Journal of KIIT. Vol. 23, No. 1, pp. 1-11, Jan. 31, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.1.1

저자원 텍스트분류의 성능 향상을 위한 KoBERT기반 카테고리 매핑과 LLM 결합 연구

곽지호*, 정유철**

Enhancing the Performance of Low-Resource Text Classification through KoBERT-based Category Mapping and LLM Integration

Jiho Gwak*, Yuchul Jung**

이 연구는 국립금오공과대학교 대학 연구과제비로 지원되었음(2022~2024)

요 약

학습 데이터가 충분하지 않은 분류 테스크는 여전히 도전적인 문제이다. 이에 본 연구는 KoBERT 기반의 미세 조정 학습과 대형 언어 모델(LLM) 기반 제로샷 분류 방법을 결합하는 기법을 제안한다. 과학기술표준분류와 미래유망신기술(6T) 체계를 대상으로, KoBERT를 미세 조정하여 과학기술표준분류 모델을 구현하고 두 체계 간 매핑 전략을 수립하였다. 매핑이 불가능한 세부 카테고리에서는 대형 언어 모델 기반 제로샷 분류를 적용하고, 설명문 기반 프롬프트와 검증 프롬프트 엔지니어링을 통해 분류 결과 성능과 신뢰성을 향상시켰다. 실험 결과, 제안 기법이 학습 데이터가 부족한 상황에 대해 효과적으로 적용될 수 있음을 확인하였다. 이를 통해 학습 데이터가 충분하지 않은 분류 테스크를 효과적으로 해결할 수 있는 방안을 제시하였다.

Abstract

Text classification tasks with insufficient training data remain a challenging problem. To address this, we propose a method that combines KoBERT-based fine-tuning with Large Language Model(LLM)-based zero-shot classification. Focusing on the Science and Technology Standard Classification and Future Emerging Technologies(6T) systems, we fine-tuned KoBERT to implement a model for the Science and Technology Standard Classification and established a mapping strategy between the two systems. For subcategories where mapping was not feasible, zero-shot classification was applied. We also employed explanation-based and verification prompts to enhance the reliability of classification results. Experimental results confirmed that the proposed method can be highly effective in scenarios with limited training data. This provides an approach for addressing classification tasks with insufficient data.

Keywords

large language model, text classification, multi-label, KoBERT, mapping, prompt engineering

* 국립금오공과대학교 컴퓨터공학과 학사과정
- ORCID: https://orcid.org/0009-0000-7524-0357
** 국립금오공과대학교 인공지능공학과 부교수(교신저자)
- ORCID: http://orcid.org/0000-0002-8871-1979

· Received: Sep. 06, 2024, Revised: Oct. 01, 2024, Accepted: Oct. 04, 2024

· Corresponding Author: Yuchul Jung
Dept. of AI Engineering, Kumoh National Institute of Technology,
61 Daehak-ro (yangho-dong), Gumi, Gyeongbuk, [39177] Korea
Tel.: +82-54-478-7536, Email: jyc@kumoh.ac.kr

Ⅰ. 서 론

텍스트 분류는 텍스트를 기준에 맞게 체계적으로 나누고 정리하는 과정으로, 데이터를 이해하는데 있어 필수적인 요소이다. 분류 클래스는 데이터의 전반적인 특성을 드러내는 중요한 역할을 하며, 이를 통해 데이터의 패턴이나 경향을 빠르게 파악할 수 있다. 과거에는 하나의 분류 체계가 과학 기술, 사회, 정책 문서 등, 다양한 도메인을 모두 포함하는 표준화된 틀로 사용되었다. 그러나 지속적인 기술의 발전으로 인해 데이터의 다양성이 증가함에 따라, 각 도메인의 복잡성이 증가하고 전문화되면서 단일한 분류 체계에 의존하는 접근 방식은 한계[1]를 드러내고 있다. 이러한 변화로 인해 새로운 분류 체계가 형성되었으며, 각각은 특정 도메인의 추가적인 요구와 특성에 맞추어 구성되었다. 그러나 이러한 분류 체계 간에는 기존의 체계와 중첩되는 영역이 존재하며, 이는 특정 데이터가 하나의 분류 체계에만 속하지 않고 여러 분류 체계의 클래스에 포함될 수 있음을 의미한다.

최근, 텍스트 분류 작업에서는 BERT와 같은 사전 학습 모델을 지도 학습 기반으로 미세 조정(Fine-tuning)하는 방법[2]이 많이 사용되었다. 이 방법은 높은 성능을 보이지만, 충분한 학습 데이터를 필요로 한다는 한계를 가지고 있다. 따라서, 학습 데이터가 충분하지 않은 새로운 분류 체계의 경우, 분류 모델을 효과적으로 구축하는 것이 어려웠다.

최근 OpenAI사의 ChatGPT[3]와 같은 대형 언어 모델(LLM, Large Language Model)의 등장으로 자연어 처리의 여러 세부 분야에서 획기적인 발전을 이루었다. 대형 언어 모델은 새로운 작업에 대한 적응 능력이 뛰어나며, 제한된 상황에서도 인상적인 성능[4]을 보인다. 이러한 모델의 다양한 능력 중 하나는 제로샷 및 퓨샷 분류로, 이는 학습 데이터를 제공하지 않거나 최소한의 데이터를 제공한 상태에서도 분류, 추론, 생성 등의 작업을 수행할 수 있는 능력이다. 이와 같은 대형 언어 모델 기반 제로샷 및 퓨샷 텍스트 분류의 성능을 최대한 끌어내기 위해서는 프롬프트 엔지니어링이 중요한 역할[5]을 한다.

본 논문에서는 학습 데이터가 거의 존재하지 않는 분류 체계를 위해 학습 자원이 충분한 분류 체계를 활용하는 방안을 제시한다. 여러 분류 체계 중, 국내 과학 기술 분야의 분류에서 많이 사용되는 분류 체계인 과학기술표준분류[6] (학습 자원이 충분한 분류 체계)와 차세대 산업을 상징하는 미래유망신기술(6T)[7] (학습 자원이 거의 없는 분류 체계)를 선정·활용하였다. 이를 위해 KoBERT와 같은 사전 학습 모델 기반으로 지도 학습을 통해 과학기술표준분류 모델을 구현하였다. 또한 해당 모델을 바탕으로 미래유망신기술(6T) 분류 체계와 매핑 가능한 부분을 식별하여, 공통부분의 매핑 테이블 구축을 위한 매핑전략을 도출하였다. 매핑이 불가능한 부분에 대해서는 대형 언어 모델 기반의 제로샷 분류를 시도하였으며, 특히 분류를 위한 설명문 기반 프롬프트와 분류 결과의 신뢰성을 높이기 위한 검증 프롬프트 엔지니어링을 적용하였다.

최종적으로, 이러한 방법들을 바탕으로 학습 자원이 없어 분류가 어려웠던 미래유망신기술(6T)에 대한 분류의 성능과 적절성을 실험을 통해 검증하고, 실험 결과의 분석 및 한계점을 설명한다.

본 논문의 구조는 다음과 같이 구성된다. 2장에서는 연구와 관련된 배경 지식과 기존 연구들을 검토한다. 3장에서는 과학기술표준분류 체계를 바탕으로 한 지도 학습 모델이 구현과 그에 따른 매핑 과정, 제로샷 분류를 위한 프롬프트 엔지니어링을 설명한다. 4장에서는 대형 언어 모델을 이용한 제로샷 분류 및 프롬프트 엔지니어링을 통한 분류 신뢰도 향상 방법을 다룬다. 마지막으로 5장에서는 본 연구에서 제안한 방법의 실험 결과를 분석하고, 미래유망신기술(6T)분류의 성능을 평가한 후, 본 연구의 한계점과 향후 연구 방향을 제시한다.

Ⅱ. 관련 연구

2.1 BERT

BERT(Bidirectional Encoder Representations from Transformers)[8]는 구글에서 개발한 자연어 처리 모델로, 트랜스포머 인코더(Transformer encoder) 부분을 기반으로 하며, 주어진 문맥의 양방향 정보를 모두 활용해 의미 파악을 가능하게 한다.

BERT는 위키피디아와 BookCorpus 등 대규모 텍스트 말뭉치를 활용하여 사전 학습(Pre-training) 되었다. 이를 기반으로 SKT에서 한국어 데이터를 활용해 사전 학습된 한국어 모델인 KoBERT[9]를 개발한 바 있다. 이러한 사전 학습 모델은 다양한 언어처리 작업에 대해 미세 조정[10]이 가능하다. 예를 들어 텍스트 분류(Classification)[11], 질의응답(Question ans wering), 개체명 인식(NER, Named Entity Recognit ion)등 다양한 작업에 특화된 모델을 생성할 수 있다. 특히 텍스트 분류 작업에서 BERT 기반 모델은 미세 조정을 통해 뉴스 기사의 제목 분류 연구[12]와 문학 감정 분류 연구[13]에서 그 유용성이 입증된 바 있다.

그러나, 미세 조정을 기반으로 하는 텍스트 분류 방식은 충분한 학습 데이터가 필요하며, 학습 데이터가 충분치 않은 분류 체계에 적용하기 위해 미세 조정하는 경우, 성능저하 또는 과적합의 문제가 발생할 수 있다. 이에 따라, 본 연구에서는 분류 체계 중 학습 데이터가 충분한 분류 체계를 통해 KoBERT 모델을 학습하고, 이를 학습 데이터가 없는 분류 체계의 분류에 적용하고, 매핑 테이블을 구축하여 분류에 활용하는 방안을 제시한다.

2.2 대형 언어 모델 제로샷 분류

대형 언어 모델(LLM)의 제로샷 분류(Zero-shot classification)와 퓨샷 분류(Few-shot classification)는 사전 학습된 모델이 추가적인 학습 없이 새로운 데이터에 대해 분류 작업을 수행하는 능력을 의미한다. 제로샷 분류는 학습 데이터를 전혀 제공하지 않고 분류 작업을 수행하는 반면, 퓨샷 분류는 소량의 데이터를 제공하여 분류 작업을 수행하는 방식이다. 그중에서도 원샷 분류(One-shot classification)는 단 하나의 데이터만을 제공하여 분류 작업을 수행하는 경우이다.

이러한 대형 언어 모델은 방대한 양의 텍스트데이터를 바탕으로 다양한 언어 패턴과 지식에 대해 학습하게 된다. 수십억 개의 파라미터(Parameter)를 가지는 모델들은 자연어 처리 작업에서 뛰어난 성능을 발휘할 수 있다. 예를 들어, OpenAI사에서 개발한 대형 언어 모델인 GPT-4는 이러한 특징을 보여주는 대표적인 예시이다. GPT-4는 복잡한 문맥을 이해하고, 다양한 언어적 표현을 분석하는 능력이 뛰어나 새로운 작업에 대한 적응력이 높다. 이는 모델이 사전에 학습한 광범위한 언어 지식을 바탕으로 학습 데이터를 통한 별도의 학습 과정 없이 새로운 작업에 효과적으로 활용 가능하다는 것을 의미한다. 본 연구에서는 학습 데이터가 없는 분류 체계의 분류를 목적으로 하기 때문에, 퓨샷이나 원샷 방식을 사용하지 않고 제로샷 방식을 활용하여 대형 언어 모델 분류를 진행하였다.

최근 연구에서는 긍정, 부정, 중립이라는 세 가지 레이블을 가지는 트윗 데이터와 페이스북 댓글 데이터를 대상으로 대형 언어 모델의 제로샷 분류 성능이 검증되었다[14]. 이 연구는 대형 언어 모델이 별도의 추가 학습 없이도 문장을 높은 정확도로 분류할 수 있음을 입증하였으며[15], 대형 언어 모델의 일반적인 언어 이해 능력을 확인할 수 있었다. 그러나 다중 레이블을 가지는 복잡한 데이터셋에 대해서는 여전히 사전 학습 모델(PLM, Pre-train ed Language Model)의 미세 조정 방식이 대형 언어 모델의 제로샷 분류 능력보다 우수한 성능을 보인다는 연구 결과도 존재한다[16]. 이러한 결과는 대형 언어 모델의 분류 성능이 특정 작업이나 데이터셋의 특성에 따라 크게 달라질 수 있음을 의미한다.

2.3 텍스트 분류에서의 프롬프트

프롬프트는 대형 언어 모델이 주어진 작업에 대해 원하는 응답을 이끌어내기 위해 사용되는 중요한 도구로, 프롬프트 엔지니어링은 이러한 프롬프트를 효과적으로 설계하여 모델의 성능을 극대화하는 과정을 의미한다. 자연어 처리(NLP) 분야에서 프롬프트의 중요성은 날로 증가[17]하고 있으며, 특히 사전 학습된 언어 모델의 활용이 보편화됨에 따라 프롬프트 설계는 모델 성능에 중대한 영향을 미치고 있다. 기존의 NLP 패러다임은 사전 학습 모델의 미세 조정에서 사전 학습 모델의 프롬프트 접근 방향으로 집중[18]되고 있다. 이와 같은 변화는 모델의 활용 가능성을 한층 더 확장시키고 있으며, 프롬프트의 설계와 적용 방식이 모델 성능에 결정적인 역할을 한다는 의미로 해석될 수 있다.

최근 연구에서는 프롬프트의 구성 방식에 따라 모델의 분류 성능이 크게 달라질 수 있다는 것[19]과 CoT(Chain of Thought Prompting)[20]와 같은 단계별 추론 과정[21]의 프롬프트 기법을 적용하여 모델의 정확한 답을 도출하는 연구 등에서 프롬프트의 효과가 입증되고 있다. 본 연구에서는 대형 언어 모델의 분류 작업에서 설명문을 기반으로 한 프롬프트 방법을 제시한다.

III. 방법론

본 섹션은 한정적인 학습 데이터셋으로 인해 학습 제한되는 분류 체계에 대한 효과적인 분류를 위해 미세 조정된 KoBERT[9] 모델과 설명문을 기반으로 한 대형 언어 모델 제로샷 분류 기법의 결합 방식을 진행함에 사용된 분류 체계 및 방법론에 대한 설명이다. 1) 선정한 분류 체계, 2) 과학기술표준 분류 데이터를 활용한 KoBERT 모델의 미세 조정, 3) 과학기술표준분류와 미래유망신기술 매핑 테이블 구축, 4) 관계에 따른 분류 (종속적 관계분류 및 독립적 관계분류), 그리고 5) 자동 검증기법에 관해 설명한다. 그림 1은 방법론에 대한 전반적인 개요도이다.

3.1 분류 체계

본 연구에서는 과학기술 분야에서의 분류 작업을 수행하기 위해, 국가과학기술의 기획·평가·관리의 기본체계로 활용되는 한국과학기술기획평가원(KISTEP)의 과학기술표준분류[6] 분류 체계 그리고 중소벤처기업부의 미래유망신기술(6T)[7] 분류 체계를 사용하였다. 표 1은 각 분류 체계를 구성하는 대분류와 중분류 수를 보여주며, 분류 체계 구성의 기준을 나타낸다.

표 1. 분류 체계의 수와 기준
Table 1. Number and standards of classification systems

Classification system	Major categories	Medium categories	Classification system criteria
Science and technology standard classification	17	188	Organizes research topics and fields related to science and technology
Future promising new technologies (6t)	6	23	Focuses on cutting-edge technology industries

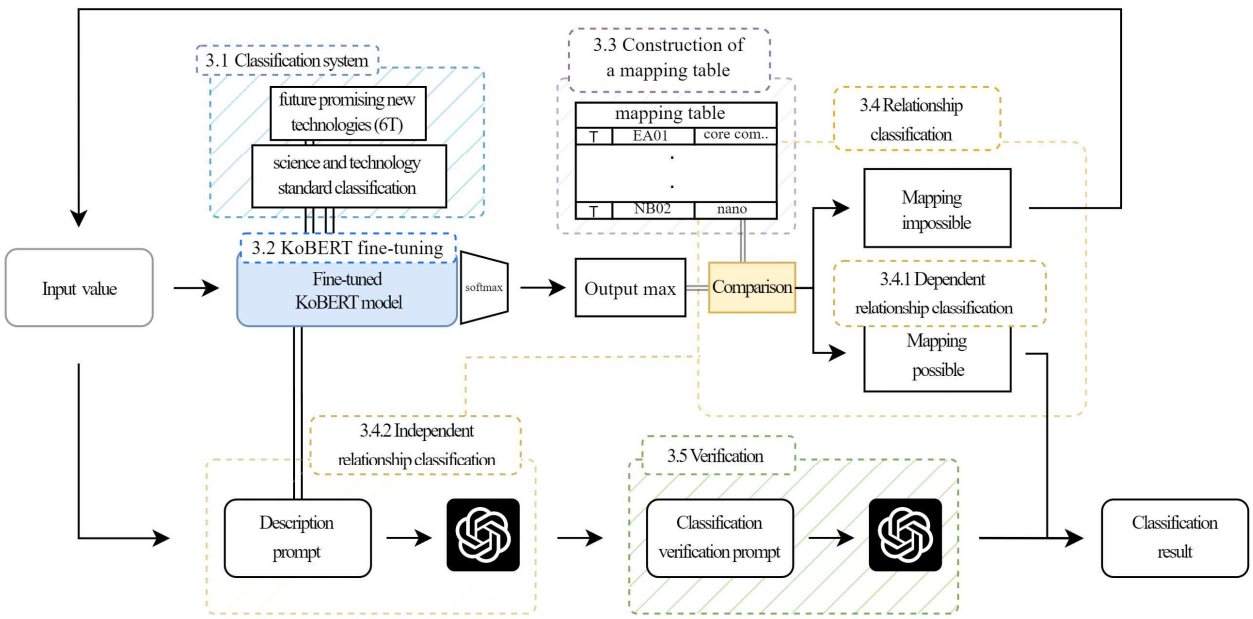

그림 1. 방법론 개요도
Fig. 1. Methodology overview

3.1.1 과학기술표준분류

과학기술표준분류는 과학기술정보통신부(MSIT)와 한국과학기술정보연구원(KISTI)에서 관리하는 분류 체계이다. 해당 분류 체계는 과학 기술의 연구 주제와 분야를 정리하기 위한 중요한 역할을 한다.

3.1.2 미래유망신기술(6T)분류

미래유망신기술(6T) 분류는 정보통신, 생명공학, 나노, 환경공학, 항공우주, 문화콘텐츠 등을 포함하는 6개의 첨단기술산업을 기반으로 구성된 분류 체계이다. 해당 분류 체계는 6개의 대분류, 23개의 중분류로 구성되어 있으며, 분류 체계의 세부 구성 요소는 표 2에 기재되어 있다.

표 2. 미래유망신기술(6T) 분류 체계
Table 2. Future emerging technologies (6T) systems

6T (Future promising new technologies) classification system components			
IT	Core components	ST	Satellite technology
	Next-generation network base		Launch vehicle technology
	Information processing systems and s/w		Aircraft technology
	Other information technologies		Other
BT	Basic·fundamental technology	ET	Environmental base
	Health and medical related applications		Energy
	Agriculture·marine·environment related applications		Clean production
			Marine environment
NT	Nano devices and systems	CT	Cultural contents
	Nanomaterials		Lifestyle culture (cyber communication)
	Nano bio health		Cultural heritage
	Nano base·process		

3.2 KoBERT 미세 조정

서로 다른 분류 체계를 매핑하기 위해 과학기술표준분류 모델을 구축하였다. 이 모델은 KoBERT 모델을 기반으로 하며, AI Hub에서 제공하는 과학기술표준분류 대응 특허 데이터[22]를 활용하여 지도 학습을 진행하였다. 데이터셋은 총 300,240건으로 구성되어 있으며, 이를 8:2 비율로 나누어 80%는 학습 데이터로, 나머지 20%는 평가 데이터로 사용하였다. 188개의 중분류 모델을 구축하기 위해, 사전 학습된 KoBERT 모델의 출력층에 188개의 Dense Layer를 추가하고, Softmax 활성화 함수를 사용하여 각 클래스에 대한 확률 분포를 계산할 수 있도록 설계하여 미세 조정을 진행하였다. 해당 모델은 입력된 텍스트가 확률적으로 어느 중분류에 속하는지 예측하는 역할을 하며, 매핑 테이블 구축 및 매핑 가능 여부 판단을 위해 사용된다. 학습 시 사용된 주요 파라미터와 환경은 표 3에 기재되어 있다.

표 3. KoBERT 학습 파라미터 및 환경
Table 3. KoBERT training parameters and environment

epoch	12
batch_size	16
max_length	256
optimizer	Adam
learning_rate	5e-05

3.3 매핑 테이블 구축

매핑 테이블은 서로 다른 분류 체계 간 유사한 클래스의 연결 정보를 가지며 과학기술표준분류와 미래유망신기술(6T) 분류 체계가 1대 다의 관계를 가질 수 있는 테이블이다. 클래스의 연결은 다음과 같은 과정을 통해 진행되었다. 임곗값을 {95, 90, 85, 70, 60} 단위로 설정하고, 각 임곗값(T)과 출력값의 최댓값(M)을 비교하여 클래스의 고윳값에 대응한다. 임곗값은 클래스별로 고유한 값을 매핑하기 위한 척도이며 임곗값을 설정하지 않을 때 잘못된 매핑이 발생될 가능성이 있으므로 단계별로 임곗값을 설정하였다. 그림 2는 매핑 테이블을 구축하는 주요 과정을 순서도로 나타낸 것이다. 새로운 입력이 발생하면 우선 미세 조정된 KoBERT 모델에 의해 클래스를 예측한다. 예측된 결과 벡터에서 큰 예측값(M)을 찾아낸 후 이를 사전에 정의된 임곗값(T)과 비교한다.

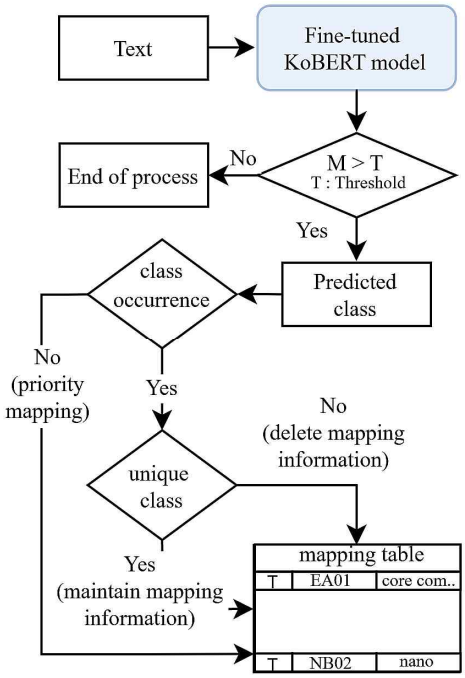

그림 2. 매핑 과정
Fig. 2. Mapping process

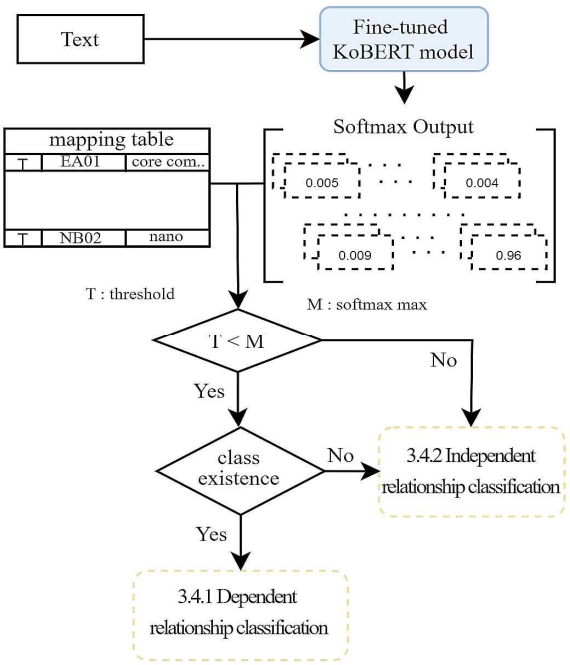

그림 3. 관계 분류 과정
Fig. 3. Process of categorizing relationships

예측값이 임곗값보다 큰 경우 해당 값이 이전에 나타나지 않았는지 검사하며, 나타나지 않으면 매핑 테이블에 우선 매핑한다. 만약 나타난 경우라면 매핑 테이블에 매핑이 고유한지 검사한다. 고유하지 않은 매핑이라면 매핑 테이블에서 해당 정보를 삭제하고, 아닌 경우 매핑 정보를 유지한다. 모든 입력값에 반복적으로 수행되면 최종적으로 두 분류 체계에 대한 매핑 정보를 담은 매핑 테이블이 구축된다.

3.4 관계에 따른 분류

앞서 수행된 과정에서는 서로 다른 분류 체계 간 유사한 클래스 간 매핑 정보를 담고 있는 매핑 테이블을 구축하였다. 이처럼 클래스 간 매핑이 가능한 경우도 있으나, 분류 체계별로 추구하는 목적이 다르므로 입력값이 하나의 분류 체계에만 속할 수 있다. 본 절에서는 매핑이 가능한, 즉 공통적인 범위를 지니는 관계를 종속적 관계로 정의하고, 매핑이 가능하지 않아 다른 범위를 지니는 관계를 독립적 관계로 정의하였다. 그림 3은 종속적 관계와 독립적 관계를 판단하는 순서도이다.

3.4.1 종속적 관계 분류

매핑 과정과 동일하게 입력값이 발생할 경우 미세 조정된 분류 모델을 활용하여 예측을 수행한다. 예를 들어, 모델은 예측 결괏값을 벡터 형태로 출력하게 되며, 이 벡터에서 최댓값(M)과 그에 해당하는 인덱스 값을 추출한다. 최댓값(M)은 해당 입력값이 특정 클래스에 속할 가능성을 나타내는 값으로 벡터 내에서 가장 높은 값을 가지는 요소이다.

매핑 테이블은 각각 임곗값(T)을 포함하고 있으며, 이 값은 모델의 출력값인 최댓값(M)과 비교된다. 만약 최댓값(M)이 임곗값(T) 이상이며, 해당 인덱스값이 매핑 테이블에 존재하는 경우, 입력값의 분류 결과는 매핑 테이블에 정의된 대로 바로 매핑된다. 두 개의 조건에 모두 해당하지 않을 경우 독립적 관계 분류(비포함) 관계로 분류된다.

3.4.2 독립적 관계 분류

독립적 관계에 해당하는 텍스트 분류는 OpenAI 사의 GPT-4o API[23]를 기반으로 수행되었으며, 이를 위한 프롬프트 설계 방식은 다음과 같이 구성하였다.

대형 언어 모델을 활용하여 제로샷 분류를 수행하기 위해서는 분류 클래스에 대한 최소한의 정보를 제공하는 것이 중요하다. 이를 위해, 분류 체계의 각 클래스에 대한 설명문을 입력값으로 제시하여 모델이 각 클래스의 의미를 이해하고 적절한 분류를 수행할 수 있도록 한다. 클래스에 대
한 설명문은 표 4와 같이 구성하여, 23개의 클래스에 대하여 모두 작성하였다. 또한 분류 체계에 포함되지 않는 기타 항목을 다루기 위해 해당 항목들에 대한 설명과 예시를 제공하여 모델이 이를 효과적으로 분류할 수 있도록 한다. 이 외에도 분류 성능을 향상시키기 위해 계층적인 분류 방향성 설정, 분류 결과의 타당성을 검증하기 위해 각 분류에 대한 이유의 설명을 함께 제시하는 방식을 선택하였다. 프롬프트 작성 시 입력 제공 정보는 표 5와 같다. 이러한 프롬프트 설계는 모델이 입력 텍스트를 단순히 분류하는 것에 그치지 않고, 근거를 함께 제시하도록 한다. 표 6과 같이 구조화된 출력을 생성하며, 이를 추후 검증하는 과정에 사용함으로써, 일관되고 신뢰성이 높은 결과를 얻을 수 있다. 이러한 과정을 통해 모델은 최소한의 정보로 분류 체계 내에서 적합한 분류를 수행할 뿐 아니라, 분류 체계 외의 항목에 대해서도 적절한 처리를 할 수 있는 능력을 가지게 된다.

표 4. 설명문 입력값 형식
Table 4. Input format of explanation

```
description_information = {
"IT" : {
      "core components" : description of core
                          components ,
            ...,
      "information processing systems and s/w" :
        description of information processing systems
        and s/w
      }
   ...,
"CT" : {
       "cultural contents" : description of cultural
                          contents,
       ...,
       "cultural heritage" : description of cultural
                          heritage
      }
 "not applicable" : example
}
```

표 5. 프롬프트 제공 정보
Table 5. Input information of prompt

```
Classify the user_input according to the
description_information provided. (The provided
classification information consists of the main
categories, subcategories within the main categories,
and descriptions and examples of those categories.

<<Classification Guidelines>>
...

description_information : { description_information }
user_input : { user_input }
```

표 6. 구조화된 출력 구조
Table 6. Structured output format

```
major category : [major category code]\n
medium category : [medium category code]\n
classification reason: [reason for classification]\n
```

3.5 자동검증

자동검증 과정은 3.4.2 독립적 관계 분류의 결괏값을 확인하고 평가하는 단계이다. 프롬프트에 의해 대형 언어 모델에서 생성된 구조화된 출력 결괏값을 다시 대형 언어 모델의 입력값으로 활용하여 수행된다. 입력값으로는 텍스트 원문, 1차 분류 결과, 분류 이유 설명으로 설정되며, 이를 바탕으로 해당 분류가 참인지 거짓인지 판단한다. 결과가 참인 경우, 분류 결과는 그대로 유지되며, 거짓인 경우 거짓으로 판정된 이유를 설명하도록 프롬프트를 구성한다. 이로써 검증 과정에서 거짓으로 판단된 집합은 3.4.2 독립적 관계 분류 프롬프트에서 검증 결괏값을 추가하여 재분류가 진행된다. 재분류 과정을 통해 최종적인 결괏값이 도출되며, 검증과 재분류를 통해 모델의 분류 성능은 향상된다.

IV. 실험 및 분석

4.1 데이터셋

미래유망신기술(6T) 분류를 위해, 정책보고서를 작성하는 외부 전문업체에서 수동으로 수집한 376건의 정책 문서를 사용하였다.

해당 문서는 전문가의 라벨링 작업을 거쳐 분류 정답 데이터셋으로 구축되었다.

4.2 평가 기준

실험에서는 데이터셋의 불균형 문제를 고려하여 성능을 보다 정밀하게 측정하기 위해 Precision@1, Recall@1, F1_score, 그리고 Accuracy 등의 평가 지표를 사용하여 모델의 성능을 분석하였다. Precision@1은 모델이 상위 한 개의 예측에 대해 얼마나 정확한지 평가하는 지표로, 예측 결과가 실제 정답과 일치하는지 측정한다. Recall@1은 상위 한 개의 예측이 실제 정답에 속하는지 평가한다. F1-score는 Precision과 Recall 간의 조화 평균을 의미하며, 이러한 지표들은 각 클래스에 대해 개별적으로 계산되어 불균형 데이터 평가에 적합하다. 반면, Accuracy는 전체 예측 중 정답 비율을 나타내는 지표이다. Accuracy는 불균형 데이터셋에서 다수 클래스에 편향될 수 있지만, 전반적인 모델 성능을 직관적으로 평가하는데 유용하다. 이러한 다양한 평가 지표를 통해 분류 성능을 종합적으로 평가하였다.

$$\mathrm{Pr}ecision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 - score = 2\frac{\mathrm{Pr}ecision * Recall}{\mathrm{Pr}ecision + Recall} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

4.3 실험 결과

학습 자원이 없는 분류 체계를 효과적으로 분류하기 위해 학습 자원이 충분한 분류 체계와의 매핑 전략과 대형 언어 모델을 활용한 설명문 기반 제로샷 분류 및 검증을 수행하였다.

4.3.1. 제안 기법의 성능 향상

표 7은 제안 기법을 적용한 분류 성능과 적용하지 않은 경우의 성능을 비교한 결과를 나타낸다. 제안 기법을 적용하지 않을 경우는 프롬프트 적용 전/후 와 검증 전/후로 나뉜다. w/o prompt는 분류 체계에서 클래스 정보만 제공하여 대형 언어 모델을 이용한 제로샷 분류를 수행한 경우로, 34%의 F1-score를 나타내어 효과적인 분류가 어려움을 확인하였다. 설명문을 기반으로 작성한 프롬프트를 추가한 prompt w/o verification의 경우, 별도의 검증 과정 없이 48%의 F1-score를 보였으며, 설명문을 활용한 프롬프트가 분류 성능 향상에 기여하는 것을 확인하였다. 또한 검증 과정을 추가한 prompt w/ verification은 69%의 F1-score를 보이며 검증 전 대비 21%의 성능 향상을 보였다. 제안 기법을 적용한 경우, prompt w/o verification에서 58%의 F1-score를 보였다. 이는 제안 기법을 적용하지 않은 경우보다 10%의 성능 향상을 보였으며, 검증 과정을 추가한 경우 70%의 F1-score를 보이며 이전 대비 3.3%의 성능 향상을 보였다. 이를 통해 제안 기법이 모든 경우에서 긍정적인 영향을 미치는 것을 확인하였다.결과적으로, 학습 자원이 충분한 분류 체계를 활용하여 학습 자원이 없는 분류 체계와 매핑하여 분류하는 전략이 유효하다는 사실을 입증하였다.

표 7. 성능 평가
Table 7. Experimental evaluation

6T classification		Precision	Recall	F1-score	Accuracy
GPT-4o (zero-shot)	w/o prompt	0.317	0.368	0.341	0.247
	prompt w/o verification	0.416	0.583	0.486	0.601
	prompt w/ verification	0.68	0.66	0.669	0.731
KoBERT pipeline GPT-4o	prompt w/o verification	0.533	0.651	0.586	0.656
	prompt w/ verification	0.714	0.696	0.705	0.76

또한, 분류 체계의 설명문과 검증 과정에서 프롬프트 엔지니어링을 통한 대형 언어 모델의 활용이 매우 효과적임을 확인하였다.

4.3.2 분류 체계간 매핑 분석

올바르게 이루어진 매핑, 그렇지 못한 매핑에 대한 분석과 매핑의 한계점을 상세히 다룬다. 표 8은 실험 과정에서 구축된 매핑 테이블의 일부를 나타내며, 여기서 T는 미세 조정된 KoBERT 모델의 Softmax 출력 임계값을 의미한다. 이 테이블은 과학기술표준분류 모델의 결과에 따라 미래유망신기술을 매핑한 결과로, 1대 다의 매핑 관계를 가진다. 예시는 대형 언어 모델의 잘못된 분류 결과를 올바르게 수정한 대표적인 매핑 사례들로 구성되어있다. 이를 통해 서로 다른 분류 체계 간의 일부 유사성을 확인할 수 있으며, 예를 들어 "해양환경"과 같이 동일한 클래스 이름을 가진 경우도 확인할 수 있었다.

표 8. 분류 체계 간 매핑 테이블 예시
Table 8. Example of a mapping table between classification systems

T	Science and technology standard classification	Future promising new technologies (6T)
0.95	wireless communication network	next-generation network base
	greenhouse gas treatment	environmental base
	agricultural environment ecology	environmental base
	applied mathematics	no applicable classification
	weapon sensors and control	no applicable classification
0.8	convergence bio	basic·fundamental technology
	waste management/resource circulation	environmental base
	nuclear fuel cycle/radioactive waste management technology	energy
	energy/environmental system	energy
0.7	marine environment	marine environment
	artificial satellite	st (other)

그러나 모든 매핑이 정확하게 이루어진 것은 아니다. 예를 들어 미래유망신기술의 정보처리 시스템 클래스는 과학기술표준분류의 정보 이론, 정보 보호 클래스에 대부분 매핑되었지만, 일부는 기타 정보 기술로 분류되는 등 클래스 간 경계가 모호한 경우 매핑 테이블을 구성하지 못하였다. 또한, 미래유망신기술의 CT(Culture Technology)에 해당하는 문화콘텐츠, 생활문화, 문화유산 등의 중분류 클래스는 과학기술과 연관이 없으므로 매핑이 불가능하였다. 이러한 결과를 통해 매핑 과정에서 발생하는 한계점을 확인할 수 있었다.

Ⅴ. 결론 및 향후 과제

본 연구를 통해 저자원 분류 체계에 대한 분류 가능성을 확인하였으며, 서로 다른 분류 체계의 매핑 전략과 설명문을 활용한 대형 언어 모델의 분류 및 검증 프롬프트를 통해 분류 성능을 유의미하게 향상시킬 수 있음을 입증하였다. 특히, 학습 데이터가 없는 분류 체계를 학습 자원이 충분한 다른 분류 체계와 효과적으로 통합함으로써 새로운 데이터 구축의 필요성을 줄였으며, 이는 학습데이터가 부족한 상황에서도 효과적으로 분류할 수 있음을 의미한다. 이러한 접근법은 다양한 분야에서 광범위하게 활용할 수 있을 것으로 기대된다. 그러나, 불균형한 클래스 분포와 제한된 데이터셋을 바탕으로 실험과 평가를 진행하였으므로, 추후 다양한 영역과 더 큰 규모의 데이터셋을 활용한 연구를 통해 제안 기법의 적용 가능성을 넓혀 나갈 필요가 있다. 또한 프롬프트 기반의 접근법이 프롬프트의 설정 방식에 따라 결괏값이 크게 달라질 수 있다는 점을 고려해야 한다. 따라서, 다양한 프롬프트 기법을 적용하여 분류 성능을 최적화하는 방법, 매핑 전략에 대한 알고리즘 개선, 두 개 이상의 분류 체계를 활용하는 등의 후속 연구가 필요하다.

References

[1] I. Kuzminov, D. Meissner, A. Lavrynenko, and E. Khabirova, "Technology Classification for the

Purposes of Futures Studies", Higher School of Economics Research Paper, Jan. 2018. http://dx.doi.org/10.2139/ssrn.3104451.

[2] S. Mohammadi and M. Chapon, "Investigating the Performance of Fine-Tuned Text Classification Models Based on Bert", 2020 IEEE 22nd International Conference on High Performance Computing and Communications, IEEE 18th International Conference on Smart City and IEEE 6th International Conference on Data Science and Systems (HPCC-SmartCity-DSS 2020), Yanuca Island, Cuvu, Fiji, pp. 1252-1257, Dec. 2020. https://doi.org/10.1109/HPCC-SmartCity-DSS50907.2020.00162.

[3] OpenAI, "ChatGPT", https://chat.openai.com/ [accessed: Aug. 08, 2024]

[4] N. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, and A. Mian, "A Comprehensive Overview of Large Language Models", arXiv preprint, arXiv:2307.06435, Jul. 2023. https://doi.org/10.48550/arXiv.2307.06435.

[5] B. Chen, Z. Zhang, N. Langrené, and S. Zhu, "Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review", arXiv preprint, arXiv:2402.17177, Oct. 2023. https://doi.org/10.48550/arXiv.2402.17177.

[6] https://www.kistep.re.kr/board.es?mid=a10305080000&bid=0002&act=view&list_no=43328&tag=&nPage=1 [accessed: Jul. 07, 2024]

[7] https://www.mss.go.kr/common/board/Download.do?bcIdx=44887&cbIdx=207&streFileNm=20140218105317516.hwp [accessed: Jul. 07, 2024]

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv preprint, arXiv:1810.04805, Oct. 2018. https://doi.org/10.48550/arXiv.1810.04805.

[9] SKT Brain, "KoBERT: Pretrained Language Models for Korean", GitHub repository, https://github.com/SKTBrain/KoBERT [accessed: Aug. 07, 2024]

[10] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification?", Chinese Computational Linguistics, Vol. 11856, pp. 194-206, Oct. 2019. https://doi.org/10.1007/978-3-030-32381-3_16.

[11] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. A. Almazroi, "A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification", Journal of Healthcare Engineering, Jan. 2022. https://doi.org/10.1155/2022/3498123.

[12] J. Hyeon, J. Lee, and H. Cho, "Sentiment Analysis of News on Corporation Using KoBERT", Korean Accounting Review, Vol. 47, No. 4, pp. 33-54, Aug. 2022. https://doi.org/10.24056/KAR.2022.08.002.

[13] J.-Y. Kim, D. Lee, and S. Cheon, "Literary Emotion Classification using KoBert", The Journal of Korean Studies, Vol. 87, No. 12, pp. 5-31, Dec. 2023. https://doi.org/10.17790/kors.2023.12.87.5.

[14] Y. Chae and T. Davidson, "Large Language Models for Text Classification: From Zero-Shot Learning to Instruction-Tuning" SocArXiv Papers, Jul. 2024. https://doi.org/10.31235/osf.io/sthwk.

[15] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang, "Text Classification via Large Language Models", Findings of the Association for Computational Linguistics, EMNLP 2023, Singapore, pp. 8990-9005, Dec. 2023. https://doi.org/10.18653/v1/2023.findings-emnlp.603.

[16] M. J. J. Bucher and M. Martini, "Fine-Tuned 'Small' LLMs (Still) Significantly Outperform Zero-Shot Generative AI Models in Text Classification", arXiv preprint, arXiv:2406.08660, Jun. 2023. https://doi.org/10.48550/arXiv.2406.08660.

[17] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing", ACM Computing Surveys, Vol. 55, No. 9, pp. 1-35, Jan. 2023. https://doi.org/10.1145/3560815.

[18] T. B. Brown, et al., "Language Models are Few-Shot Learners", Advances in Neural Information Processing Systems, NeurIPS 2020, Vol. 33, pp. 1877-1901, May 2020

[19] Y. Fei, M. Zhao, P. Nie, R. Wattenhofer, and M. Sachan, "Beyond prompting: Making Pre-trained Language Models Better Zero-shot Learners by Clustering Representations", Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, pp. 8560-8579, Dec. 2022. https://doi.org/10.18653/v1/2022.emnlp-main.587.

[20] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models", Advances in Neural Information Processing Systems, Vol. 35, pp. 24824-24837, Jan. 2022.

[21] Z. Wang, Y. Pang, and Y. Lin, "Large Language Models Are Zero-Shot Text Classifiers", arXiv preprint, arXiv:2312.01044, Dec. 2023. https://doi.org/10.48550/arXiv.2312.01044.

[22] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=71531 [accessed: Jul. 07, 2024]

[23] https://platform.openai.com/docs/models/gpt-4o [accessed: Aug. 08, 2024]

저자소개

곽 지 호 (Jiho Gwak)

2019년 3월 ~ 현재 : 국립금오공과대학교 컴퓨터공학과 학사과정
관심분야 : 딥러닝, 자연어처리, 텍스트 분류

정 유 철 (Yuchul Jung)

2023년 2월 ~ 2005년 2월 : 한국과학기술원(KAIST) 정보통신공학과(공학석사)
2005년 2월 ~ 2011년 2월 : 한국과학기술원(KAIST) 전산학과(공학박사)
2009년 1월 ~ 2013년 7월 : 한국전자통신연구원(ETRI) 연구원/선임연구원
2013년 8월 ~ 2017년 8월 : 한국과학기술정보연구원(KISTI) 선임연구원
2017년 8월 ~ 2022년 9월 : 국립금오공과대학교 컴퓨터공학과 조교수
2022년 10월 ~ 현재 : 국립금오공과대학교 인공지능공학과 부교수
관심분야 : 거대언어모델, 자연어처리, 지식그래프, 한국어 음성 인식/합성, AI응용