

프롬프트 엔지니어링을 통한 RAG 기반 데이터셋 품질 개선과 파인 튜닝 적용 모델의 성능 비교 연구

김성언¹, 김산이², 박정현², 유승미², 이수빈³, 박정규⁴

¹ 인하대학교 인공지능공학과 학부생

² 충북대학교 정보통신공학부 학부생

³ 충남대학교 국어국문학과 학부생

⁴(주)엘지씨엔에스

ykiki5778@inha.edu, sanikani@chungbuk.ac.kr, anna010828@chungbuk.ac.kr,
ysm0909@chungbuk.ac.kr, aksgdmschdan@o.cnu.ac.kr, junggpark@lgcns.com

Fine-Tuned Models Enhanced by Dataset Quality Improvement through Prompt Engineering in RAG

Seong-Eon Kim¹, San-I Kim², Jung-Hyun Park², Seung-Mi Yu², Su-Bin Lee³, Jung-Gyu Park⁴

¹Dept. of Artificial Intelligence, In-Ha University

²Dept. of Information and Communication Engineering, Chungbuk-National University

³Dept. of Korean Language and Literature, Chungnam-National University

⁴LG CNS Co., Ltd.

요 약

본 연구는 Port MIS 보조 민원 상담 서비스로서 인공지능 모델을 제시한다. RAG 와 파인튜닝을 결합한 모델(RAG + Llama-3-PortMIS-Ko-8B)을 개발함으로써 기존의 FAQ 게시판으로 이루어졌던 민원 상담을 대체할 서비스를 소개한다. 또한, 정확도와 BLUE score 를 통해 Base Model 보다 향상된 답변을 확인할 수 있다.

서론

Port-MIS(항만운영정보시스템)은 전국 무역항의 항만 운영 업무 및 민원 업무를 처리하는 시스템이다. 현재 항만 물류 민원 상담 서비스는 Port-MIS 홈페이지의 질의응답 게시판을 통해 이루어지고 있으며, 사용자가 질문을 올리면 답변이 도착할 때까지 기다리는 방식으로 진행된다. 이에 본 연구는 RAG(Retrieval-Augmented Generation) 기술과 파인튜닝을 결합한 인공지능 모델을 구축하고, 항만 물류 도메인에 특화된 답변을 제공할 수 있는 인공지능 모델을 구현하여 상담원의 업무 부담을 줄이고, 24 시간 민원 해결의 가능성을 제시한다.

본론

2.1 데이터셋 생성

본 연구에서는 Port-MIS 데이터를 수집하고, PDF 파일, OCR(광학 문자 인식)을 이용한 이미지 데이터, CSV 파일 등 다양한 형식의 데이터를 전처리하여 LLM 파인튜닝에 활용할 데이터셋을 생성하였다. Zero-

shot 과 Few-shot 프롬프트를 활용해 고품질 Q&A 데이터를 생성하며 모델의 성능을 높이하고자 하였다. 이를 통해 다양한 시나리오에 적용 가능한 고차원적 정보를 제공하는 데이터셋을 구축하고 증강하였다.

2.2 데이터셋 품질 평가

생성된 데이터셋의 성능 평가는 RAGAS 라이브러리를 활용하여 진행하였다.

<표 1> RAGAS 평가 비교

	Zero-shot		Few-shot	
	GPT-4	GPT-4o	GPT-4	GPT-4o
Faithfulness	0.7165	0.6837	0.7357	0.7852
Answer Relevancy	0.8207	0.7940	0.8120	0.8078
Context Recall	0.9385	0.9231	0.9846	1.0000
Context Precision	0.9692	0.9231	0.9692	0.9846
Answer correctness	0.9950	0.9845	0.9971	0.9836

Few-shot 프롬프트 방식이 Zero-shot 보다 전반적으로 더 높은 성능을 나타냈으며, 특히 Faithfulness 와 지표에서 뚜렷한 성능 차이를 확인할 수 있었다. 이

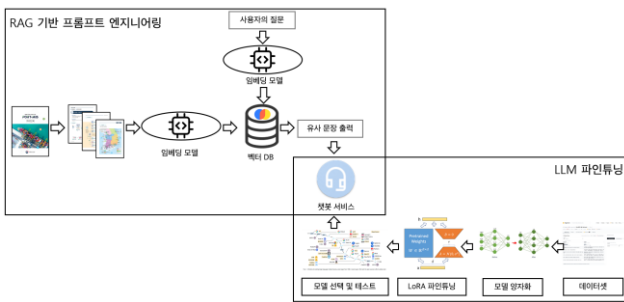
러한 결과를 바탕으로 최종 데이터 생성 방식으로 GPT-4o 와 Few-shot 프롬프트의 조합을 채택하였으며, 고품질 Port MIS 민원 데이터셋(PortMIS-QA-Dataset)을 3,000 개의 샘플로 구축하였다.

2.3 RAG 기법 활용 및 파인튜닝

본 연구는 Port MIS 민원 데이터셋(PortMIS-QA-Dataset)을 학습한 자체 LLM(Llama-3-PortMIS-Ko-8B)을 사용하며, RAG의 검색 기능을 통해 실시간으로 정보를 참조하여 응답 품질을 향상시켰다.

LoRA(Low-Rank Adaptation) 기법을 적용하고, 4 비트 양자화를 통해 모델의 학습 시간을 단축하고, 메모리 사용량을 최소화하였다. 또한, Gradient Checkpointing을 활용해 VRAM 사용을 최적화하였다. 결과적으로, 본 연구의 모델은 복잡한 도메인 특화 상담 챗봇으로 높은 정확도와 일관성을 제공할 수 있는 잠재력을 입증하며, RAG 기법을 활용한 모델의 성능 향상 효과를 확인하였다.

<그림 1> LLM 학습 파이프라인



2.4 모델별 테스트

RAG 와 파인튜닝의 성능을 확인하기 위해 Base Model 과 파인튜닝 모델, RAG + 파인튜닝 모델(RAG + Llama-3-PortMIS-Ko-8B)을 평가하였다. 성능은 Accuracy, BLEU Score, Response Time 으로 평가되었으며, 각 모델은 동일한 Port MIS 민원 데이터를 활용한 테스트 세트를 사용한 각 지표의 평균값을 기록하였다.

<표 2> 모델별 성능

Model	Accuracy	BLEU score	Response Time(s)
Base Model (Llama-3-Open-Ko-8B)	63.3	0.42	2.5
Llama-3-PortMIS-Ko-8B	75.4	0.83	2.8
RAG + Llama-3-PortMIS-Ko-8B	80.0	0.93	3.2

Base 모델에 Port MIS 데이터를 파인튜닝한 Llama-3-PortMIS-Ko-8B 모델은 Accuracy 와 BLEU Score 에서 크게 성능이 향상되었다. 추가적으로 RAG 기술을 활

용한 RAG + Llama-3-PortMIS-Ko-8B ¹모델은 두 지표에서 가장 높은 점수를 기록하였다. 다만, 응답 시간은 약간 증가하였으나 실무 적용에는 무리가 없는 수준으로 평가된다.

결론

3.1 성능 평가 결과 및 분석

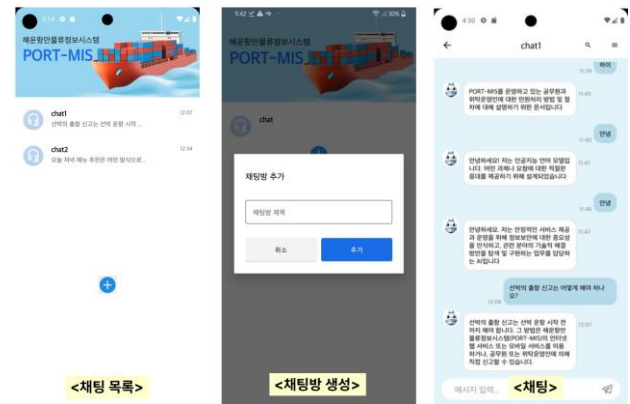
본 연구에서는 프롬프트 엔지니어링을 통해 RAG 기반 데이터셋의 품질을 개선하고, 이를 바탕으로 세 가지 모델(Base Model, 파인튜닝 모델, RAG + 파인튜닝 모델)의 성능을 비교하였다.

데이터셋 평가는 RAGAS 라이브러리를 통해 이루어졌으며, 그 결과 Few-shot 프롬프트 방식이 Zero-shot 보다 Faithfulness 에서 더 우수한 성능을 나타냈다. 이러한 결과를 바탕으로 GPT-4o 와 Few-shot 조합을 채택하여 고품질 데이터셋을 구축하였다.

모델 성능은 Accuracy, BLEU Score, Response Time 으로 평가되었으며, RAG + LLaMA-3-PortMIS-Ko-8B 모델이 복잡한 질문에 대해 정확한 답변을 생성하며 가장 우수한 결과를 보였다.

3.2 응용 가능성 및 기대 효과

<그림 2> 항만 물류 민원 챗봇 서비스 구현 화면



본 연구는 RAG + LLaMA-3-PortMIS-Ko-8B 모델이 챗봇 서비스에 적용 가능성을 제시하며, 특히 항만 물류 민원 서비스 분야에서 신속하고 정확한 답변을 통해 상담 업무 효율화와 사용자 만족도 향상에 기여할 잠재력이 기대된다.

참고문헌

[1] Roger S. Pressman "Software Engineering A Practliners' Approach" 3rd Ed. McGraw Hill

※ 본 논문은 해양수산부 실무형 해상물류 일자리 지원사업(스마트해상물류 x ICT 멘토링)을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

¹ <https://huggingface.co/Port-Pilot/Llama-3-PortMIS-Ko-8B>