국내 서비스에 도입하기 위한 구체적인 아키텍처를 제안한 연구는 아직 수행되지 않았다. 국영문을 모두 고려한 서비스 아키텍처 구성을 통해 국내 연구자의 학술출판 지원을 고려할 필요가 있다.

본 연구는 국내 학술 생태계를 반영한 학술지 추천 시스템을 제안하는 것을 목표로 한다. 제안된 시스템 아키텍처는 학술지 데이터셋 구축, 학술지 추천을 위한 모델 학습, 이용자 입력에 대한 학술지 추천으로 구성된다. 데이터셋 구축 단계에서는 서비스 범위에서 벗어난 논문을 분류할 수 있는 더미 학술지를 반영한다. 학술지 추천 단계에서는 논문의 의미론적인 특징을 고려하기 위하여 언어모델을 추가 학습한다. 이후 학습된 언어모델로 논문의 제목과 초록을 임베딩하고 이를 통해 학술지를 추천하는 기계

학습 분류 모델을 구축한다. 마지막으로 서비스 단계에서는 거대언어모델을 활용하여 이용자의 국영문 입력값에 대해 학술지를 추천할 수 있는 프로세스를 구성한다. 본 연구는 국내 학술 서비스를 통해 수집한 실제 데이터와 국내 논문에 대한 공개 데이터를 활용하여 제안한 추천 시스템의 성능을 평가한다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 연구에 관한 선행연구를 짚어본다. 제3장에서는 본 연구가 제안하는 방법을 상세히 소개한다. 제4장에서는 본 연구에서 수행한 실험의 구성에 관하여 기술한다. 제5장에서는 제3장과 제4장을 통해 실험한 결과를 해석하고, 마지막으로 제6장에서는 본 연구의 결론 및 추후연구에 관해 기술한다.

2. 배경연구

2.1 문서 분류

인공지능 기술이 발전함에 따라 문서 분류는 자연어 처리 분야의 주요 연구 주제로 자리 잡았다. 초기 문서 분류 연구에서는 주로 나이브 베이즈 혹은 서포트 벡터 머신과 같은 전통적인 기계학습 알고리즘이 활용되었다(Xu, 2018; Zhang et al., 2008). 이러한 기계학습 알고리즘은 텍스트 데이터에서 통계적인 특징을 추출하여 분류를 수행하였는데, 고차원의 희소 데이터 문제 및 자연어의 복잡성 문제로 인해 한계가 존재한다. 이후에는 앙상블 기반의 모델과 특징 선택 기법 등을 통해 더욱 정교한 기계학습 모델들이 문서 분류에 활용되었다(Chen et al., 2022; Kang et al., 2018; Silva et al., 2010).

최근에는 딥러닝을 기반으로 문서를 분류하는 연구도 다수 수행되었다. 합성곱 신경망과 순환 신경망은 텍스트에서 지역적 혹은 순차적 패턴을 포착하는 능력으로 많은 주목을 받았다(Chung & Sohn, 2020). 특히 BERT(Bidirectional Encoder Representations from Transformers) 및 GPT(Generative Pre-trained Transformer) 계열의 아키텍처들은 다양한 문서 분류 작업에서 뛰어난 성능을 보여주었다(Brown et al., 2020; Devlin et al., 2019). 이러한 트랜스포머 기반 모델들은 어텐션 메커니즘을 활용하여 타겟 시퀀스와 관련도가 높은 소스 시퀀스에 높은 가중치를 부여하는 방식으로 텍스트의 의미론적 맥락에 대한 이해를 크게 향상했다(Vaswani et al., 2017).

학술논문 데이터에 문서 분류를 적용한 연구들도 다양하게 수행되고 있다. Zhang et al.(2020)은 논문에 적합한 심사위원을 추천하는 다중 분류 모델을 제안하였다. 제안된 모델은 단어를 통해 문장을 표현하고 문장을 통해 문서를 표현하는 계층적 구조에 어텐션 메커니즘을 적용하여 강건하고 우수한 분류 성능을 도출하였다. Da et al.(2022)는 인용 문헌 추천을 위해 텍스트의 의미를 고려한 분류 모델을 제안하였다. 해당 연구에서는 LSTM(Long Short-Term Memory)를 통해 입력된 값과 논문의 텍스트를 각각 인코딩하고 다층 퍼셉트론을 통해 추천 여부를 분류하였다. 국내 사례로, 김판준(2022)은 국내 학술논문에 해당하는 주제를 식별하는 자동분류 방안을 제시하였다. 위 연구에서는 필터 방법에 해당하는 자질 순위화 기법인 문헌빈도, 자카드 계수, 카이제곱 통계량,

상대적 상호정보량, GSS 계수, 피어슨 계수, 상호정보량, 로그승산비를 통해 2002년부터 2015년 사이에 정보관리학회지에서 출판된 논문을 한국연구재단의 학술연구분야분류에 자동으로 할당하였다. 국희진 외(2024)는 학술논문의 문장 의미를 3개 상위 카테고리 및 9개 하위 카테고리로 분류하는 모델을 제안하였다. 제안된 모델은 문장 의미 레이블의 임베딩을 예측에 활용하였으며 문장 의미의 계층 구조를 반영하기 위해 계층적 손실함수를 적용하였다는 특징을 가진다.

2.2 학술지 추천

연구자가 논문에 적합한 학술지를 선택하는 것은 논문의 게재 가능 여부뿐만 아니라 연구 결과의 가시성과 영향력을 극대화하는 데에도 큰 영향을 미친다. 학술지 선정은 연구자의 주관성이 개입되며 심사 및 출판 소요기간이나 게재료와 같은 외부 인자에 영향을 받기 때문에, 학술지 추천은 학술 연구 분야의 다른 추천 과제보다 복잡도가 높다(Liu et al., 2022). 뿐만 아니라 다양한 학문 분야의 연구논문을 출판하는 다학제 학술지가 등장함에 따라 연구자들이 논문을 투고할 적절한 학술지를 선택하는 일이 보다 어려워졌다. 특히 연구 경험이 적은 신진 연구자들이나 새로운 분야로 연구 주제를 변경한 연구자들은 투고할 학술지를 찾는 것이 연구를 수행하는 것만큼이나 어려운 일이다(김용우 외, 2023).

연구 결과의 가시성과 영향력을 극대화하기 위하여 연구자들이 적절한 저널을 선택하는 데 도움을 줄 수 있는 데이터 기반의 학술지 추천 시스템을 제안하는 연구들이 활발히 수행되고 있다. Pradhan et al.(2020)은 논문의 초록, 제목, 키워드, 주제 분야, 저자, 학술지 정보를 활용하여 상위 K개 학술지를 추천하는 모델을 제안하였다. 해당 연구는 텍스트에 내재되어 있는 의미를 보다 정확하게 표현하기 위해 단어 수준의 어텐션을 계산하고 문장 수준의 어텐션을 계산하는 절차를 통해 논문의 특징을 추출하는 계층적 어텐션 네트워크를 활용하였다. Entrup et al.(2024)는 웹 기반의 오픈액세스 학술지 추천 시스템 B!SON을 제안하였다. 해당 시스템은 엘라스틱서치를 통해 제목과 초록의 유사도를 산출하고, 서지 결합 기법을 통해 인용 문헌 유사도를 산출한 뒤에 유사도 값을 더하여 추천 점수를 도출하는 방식으로 학술지를 추천한다. ZhengWei et al.(2022)는 논문의 제목과 초록을 임베딩하는 Doc2Vec 모델을 학습하고, 임베딩 벡터를 통해 학술지를 추천하는 XGBoost(eXtreme Gradient Boost)를 학습하는 두 단계의 학술지 추천 모델을 제안하였다. 이 연구는 논문의 저자, 분야, 인용 등의 메타데이터 없이 텍스트 데이터만으로도 높은 성능의 추천이 가능하다는 장점이 있다.

김용우 외(2023)는 국내 학술논문을 대상으로 한 학술지 추천 연구를 수행하였다. 해당 연구는 텍스트의 유사도 산출에 특화된 모델인 Sentence-BERT를 통해 초록 임베딩하여 벡터 유사도를 기준으로 학술지를 정렬하고, 제목과 저자 키워드를 통해 저널-키워드 매트릭스를 구축하고 학술지 유사도를 산출하여 학술지 리스트를 재정렬한 뒤, 1순위로 추천된 학술지와 가장 유사한 제목과 키워드 분포를 보이는 학술지를 추천 리스트에 추가하는 절차로

학술지 추천을 수행한다. 이 연구는 국문 논문을 대상으로 한 학술지 추천 방법을 최초로 제시하였지만 국문지와 영문지가 혼재된 국내 학술 서비스에 제안된 방법을 직접 적용하기는 어렵다는 한계점을 가진다.

3. 학술지 추천 시스템

본 연구에서 제안하는 학술지 추천 시스템은 크게 데이터셋 수집 및 처리, 학술지 추천 모델 구축, 학술지 추천 서비스 프로세스 구축으로 구성된다. 본 장에서는 각 구성에 대하여 상세히 기술한다.

3.1 데이터셋 구축

본 연구가 제안하는 시스템은 데이터셋을 구축하는 단계로부터 시작한다. 수집된 논문 데이터는 학술지, 국문 제목, 국문 초록, 영문 제목, 영문 초록, 발행일 등을 포함하며, 학술지 혹은 논문에 따라 국문 제목이나 국문 초록이 없는 경우가 존재할 수 있다. 수집한 데이터 중 영문 제목과 영문 초록이 없는 데이터는 분석에서 제외하였다. 본 연구는 영문 제목과 영문 초록을 하나의 문자열로 연결하여 분석에 활용한다.

모델 학습을 위한 데이터셋은 다음의 순서로 구축한다. 첫째, 추천 대상 학술지의 논문 수를 검토하여 적정한 논문 수 기준을 정한다. 기계학습을 통한 분류를 위해서는 레이블별 데이터 수에 균형을 유지하는 것이 중요하다. 학술지마다 발행된 논문의 수가 상이하기 때문에 추천 대상 학술지를 늘리면 논문 수가 적은 학술지가 포함되어 학습을 위한 데이터가 적어지고 이는 추천 성능의 하락을 초래한다. 반면 학술지별 논문 수 임계치를 높이면 추천 대상 학술지가 줄어들기 때문에 추천 시스템의 활용성이 떨어질 수 있다. 둘째, 학술지별 논문 수가 정해지면 추천 대상 학술지별로 가장 최근에 출판된 논문을 선별한다. 학술지마다 출판되는 논문의 전반적인 트렌드는 시간에 따라 조금씩 변화한다. 따라서 최신의 논문만을 학습 대상으로 사용함으로써 학술지별 최신의 연구 트렌드를 추천 모델에 반영할 수 있다. 셋째, 서비스 범위를 벗어나는 논문을 분류하기 위하여 더미 학술지를 추가한다. 더미 학술지에 포함되는 논문은 수집한 데이터에서 추천 대상 학술지에 포함되지 않은 학술지 논문을 임의로 선별하여 구성한다. 마지막으로 수집한 데이터를 레이블별 데이터 분포를 고려하여 학습 데이터와 테스트 데이터로 구분한다.

3.2 학술지 추천

본 연구는 과학기술 문헌의 의미론적인 문맥을 고려한 학술지 추천 시스템을 구축하기 위하여 BERT와 XGBoost를 사용한다. BERT는 트랜스포머 아키텍처의 인코더를 기반으로 하는 사전학습 언어모델로 입력된 텍스트의 문맥 정보를 양방향으로 학습하여 의미론적인 문맥을 효과적으로 표현한다는 특징을 가진다(김인후, 김성희, 2022). BERT는 분석 목적에 따라 모델을 추가로 학습했을 때 우수한 성능을 도출한다는 장점 때문에 자연어 처리 분야에서 널리 활용되고 있다.

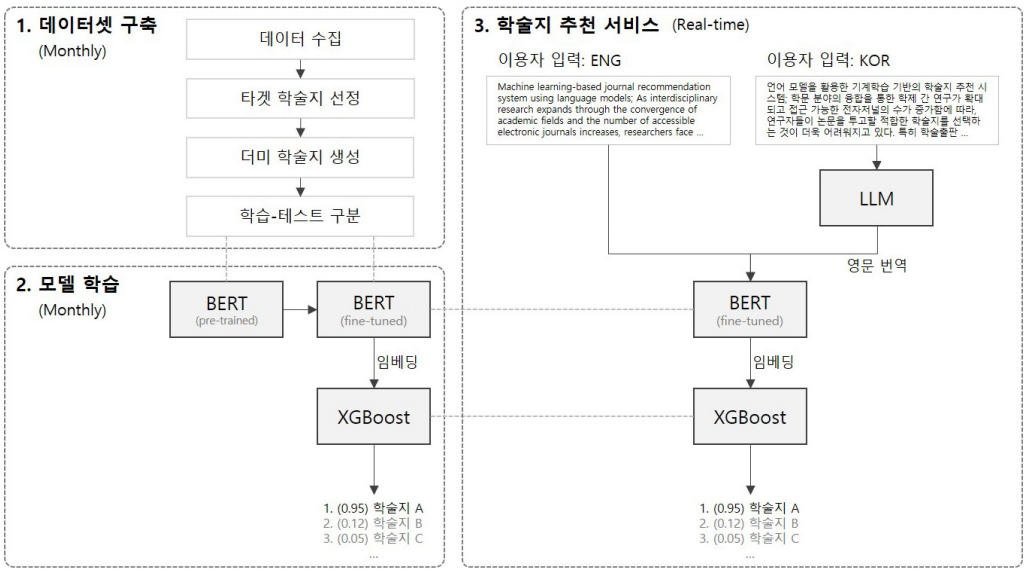
기존에 공개된 BERT 계열의 모델들은 해외 영문 텍스트 데이터를 기반으로 학습되었다. 본 연구에서 분석하고자 하는 국내 과학기술 논문의 문맥은 사전학습 모델의 가중치에 반영되어 있지 않은 상태이다. 따라서 본 연구는 수집한 학습 데이터를 통해 BERT 계열 사전학습 모델을 추가로 학습하여 국내 논문들의 문맥 정보를 더 정확하게 표현할 수 있도록 하였다. 이후 학습 데이터와 테스트 데이터를 미세조정된 BERT 모델에 입력하고, 모델의 마지막 레이어의 CLS 토큰에 해당하는 벡터를 추출하여 임베딩 데이터를 구성하였다.

임베딩된 벡터와 학술지 레이블을 통해 논문의 학술지를 예측하는 XGBoost 분류기를 학습한다. XGBoost는 병렬 처리, 트리 기반 구조, 그리고 정규화를 기반으로 한 과적화 방지를 통해 다른 기계학습 모델보다 일반적으로 우수한 예측 결과를 제공한다(Chen & Guestrin, 2016). 본 연구는 5-fold 교차검증과 그리드 서치를 통하여 XGBoost 모델의 학습률, 최대 트리의 깊이, 서브 샘플 비율, L1 정규화, L2 정규화 파라미터를 최적화한다.

3.3 서비스 아키텍처

학술지는 지속적으로 새로운 논문을 출판하기 때문에 학술지 추천 서비스도 주기적으로 업데이트되어야 한다. 특히 추천 서비스는 학술지별 최신 논문의 트렌드를 반영할 수 있어야 하고 발행된 논문의 수가 충분한 학술지는 추천 서비스의 대상으로 포함되어야 한다. 이러한 동적인 모델 업데이트를 반영한 학술지 추천 서비스 아키텍처는 〈그림 1〉과 같다.

전체 서비스는 크게 세 부분으로 구분된다. 데이터셋 구축 단계와 모델 학습 단계는 서비스를 위한 추천 모델을 구축하기 위해 사전에



〈그림 1〉 학술지 추천 서비스 프로세스

준비하고 학습하는 단계로, 앞선 두 절에서 설명한 바와 같다. 일반적으로 학술지들은 연간, 반연간, 계간, 격월간, 혹은 월간 발행주기를 따르므로 학술지별로 매달 새로운 데이터가 추가될 수 있다. 따라서, 데이터셋 구축 단계와 모델 학습 단계는 논문 데이터가 추가되는 시기에 맞추어 월 1회 업데이트된다.

실제 이용자 입력에 대하여 적절한 학술지를 추천하는 서비스 단계는 실시간으로 수행된다. 본 연구에서 제안한 방법은 영문을 처리할 수 있는 BERT 모델을 기반으로 하기 때문에 국문 입력을 직접 처리할 수 없다. 따라서 이용자가 국문 초록을 입력하는 경우 거대 언어모델을 활용해서 국문 초록을 영문으로 번역하는 과정을 거친다. 이용자의 영문 입력 혹은 영문으로 번역된 국문 입력은 학습이 완료된 BERT 모델과 XGBoost 분류기를 거쳐 학술지별 확률값으로 변환되고, 가장 높은 확률값을 가지는 학술지들이 이용자들에게 추천된다.

4. 데이터 및 실험 모델

본 연구는 학술지 추천 시스템의 실효성을 보이기 위해 두 가지 데이터를 제안된 방법에 적용해보는 실험을 진행한다. 본 장에서는 데이터의 종류 및 구성, 베이스라인 모델과 제시한 모델 및 파라미터에 대한 설명, 학술지 추천 시스템에 대한 평가 지표와 관련한 세부적인 내용을 기술한다.

4.1 데이터

우선, 실제 학술 서비스의 데이터를 활용하고자 한국과학기술정보연구원에서 운영 중인 국가오픈액세스플랫폼 AccessON[1]의 학술출판 지원 서비스를 사용하는 학술지 논문 데이터를 수집하였다. 데이터는 2024년 9월 20일 추출되었으며 60개 학술지에서 발행된 논문 총 23,107건의 제목, 초록 데이터를 포함한다. 이 중 분석에 활용할 수 있는 논문이 100건 이상인 학술지 32종을 선별하였고 선별된 학술지에서 가장 최근에 발행된 논문 100건씩을 데이터셋에 포함하였다. 그리고 선별되지 않은 학술지 중 논문 수가 많은 25종의 학술지에서 각각 4건의 논문을 랜덤으로 추출하여 더미 학술지를 구성하였다.

다음으로 한국과학기술정보연구원에서 공개한 국내 논문 전문 텍스트 데이터[2]를 수집하였다. 데이터는 935개의 학술지에서 발행된 논문 총 481,578건의 제목, 초록 등의 데이터를 포함한다. 수집한 공개 데이터로 총 두 가지 데이터셋을 구축하였다. 우선 발행된 논문이 300건 이상인 학술지 450종에서 가장 최근에 발행된 논문 300건씩을 선별하였다. 그리고 선별되지 않은 학술지에서 논문을 1건씩 랜덤 추출하여 300건의 논문을 가지는 더미 학술지를 구성하였다. 다음으로 발행된 논문이 100건 이상인 학술지 655종에서 가장 최근에 발행된 논문 100건씩을 선별하였다. 그리고 선별되지 않은 학술지에서 논문을 1건씩 랜덤 추출하여 100건의 논문을 가지는 더미 학술지를 구성하였다.

1) https://accesson.kisti.re.kr/
2) https://doi.org/10.23057/38

구축된 데이터셋은 80:20의 비율로 학습 데이터와 테스트 데이터로 구분하였다. 이때 균형된 학습 및 테스트 데이터를 구성하기 위하여 계층적 데이터 추출을 적용하였다. 구축 데이터셋에 대한 상세한 정보는 〈표 1〉과 같다.

4.2 비교 모델

본 추천 시스템의 핵심은 논문이 가지고 있는 문맥을 잘 반영할 수 있도록 텍스트를 벡터화하는 것이다. 따라서 본 연구는 전통적인 문서 표현 기법인 TF-IDF(Term Frequency-Inverse Document Frequency)와 Doc2Vec을 베이스라인으로 삼아 다양한 임베딩 기법을 기반으로 한 결과를 비교한다. 본 연구에서 활용된 임베딩 모델의 파라미터는 선행연구를 바탕으로 초기 설정되었으며 학습 결과를 바탕으로 일부 조정되었다.

(1) TF-IDF: TF-IDF는 단어의 문서 내 등장 빈도와 단어가 포함된 문서의 빈도를 기반으로 단어의 가중치를 계산하는 방법이다. 본 연구에서는 불용어, 3자 이하의 단어, 그리고 전체 문서의 0.1% 이하의 문서에서만 등장한 단어들을 제외한 52,480개의 단어를 기반으로 TF-IDF 가중치를 산출하였다. 이후 효율적인 학습을 위하여 오토인코더를 통해 TF-IDF 벡터의 차원을 축소하였다. 인코더는 총 4개 레이어로 52,480, 2,048, 1,024, 512, 128차원으로 축소되고, 디코더 역시 4개 레이어로 128, 512, 1,024, 2,048, 52,480차원으로 확대된다. 모델의 배치 크기와 학습률은 각각 64, 0.001로 설정되었으며 Adam 옵티마이저로 100 에폭에 걸쳐 학습되었다.

(2) Doc2Vec: Doc2Vec은 Word2Vec을 확장한 방법으로, 신경망 기반의 문서 임베딩 기법이다. 본 연구는 불용어와 3자 이하의 단어를 제거한 뒤 PV-DM(Distributed Memory Model of Paragraph Vector) 방식의 Doc2Vec 모델을 학습하였다. 모델의 임베딩 차원은 768, 윈도우는 15, 단어의 최소 등장 빈도는 5로 설정 후 100 에폭에 걸쳐 학습되었다.

(3) BERT: BERT는 트랜스포머 아키텍처의 인코더를 기반으로 한 사전학습 언어모델이다. 본 연구는 수집한 학습 데이터를 활용하여 BERT, RoBERTa, SciBERT 모델을 각각 추가 학습하였다. BERT는 가장 기본이 되는 모델이며, RoBERTa는 학습 시간, 데이터셋, 동적 마스킹 등의 최적화 기법을 통해 성능을 개선한 모델이다. SciBERT는 다양한 과학기술 문헌을 학습한 언어모델로 과학적 텍스트를 처리하는데 최적화된 모델이다. 각 모델은 최대 토큰 길이 512, 마스킹 토큰의 비율은 0.15, 배치 크기는 4로 설정하여 5 에폭에 걸쳐 추가 학습되었다.

〈표 1〉 학술지 추천 데이터셋

데이터셋	학술지별 논문 수	학술지 수[1)]	학습 데이터	테스트 데이터	출처
#1	100	33	2,640	660	AccessON
#2	300	451	108,940	27,060	AIDA
#3	100	656	52,570	13,210	AIDA

1) 더미 학술지 1개가 포함된 값임.

4.3 평가 지표

본 연구가 제안한 방법의 추천 성능을 평가하기 위하여 학술지 추천 시스템에서 널리 활용되는 두 가지 평가 지표를 사용한다.

(1) Acc@k: 이 지표는 시스템이 추천한 상위 k개 학술지에 실제 학술지가 포함되어있는 데이터의 비율을 의미한다(수식 1). 예를 들어 Acc@10은 XGBoost 모델이 가장 확률이 높을 것으로 예측한 10개 레이블 중 실제 정답이 포함되었을 때 예측에 성공하였다고 간주한다.

$$Acc@k = \frac{count_{top-k}}{count_{total}} \quad \text{(수식 1)}$$

(2) MRR(Mean Reciprocal Rank): 모델이 예측한 결과를 확률에 따라 내림차순으로 정렬하고, 실제 정답이 위치한 순위에 역수를 취한다. 이후 모든 테스트 데이터에 대하여 순위 역수 값의 산술평균을 계산하면 MRR 값을 얻을 수 있다(수식 2). 이 지표는 추천 시스템이 실제 정답을 얼마나 높은 순위로 예측하였는지를 평가할 수 있는 지표이다.

$$MRR = \frac{1}{|J|}\sum_{j \in J}\frac{1}{rank_j} \quad \text{(수식 2)}$$

5. 결과 분석 및 논의

5.1 학술지 추천 성능

각 데이터셋에 대한 모델별 성능은 〈표 2〉와 같다. 모든 실험에서 TF-IDF, Doc2Vec 모델을 활용했을 때보다 BERT 계열의 모델을 활용했을 때 성능이 더 우수했으며 BERT 계열

〈표 2〉 학술지 추천 성능평가 결과

데이터셋	모델	Acc@1	Acc@2	Acc@3	Acc@4	Acc@5	Acc@10	MRR
#1	TF-IDF	0.4685	0.6329	0.7035	0.7680	0.8157	0.9293	0.6206
	Doc2Vec	0.3810	0.5361	0.6175	0.6790	0.7465	0.8909	0.5414
	BERT	0.5818	0.7530	0.8136	0.8621	0.8955	0.9682	0.7185
	RoBERTa	**0.6212**	**0.7924**	**0.8561**	**0.9000**	**0.9273**	**0.9848**	**0.7535**
	SciBERT	0.5894	0.7409	0.8076	0.8455	0.8742	0.9606	0.7170
#2	TF-IDF	0.2215	0.3259	0.3982	0.4506	0.4943	0.6212	0.3507
	Doc2Vec	0.1409	0.2141	0.2644	0.3033	0.3388	0.4506	0.2424
	BERT	0.2789	0.3965	0.4743	0.5332	0.5761	0.7136	0.4177
	RoBERTa	**0.3715**	**0.5085**	**0.5863**	**0.6447**	**0.6868**	**0.8038**	**0.5141**
	SciBERT	0.3309	0.4579	0.5386	0.5941	0.6375	0.7667	0.4720
#3	TF-IDF	0.1685	0.2537	0.3107	0.3586	0.3932	0.5107	0.2796
	Doc2Vec	0.1569	0.2402	0.2956	0.3413	0.3769	0.4979	0.2681
	BERT	0.2528	0.3691	0.4424	0.4957	0.5355	0.6654	0.3871
	RoBERTa	**0.3098**	**0.4316**	**0.5054**	**0.5601**	**0.6008**	**0.7221**	**0.4449**
	SciBERT	0.2835	0.4022	0.4724	0.5213	0.5628	0.6883	0.4162

모델 중에서도 RoBERTa 모델을 기반으로 한 추천 시스템이 가장 좋은 추천 결과를 도출하였다. Acc@5를 기준으로 RoBERTa 모델은 TF-IDF, Doc2Vec 모델 대비 데이터셋 #1에서는 13.68%, 24.22%, 데이터셋 #2에서는 38.94%, 102.72%, 데이터셋 #3에서는 52.89%, 59.41% 더 높은 성능을 보였다.

데이터 수와 학술지 수를 기준으로 보았을 때 RoBERTa 기반 모델은 학술지 종 수가 가장 적은 데이터셋 #1에서 Acc@1, Acc@3, Acc@5가 각각 0.6212, 0.8561, 0.9273으로 높게 나타났다. 반면 학술지 수가 상대적으로 많은 데이터셋에 대해서는 추천 성능이 다소 떨어지는 양상을 보였다. 451종 학술지별로 각각 300건의 논문을 선별한 데이터셋 #2와 656종 학술지별로 각각 100건의 논문을 선별한 데이터셋 #3에 대한 추천 성능 비교를 통해 이를 확인할 수 있다. 이는 분류 모델이 학습해야 하는 학술지의 수는 10배 이상 늘어난 데 비해 학술지별 논문의 수는 크게 증가하지 않았기 때문에 학술지별 샘플이 상대적으로 적어지고 이로 인해 일반화 성능이 떨어진 결과이다. 더 많은 학술지를 추천 서비스 대상으로 포함하기 위해서는 조금 낮은 추천 성능을 감수해야 하며 높은 성능의 추천 서비스를 위해서는 대상 학술지의 수를 줄이고 데이터의 수를 늘려야 한다. 따라서 각 학술 서비스의 목적에 맞게 서비스 범위와 정확도의 상충관계에서 균형을 잘 유지할 필요가 있다.

〈표 3〉은 데이터셋 #1의 사례에 대하여 모델에 따른 학술지 추천 결과를 비교한 표이다. 첫 번째 사례 논문 “Word Embeddings-Based Pseudo Relevance Feedback Using Deep Averaging Networks for Arabic Document Retrieval”에 대하여 RoBERTa 기반 추천 모델과 SciBERT 기반 추천 모델은 정답 학술지인 “Journal of Information Science Theory and Practice”를 1순위로 식별하였다. 반면 BERT 기반 모델은 “SUVANNABHUMI”와

〈표 3〉 학술지 추천 결과 예시

사례 1	제목	Word Embeddings-Based Pseudo Relevance Feedback Using Deep Averaging Networks for Arabic Document Retrieval				
	학술지	Journal of Information Science Theory and Practice				
	모델	1순위	2순위	3순위	4순위	5순위
	BERT	X	X	O	X	X
	RoBERTa	O	X	X	X	X
	SciBERT	O	X	X	X	X

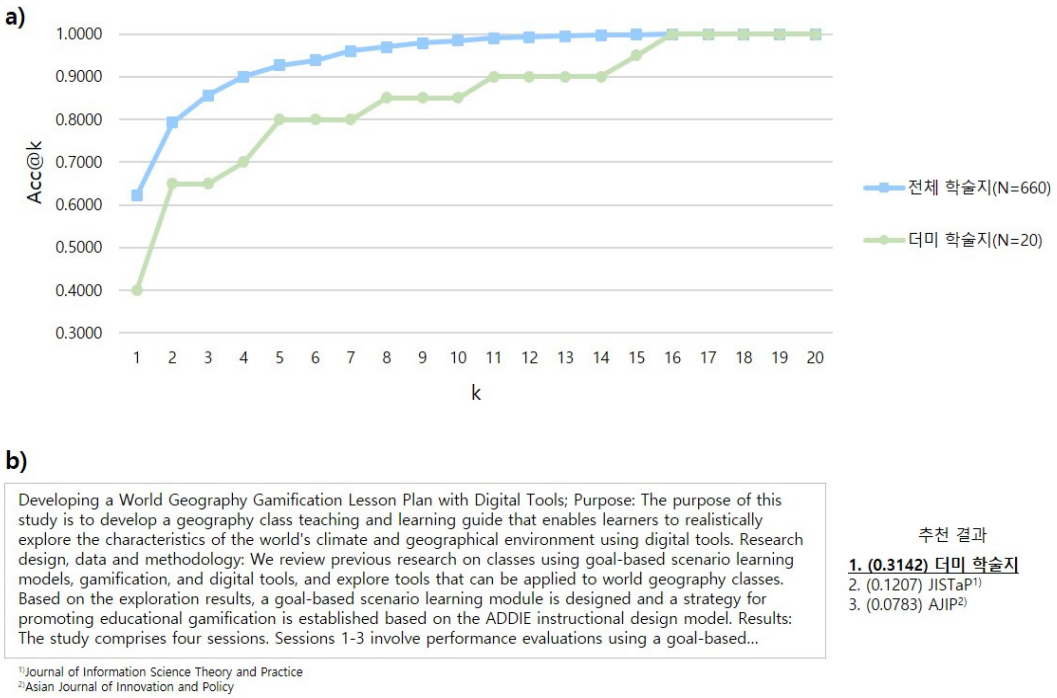
사례 2	제목	Efficient and Accurate Finite Difference Method for the Four Underlying Asset ELS				
	학술지	Journal of the Korean Society of Mathematical Education Series B: The Pure and Applied Mathematics				
	모델	1순위	2순위	3순위	4순위	5순위
	BERT	X	X	X	X	X
	RoBERTa	O	X	X	X	X
	SciBERT	X	O	X	X	X

“Asian Journal of Innovation and Policy”를 각각 1순위와 2순위로 식별하고 정답 학술지를 3순위로 식별하였다. 두 번째 사례 논문 “Efficient and Accurate Finite Difference Method for the Four Underlying Asset ELS”의 경우 RoBERTa 기반 모델은 1순위로 정답 학술지인 “Journal of the Korean Society of Mathematical Education Series B: The Pure and Applied Mathematics”를 식별했지만 SciBERT 기반 모델은 2순위로, BERT 기반 모델은 8순위로 정답 학술지를 식별하였다.

5.2 범위 외 논문에 대한 추천

본 연구는 학술 서비스의 대상 학술지 범위에서 벗어나는 논문들에 대한 추천을 고도화하기 위하여 더미 학술지를 추가하였다. 더미 학술지의 추천 결과는 〈그림 2〉와 같다. 그림 상단의 그래프는 더미 학술지에 해당하는 테스트 데이터에 대한 예측 성능을 나타내고 있다. 더미 학술지에 대한 Acc@1, Acc@3, Acc@5는 각각 0.4000, 0.6500, 0.8000이다. 더미 학술지는 서비스 대상 학술지와 비교했을 때 다양한 분야의 논문을 포함하기 때문에 상대적으로 정확한 추천이 어렵다. 그럼에도 불구하고 본 연구는 서비스 범위에서 벗어나는 논문을 충분히 구별해낼 수 있다는 점을 확인하였다.

본 연구에서 활용한 더미 학술지에 해당하는 논문들은 실제로 타겟 학술지에 게재되지 않은 논문이다. 따라서 실제로 타겟 학술지와



〈그림 2〉 범위 외 논문에 대한 추천 결과 및 예시

관련이 없을 수 있지만, 상당히 유사한 분야의 논문일 수 있다. 그림 하단의 예시를 보면, 첫 번째 예시는 “4차산업연구”의 4권 1호에 출판된 “Developing a World Geography Gamification Lesson Plan with Digital Tools”에 대한 예시이다. 해당 학술지는 본 연구의 추천 대상 학술지의 범위에 포함되지 않으며 모델이 이를 잘 예측한 것을 볼 수 있다. 하지만 두 번째와 세 번째로 추천된 학술지는 각각 정보학, 기술정책 분야의 학술지로 게임화나 디지털 툴에 대한 연구논문들을 출판한 이력을 가지고 있기 때문에 저자의 선택에 따라 투고해볼 수 있는 대안으로 고려해볼 수 있다.

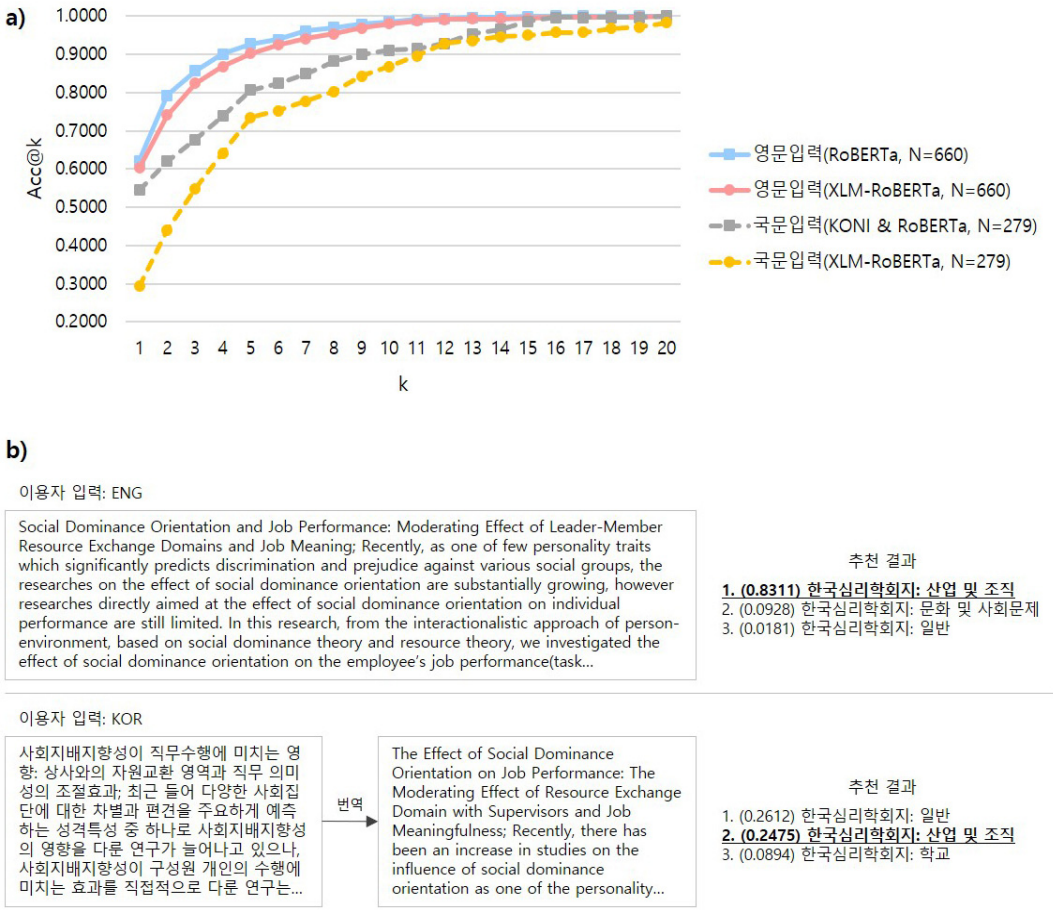
5.3 국문 논문에 대한 추천

본 연구에서 제안한 서비스 아키텍처의 실현 가능성을 보이기 위해 데이터셋 1의 국문 데이터를 활용하여 추가 실험을 진행하였다. 본 연구는 과학기술 분야 특화 거대 언어모델인 KONI(KISTI Open Natural Intelligence)를 활용하여 국문 제목과 초록을 영문으로 번역한 뒤에 학습된 RoBERTa 모델과 XGBoost 분류기에 번역된 텍스트를 입력하여 학술지 추천 결과를 도출하였다. 제안된 서비스 아키텍처의 성능을 비교하기 위하여 다국어 언어모델인 XLM-RoBERTa 모델을 동일한 방법으로 추가 학습하고 국문 제목과 초록을 모델에 입력하여 얻은 임베딩 벡터를 통해 XGBoost 분류기를 학습하였다.

데이터셋 #1의 테스트 데이터 660건 중 국문 초록이 존재하는 논문은 279건에 대하여 제안된 아키텍처를 적용한 결과 및 예시는 〈그림 3〉과 같다. 그림 상단의 그래프는 입력 언어에 따른 예측 성능을 비교한 결과를 나타내고 있다. 영문 입력에 대하여 RoBERTa 기반 추천 모델이 XLM-RoBERTa 기반의 추천 모델보다 더 높은 성능을 보였다. 국문 입력에 대해서도 KONI와 RoBERTa를 결합하여 활용한 경우가 XLM-RoBERTa를 사용했을 때보다 더 좋은 추천 결과를 도출하였다. 제안된 아키텍처의 국문 입력에 대한 Acc@1, Acc@3, Acc@5은 각각 0.5448, 0.6774, 0.8065로, XLM-RoBERTa 대비 약 85.37%, 23.53%, 9.75% 높은 성능이다.

그림 하단은 “한국심리학회지: 산업 및 조직”의 34권 3호에 출판된 “사회지배지향성이 직무수행에 미치는 영향: 상사와의 자원교환 영역과 직무 의미성의 조절효과”를 제안된 서비스 아키텍처에 적용한 예시를 나타내고 있다. 영문 제목과 초록을 기반으로 추천했을 때에는 정답 학술지를 0.8311의 높은 점수로 정확히 추천해준다. 반면 국문 제목과 초록을 번역한 뒤에 추천했을 때에는 정답 학술지가 0.2475의 점수로 2순위에 위치하였는데, 1순위 학술지의 점수와 큰 차이를 보이지 않으며 심리학 분야의 유사 저널들이 상위에 있는 결과가 도출되었다. 실제 학술지 추천 서비스 제공 시 이용자에게 추천 점수를 결과와 함께 제공함으로써 이용자가 투고할 학술지를 선택하기 위한 추가적인 정보로 활용할 수 있을 것이다.

연구자가 초록을 작성할 때 국문 초록과 영문 초록을 똑같이 작성하지는 않는다. 그뿐만 아니라 모델에 따라서 번역된 문장의 구조는 실제 사람이 작성하는 문장의 구조와 다소 차이가 있을 수 있다. 그럼에도 불구하고 본 연구는 거대 언어모델을 활용하여 이용자의 국문



〈그림 3〉 국문 입력에 대한 추천 결과 및 예시

입력에 대해서도 학술지 추천이 가능하다는 점을 확인하였다.

6. 결 론

본 연구는 국내 학술논문을 대상으로 학술지 추천 아키텍처를 제안하였다. 제안된 아키텍처는 더미 학술지 생성을 포함하는 데이터 구축 단계, 학술논문의 벡터 표현을 위한 BERT 추가 학습 및 논문 임베딩 벡터를 통해 학술지를 예측하는 XGBoost 분류기 학습을 포함하는 모델 학습 단계, 이용자의 입력 언어에 따라 국문 입력인 경우 거대 언어모델을 추가로 활용하여 실제 학술지를 추천해주는 서비스 단계로 구성된다. 본 연구는 제안된 아키텍처의 활용 가능성을 평가하기 위해 두 가지의 국내 학술논문 데이터를 활용한 실험을 진행하였다. 그 결과 전통적인 텍스트 마이닝 기법을 활용한 방법보다 RoBERTa 기반의 예측 프로세스가 가장 좋

은 추천 성능을 보였다. 또한 더미 학술지의 추천 결과와 국문 입력에 대한 추천 결과를 실제 사례를 통해 제시하여 제안된 아키텍처가 실제 국내 학술 서비스에 적용될 수 있음을 보였다.

본 연구는 국내 학술 생태계를 고려한 학술지 추천 시스템을 제시하였다는 기여점을 가진다. 국내 학회에서는 발행되는 학술지에는 국문지와 영문지가 혼재되어 있다. 학술지 추천 시스템 이용자가 입력하는 언어에 상관없이 서비스 범위 내에 존재하는 학술지를 모두 추천해줄 수 있어야 한다. 이러한 점을 고려해 본 연구는 영문 입력을 기반으로 추천 모델을 학습하였으며 국문 입력에 대한 처리도 가능한 아키텍처를 제안하였다. 제안된 아키텍처는 국문지와 영문지를 모두 서비스하는 국내 모든 학술 서비스에 즉각적으로 활용될 수 있을 것으로 기대된다. 또한 본 연구에서 제안한 아키텍처는 사전 학습된 거대 언어모델을 활용하고 모델을 직접 개발하거나 학습하지 않기 때문에 경제적인 이점을 가진다. 본 연구에서 추가 학습한 BERT 계열의 모델은 고가의 그래픽 처리 장치 없이도 추가 학습할 수 있을 뿐만 아니라 우수한 추천 성능을 도출할 수 있다. 그뿐만 아니라 제안된 아키텍처는 최소한의 이용자 입력만으로도 학술지를 추천해줄 수 있다. 마지막으로, 본 연구는 국내 연구자들의 연구 활동과 학술지의 운영 및 활용 확산에도 기여를 할 것으로 기대된다. 고도화된 학술지 추천 시스템을 통해 연구자들은 수행한 연구를 투고하기에 적절한 학술지를 더욱 쉽게 탐색할 수 있으며 학술지의 분야를 오인해서 발생하는 시간적 손실을 줄일 수 있다. 특히 신진 연구자 및 분야를 변경한 연구자들이 논문을 투고할 새로운 학술지를 발굴하는 데 도움을 줄 수 있을 것이다. 학술지 입장에서도 분야에 맞지 않아 게재 불가 판정되는 논문의 수가 줄어들어 학술지 편집위원들의 불필요한 노력을 최소화할 수 있다.

후속 연구에서는 다음과 같은 사항들을 고려하여 학술지 추천 시스템을 고도화할 수 있다. 첫째, 거대 언어모델을 기반으로 한 추천 시스템을 개발할 수 있을 것이다. 거대 언어모델은 BERT와 같은 일반 언어모델들 대비 월등히 많은 파라미터를 가지고 있기 때문에 자연어를 이해하고 처리하는 능력이 우수하다. 따라서 학술지 추천을 위해 구조화된 프롬프트를 개발하거나 별도의 데이터셋을 구축하여 학습한다면 더욱 일반화된 성능을 도출할 수 있을 것으로 기대된다. 하지만 앞서 기술한 것처럼 거대 언어모델을 활용 및 학습하기 위한 추가적인 비용이 들기 때문에 실제 학술 서비스에서 활용 가능한 추천 시스템을 제공하기 위한 효율화 연구를 수행할 필요가 있다. 둘째, 본 연구의 서비스 단계는 거대 언어모델을 활용하여 국문 입력을 영문으로 번역하고 이를 추천 모델에 입력하는 절차를 따른다. 사례 연구를 통해 국문 입력에 대한 적절한 추천이 가능하다는 점을 보였지만 추천 성능은 여전히 영문 입력 대비 낮은 상황이다. 따라서 상대적으로 적은 국문 데이터에 대해서도 더 높은 추천 성능을 도출할 수 있는 시스템을 개발하는 연구가 필요할 것으로 보인다. 셋째, 본 연구는 추천 대상 학술지를 선정하기 위해 최소한의 논문 수를 제한하였다. 따라서 출판된 논문의 수가 충분하지 않은 학술지는 추천 대상에 포함할 수 없다는 한계점을 가진다. 추후에는 본 연구가 제안한 아키텍처와 전통적인 내용 기반 추천 방

법을 결합하는 하이브리드 방식의 추천 시스템에 관한 연구를 통해 콜드 스타트 문제를 완화할 수 있을 것이다. 넷째, 본 연구는 범위 외 논문에 대한 적절한 추천 결과를 제공하고자 하는 첫 시도이다. 하지만 추천 결과로 제시한 사례의 수가 적어 그 결과를 일반화하기에는 어려움이 있다. 추후연구에서는 데이터의 개수를 확장하고 구체적인 사례를 추가 분석하여 추천 시스템별 결과를 비교하는 연구를 수행해볼 수 있을 것이다. 마지막으로, 본 연구가 제시한 방법과 선행연구의 학술지 추천 성능을 통계적으로 비교하여 추천 시스템의 타당성을 검증하는 연구를 수행할 수 있을 것으로 사료된다.

참 고 문 헌

국희진, 김영화, 윤세휘, 강병하, 신유현 (2024). 계층적 표현 및 레이블 임베딩을 활용한 국내 논문 문장 의미 분류 모델. 정보과학회논문지, 51(1), 41-48. https://doi.org/10.5626/JOK.2024.51.1.41

김용우, 김대영, 서현희, 김영민 (2023). Sentence BERT를 이용한 내용 기반 국문 저널추천 시스템. 지능정보연구, 29(3), 37-55. https://doi.org/10.13088/jiis.2023.29.3.037

김인후, 김성희 (2022). 딥러닝 기반의 BERT 모델을 활용한 학술 문헌 자동분류. 정보관리학회지, 39(3), 293-310. https://doi.org/10.3743/KOSIM.2022.39.3.293

김판준 (2022). 자질선정을 통한 국내 학술지 논문의 자동분류에 관한 연구. 정보관리학회지, 39(1), 69-90. https://doi.org/10.3743/KOSIM.2022.39.1.069

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

Chen, H., Wu, L., Chen, J., Lu, W., & Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. Information Processing & Management, 59(2), 102798. https://doi.org/10.1016/j.ipm.2021.102798

Chen, T. & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. https://doi.org/10.1145/2939672.2939785

Chung, P. & Sohn, S. Y. (2020). Early detection of valuable patents using a deep learning model: case of semiconductor industry. Technological Forecasting and Social Change, 158, 120146. https://doi.org/10.1016/j.techfore.2020.120146

Da, F., Kou, G., & Peng, Y. (2022). Deep learning based dual encoder retrieval model for citation recommendation. Technological Forecasting and Social Change, 177, 121545. https://doi.org/10.1016/j.techfore.2022.121545

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 4171-4186. https://doi.org/10.18653/v1/N19-1423

Dinh, T. N., Pham, P., Nguyen, G. L., Nguyen, N. T., & Vo, B. (2025). A hybrid citation recommendation model with SciBERT and GraphSAGE. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 55(2), 852-863. https://doi.org/10.1109/TSMC.2024.3490774

Entrup, E., Eppelin, A., Ewerth, R., Hartwig, J., Tullney, M., Wohlgemuth, M., & Hoppe, A. (2024). Comparing different search methods for the open access journal recommendation tool B!SON. International Journal on Digital Libraries 25, 505-516. https://doi.org/10.1007/s00799-023-00372-3

Kang, M., Ahn, J., & Lee, K. (2018). Opinion mining using ensemble text hidden Markov models for text classification. Expert Systems with Applications, 94, 218-227. https://doi.org/10.1016/j.eswa.2017.07.019

Li, X., Wang, M., & Liu, X. (2024). Predicting collaborative relationship among scholars by integrating scholars' content-based and structure-based features. Scientometrics, 1-20. https://doi.org/10.1007/s11192-024-05012-4

Liao, W., Zhu, Y., Li, Y., Zhang, Q., Ou, Z., & Li, X. (2024). RevGNN: Negative Sampling Enhanced Contrastive Graph Learning for Academic Reviewer Recommendation. ACM Transactions on Information Systems, 43(1), 1-26. https://doi.org/10.1145/3679200

Liu, C., Wang, X., Liu, H., Zou, X., Cen, S., & Dai, G. (2022). Learning to recommend journals for submission based on embedding models. Neurocomputing, 508, 242-253. https://doi.org/10.1016/j.neucom.2022.08.043

Liu, X., Wu, K., Liu, B., & Qian, R. (2023). HNERec: Scientific collaborator recommendation model based on heterogeneous network embedding. Information Processing & Management, 60(2), 103253. https://doi.org/10.1016/j.ipm.2022.103253

Lu, Y., Yuan, M., Liu, J., & Chen, M. (2023). Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion. Scientometrics, 128(2), 1367-1393. https://doi.org/10.1007/s11192-022-04566-5

Pradhan, T., Gupta, A., & Pal, S. (2020). HASVRec: A modularized Hierarchical Attention-based Scholarly Venue Recommender system. Knowledge-Based Systems, 204, 106181. https://doi.org/10.1016/j.knosys.2020.106181

Silva, C., Lotric, U., Ribeiro, B., & Dobnikar, A. (2010). Distributed Text Classification With an Ensemble Kernel-Based Learning Approach. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(3), 287-297. https://doi.org/10.1109/TSMCC.2009.2038280

Tan, S., Duan, Z., Zhao, S., Chen, J., & Zhang, Y. (2021). Improved reviewer assignment based on both word and semantic features. Information Retrieval Journal, 24(3), 175-204. https://doi.org/10.1007/s10791-021-09390-8

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. Advances in Neural Information Processing Systems, 30.

Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. Journal of Information Science, 44(1), 48-59. https://doi.org/10.1177/0165551516677946

Zhang, D., Zhao, S., Duan, Z., Chen, J., Zhang, Y., & Tang, J. (2020). A Multi-Label Classification Method Using a Hierarchical and Transparent Representation for Paper-Reviewer Recommendation. ACM Transactions of Information Systems, 38(1), 1-20. https://doi.org/10.1145/3361719

Zhang, W., Yoshida, T., & Tang, X. (2008). Text classification based on multi-word with support vector machine. Knowledge-Based Systems, 21(8), 879-886. https://doi.org/10.1016/j.knosys.2008.03.044

Zhang, Z., Patra, B. G., Yaseen, A., Zhu, J., Sabharwal, R., Roberts, K., Cao, T., & Wu, H. (2023). Scholarly recommendation systems: A literature survey. Knowledge and Information Systems, 65(11), 4433-4478. https://doi.org/10.1007/s10115-023-01901-x

ZhengWei, H., JinTao, M., YanNi, Y., Jin, H., & Ye, T. (2022). Recommendation method for academic journal submission based on doc2vec and XGBoost. Scientometrics, 127(5), 2381-2394. https://doi.org/10.1007/s11192-022-04354-1

• **국문 참고자료의 영어 표기**
(English translation / romanization of references originally written in Korean)

Kim, Inhu & Kim, Seonghee (2022). Automatic classification of academic articles using BERT model based on deep learning. Journal of the Korean Society for Information Management, 39(3), 293-310. https://doi.org/10.3743/KOSIM.2022.39.3.293

Kim, Pan Jun (2022). An experimental study on the automatic classification of Korean journal articles through feature selection. Journal of the Korean Society for Information Management, 39(1), 69-90. https://doi.org/10.3743/KOSIM.2022.39.1.069

Kim, Yongwoo, Kim, Daeyoung, Seo, Hyunhee, & Kim, Young Min (2023). Conent-based Korean journal recommendation system using Sentence BERT. Journal of Intelligence and Information Systems, 29(2), 37-55. https://doi.org/10.13088/jiis.2023.29.3.037

Kook, Heejin, Kim, Yeonghwa, Yoon, Sehui, Kang, Byungha, & Shin, Youhyun (2024). Hierarchical representation and label embedding for semantic classification of domestic research paper. Journal of KIISE, 51(1), 41-48. https://doi.org/10.5626/JOK.2024.51.1.41