

공공분야 행정 전자문서 요약 생성을 위한 sLLM 모델에 관한 연구

황현준, 윤수연*
국민대학교, *국민대학교

acabar0105@kookmin.ac.kr, *1104py@kookmin.ac.kr

A Study on sLLM-Based Summarization of Administrative Electronic Documents in the Public Sector

Hwang Hyeon Jun, Soo-Yeon Yoon*
Kookmin Univ., *Kookmin Univ.

요 약

공공기관에서는 업무 중 많은 시간을 전자 결재 업무에 사용할 수 밖에 없으며, 이러한 전자 결재 문서는 방대한 내용 또는 다수의 문서로 구성될 수 있다. 따라서 전자 문서의 요약을 제공하는 것이 업무 시간 단축에 효과적일 것으로 기대되며 이를 위해 행정 문서 도메인 특화 요약 sLLM 을 제안한다. 행정 도메인 전용 데이터셋을 구축하고 LLama 3.1 8B 모델을 QLoRA 로 파인튜닝한 뒤 map-reduce 전략으로 2~3 줄 이내의 요약본을 생성하도록 하였다. 이러한 연구를 통해 문서 내의 텍스트 뿐만 아니라, 표, 서식 등의 내용을 효과적으로 요약할 수 있음을 확인하였다.

I. 서 론

공공기관에서는 하루 업무 중 많은 시간을 전자결재 업무에 사용할 수밖에 없다. 한국국토정보공사에서 2023 년 생산된 전자결재 문서는 1,653,610 건으로 사용자 4,600 명을 기준으로 단순 계산상 연간 1 인당 약 360 건의 문서가 생산되며 평균 결재선의 결재자가 3 명만 초과해도 연간 1 인당 약 1,000 번의 전자 결재가 진행된다고 볼 수 있다. [1]

공공기관의 주요 의사 결정 과정에서 활용되는 전자 결재 문서는 계획, 법규, 정부 정책, 규정 등의 핵심 정보만을 담고 있지만, 이와 함께 첨부되는 불임 문서의 경우 방대한 내용으로 이루어 졌거나 다수의 문서로 구성되어 있는 경우도 있다.

이러한 결재 문서와 불임 문서를 AI 기술을 활용하여 미리 요약하여 요약문을 제공해 준다면, 기안자와 결재자가 관련 내용을 파악하는 시간을 단축시키는 데 기여를 할 수 있을 것으로 기대된다.

II. 본론

1. 관련 연구

1.1 LLM 기반 전자 문서 요약

문서 요약은 원본 문서에서 핵심 내용만을 요약하여 문서를 구성하는 기법이다. 이러한 요약 방식은 추출 요약(Extractive Summarization)과 생성 요약(Abstractive Summarization)으로 구분할 수 있다. 추출 요약은 문서 내에 있는 문장들 가운데 중요한 문장을 선정하고 선정한 문장들을 추출 및 조합하는 방식으로 대표적인 모델로 TextRank 가 있다. 추상 요약은 모델을 통해 새로운 문장을 생성하여 문서의 요약문을 제시하는 기법으로 원문에 없는 새로운 문장을

생성함으로써 추출 요약의 한계를 극복할 가능성을 가지고 있다. 최근에는 대형 언어 모델(LLM)이 활용되고 있으며, 문맥을 깊이 이해하고, 의미를 파악하여 문장을 생성하는 능력을 토대로 문서 요약에서 뛰어난 성능을 보이고 있다.

2. 실험 설계

2.1 실험 환경 및 데이터셋 구성

본 실험은 RunPod 의 클라우드 서버를 대여하여 다음과 같이 구성하여 진행하였다.

[표 1] 실험 환경

구분	내용
GPU	A100 SXM
Python	3.11.10
Pytorch	2.4.1 + cu124
Transformers	4.51.3

본 연구에서는 ‘정보공개포털’에서 공개한 전자 문서 500 여건을 전처리 및 증강하여 학습에 활용하였다. 전자 문서는 도메인을 교육청으로 제한하고 첨부 문서가 존재 케이스만으로 한정하여 총 결재문서 520 건, 불임 문서 570 건 합쳐서 총 1090 건의 문서를 수집하였다.

문서 전처리 과정에서 표의 경우 Markdown 형식으로 변환하여 추출하고 의미 없는 텍스트를 제거하였다. 이후 결재 문서와 불임 문서를 일대일 매칭하여 570 쌍의 요약 문서를 구성하였다.

요약 문서 중 95%는 학습 데이터로 활용하기 위해 gpt-4o-mini 를 사용하여 요약 문서에 대응하는 요약문을 생성하여 약 3200 건, 270 만 개 토큰으로 이루어진 문서 요약 데이터셋으로 재구성하였다.

2.2 실험 방법 및 학습 모델 선정

공공 분야 전자 문서 요약 생성을 위해 Llama 를 활용하여 결재문서 520 건, 불입 문서 570 건 합쳐서 총 1090 건의 문서를 대상으로 도메인에 맞는 전처리를 수행한 후 사전학습을 실시하였고, 이를 기반으로 PEFT 파인튜닝을 QLoRA 기법으로 진행한 후 성능 평가를 진행하여 도메인에 맞는 문서 요약 및 정확도에 대한 평가를 수행하였다.

2.2.1 Llama

LLama 는 미국의 META 에서 출시한 대형 언어 모델로 비상업적 라이선스로 배포되어 누구나 연구, 개발, 상업적 목적으로 사용할 수 있다. 2024 년 7 월에 공개된 Llama 3.1 은 거대 모델 2 중(70B, 405B)과 경량 모델(8B)을 제공하고 있으며 본 논문에서는 경량 모델인 8B 모델을 선택하여 실험을 진행하였다.

2.2.2 QLoRA(Quantized Low-Rank Adaptation)

QLoRA 는 PEFT(Parameter-Efficient Fine-Tuning) 의 방법 중 하나로 LoRA 에 양자화를 추가하여 메모리의 효율성을 높인 방법이다.

2.2.3 Map Reduce

긴 문서 요약에서 생기는 LLM 의 높은 계산 비용과 Lost-in-the-middle 문제를 해결하기 위해 LangChain 라이브러리를 활용하여 prompt chaining 과 map-reduce 전략을 구현하였다.

문서를 1024 token length 로 chunking 하여 각 chunk 에 개별적으로 LLM 을 적용하여 요약하는 map 단계와, map 단계에서 생성된 1 차 요약문들의 token length 를 지정된 값 이하가 될 때까지 재귀적으로 압축하는 reduce 단계가 있다. 최종적으로 reduce 단계와 동일한 prompt 를 적용하여 단일의 요약문을 얻어내었다. [2] [3]

2.3 평가 방법

2.3.1 ROUGE

문서 요약 품질을 평가하는 지표 중 하나로 사람이 작성한 기준요약(이하 참조 요약)과 모델이 생성한 요약 간의 일치도를 측정하는 지표이다. N-gram 중복도로 문서 요약의 품질을 측정하고, ROUGE-L 은 가장 공통 부분문자열을 기반으로 평가하는 방식이다.

2.3.2 BERTScore

후보 문장과 참조 문장 간의 유사도를 평가하는 NLP 평가 지표이다. BERT 와 같은 언어 모델을 활용해 평가하고자 하는 각 문장의 문맥 임베딩을 생성한 뒤 코사인 유사도 측정을 통해 문장 간 유사성을 비교한다.

3. 실험 및 성능 평가 분석

3.1 실험 및 결과

[표 2] 문서 요약 성능 평가

Method	Fine-Tuning model (A)	LLama 3.1 8B Instruct 4bit (B)
ROUGE-1	0.463957	0.402273
ROUGE-2	0.250861	0.166044
ROUGE-L	0.313430	0.254812
BLEU	0.154359	0.093442
BERTScore	0.719139	0.706657

공공분야 행정 전자문서의 요약 성능을 평가하기 위하여 Fine-Tuning 을 한 모델과 대조군으로 LLama 3.1 8B Instruct 모델을 대상으로 20 여개의 검증용 전자문서를 Map Reduce 전략으로 요약하도록 하였다, 요약 중 생성된 1 차 요약문과 최종 요약문을 대상으로 요약 품질을 평가하였다.

3.2 성능 평가 분석

표 2 에서 보이는 바와 같이 모든 지표에서 Fine-Tuning 을 한 모델이 우수한 것을 확인할 수 있다. 특히 ROUGE-1,2 와 BLEU 에서 각각 6.1%p, 8.4%p, 6%p 높은 수치를 기록하였으며, 이는 문서 요약 과정에서 핵심 키워드와 구문의 일관성을 잘 유지하고 불필요한 단어를 덜 사용했다는 것을 의미한다.

특히 표 3 과 같이 불입 문서가 문서작성을 위한 서식일 경우 Instruct 모델의 경우 서식 작성을 위한 정보가 아닌 시각적 구조를 요약하려고 하는 것에 반해 Fine-Tuning 모델의 경우 양식에 어떤 항목이 있고, 각 항목을 작성할 때의 주의사항 등의 핵심적인 내용을 요약하는 것을 확인할 수 있었다.

<표 3> 불입 문서가 서식 유형일 때 요약결과

(A)	본 문서는 2025 년도 지원 대상을 위한 보조사업 우선순위와 관련된 서식을 포함하고 있습니다. 주요 내용으로는 사업명, 단체명, 예산액, 보조금 신청액 등이 있으며, 각 항목에 대한 세부 사항을 기록해야 합니다. 또한, 사업계획서는 부서별로 작성되어야 하며, 사업 목표, 내용, 비용 신청 내역, 효과 등의 정보를 포함해야 합니다. 기대효과도 명확히 기술 필요합니다.
(B)	* 기대효과 + 사업 목적 달성 시 나타날 실제 효과 기재 * 우선순위 목록 (사업부서 작성) + 2025 년 지원 대상을 위한 보조사업 우선순위 목록 + 순위 사업 이름 단체 명 2025 년 예산 액 보조 금 신청 액 사 전 심 의 조 정 액

III. 결론 및 향후 연구

본 연구에서 sLLM 을 활용하여 행정 전자문서의 텍스트 뿐만 아니라, 표, 문서 서식, 다수의 문서로 구성된 내용 등을 파악하기 위한 짧으면서 효과적인 요약문을 생성할 수 있는 것을 확인하였다.

다만 본 연구에서 BERTScore 가 71.9%로 참조요약과 의미적으로는 일치하며 문서의 핵심 정보를 담았을 수는 있지만 사람 기준으로 부자연스러운 표현과 문체를 사용하는 것으로 해석할 수도 있다.

향후 연구에서는 이러한 부분을 보완하기 위해 few-shot, RAG 와 같은 방법론을 적용하여 안정적인 모델을 설계할 필요가 있다.

참고 문헌

- [1] 규. 조, “LLM 기반 전자결재 문서 요약전,” 숭실대학교, 석사학위논문, 2024.
- [2] W. Justin, K. Perez, A. Ali-Gombe, “Enhancing privacy policy comprehension through privacy: A user-centric approach using advanced language models,” *Computers & Security, Vol 145*, pp: 103997, 8 October 2024.
- [3] Y. Wu, H. Iso, P. Pezeshkpour, “Less is More for Long Document Summary Evaluation by LLMs,” *arXiv preprint arXiv:2309.07382*, 2023.