

학부생 인턴

결과보고서

자율운항 선박을 위한 LLM 기반 자동 대화형
통신 시스템 개발

2025 . 08. 25.

소속	딥러닝 및 과학계산 연구실
지도교수	김 경 섭
작성자	김 연 범

목 차

I. 서론	3
1. 연구목적 및 필요성	3
II. 관련연구	6
1. 관련 연구 조사 및 기술 분석	6
2. 해상 특화 데이터셋 구성	9
3. 모델 구조 및 학습 방법	10
4. 실험 및 평가	11
III. 결론	14
1. 연구 요약 및 시사점	14

【참고문헌】

【부록】

I. 서론

1. 연구목적 및 필요성

1) 연구목적

본 연구의 궁극적인 목적은 해상 교통 환경에서 실시간 의사소통 및 의사결정 지원이 가능한 대규모 언어모델(LLM)을 구축하고, 나아가 자율운항의 초석을 마련하는 것이다. 이를 위해 다음과 같은 세부 목표를 설정하였다:

- 1) 도메인 특화 언어모델 개발: 선박 운항, 충돌 회피, 해양 통신 등 해상 도메인에 특화된 데이터를 기반으로, 실제 해양 교신 시나리오를 반영한 고성능 언어모델을 구축한다.
- 2) Fourier 기반 파인튜닝(FourierFT) 기법 적용: 기존의 Parameter-Efficient Fine-Tuning(PEFT) 방식인 LoRA 등과 차별화된, Fourier Transform을 활용한 파인튜닝 기법을 재현하고 이를 LLaMA3-8B 모델에 적용하여 메모리 효율성과 성능을 동시에 확보한다.
- 3) 실제 해상 통신 시나리오 반영 데이터셋 구축 및 활용: 선박명, 콜사인, MMSI 등 다양한 해상 정보에 기반한 프롬프트-응답 형식의 시뮬레이션 데이터를 제작하고, 이를 통해 모델의 도메인 적응력을 강화한다.
- 4) 최신 LLM 아키텍처 분석 및 구현: LLaMA-3 아키텍처의 구조적 특성과 학습 방식에 대한 분석을 바탕으로, 적절한 사전 처리 및 미세조정 전략을 설계하고 직접 구현함으로써 모델의 내부 작동 원리에 대한 심도 있는 이해를 확보한다.

본 연구는 해상 운송 및 항만 관리에서 발생하는 다양한 자연어 기반 의사소통 문제를 해결하고, 선박 간 자동화된 정보 교환 및 판단 보조를 가능하게 하는 인공지능 기반 시스템의 초석을 마련하고자 한다.

2) 필요성

1) 해상 통신의 특수성과 자동화의 한계

선박 간의 통신은 주로 VHF 무선 음성 통신이나 제한된 메시지 포맷에 의존하고 있으며, 그 내용은 구조화되어 있지 않아 자동화 및 해석이 어렵다[1]. 이러한 상황은 위급 상황에서의 판단 지연과 오해로 이어질 수 있다. 이를 해결하기 위해서는 해상 도메인에 특화된 언어 이해 및 생성 능력을 갖춘 모델이 필수적이다.

2) 기존 LLM의 일반성 한계

GPT, LLaMA 등 범용 LLM들은 해상 운항, AIS 정보, 콜사인, MMSI 등과 같은 전문 용어와 통신 양식을 충분히 이해하지 못한다[2]. 따라서 이들을 그대로 활용할 경우 도메인 적합성 부족, 정보 왜곡, 오답 생성 등의 문제가 발생한다. 이에 따라 해상이라는 특수한 문맥을 이해하고 적절하게 반응할 수 있는 LLM의 필요성이 대두되고 있다.

3) 지속적으로 증가하는 해상 교통량과 위험성

글로벌 해운 운송량은 지속적으로 증가하고 있으며, 이에 따라 해상에서의 충돌 위험, 통신 오류, 대웅 지연 등 다양한 문제가 발생하고 있다. 이러한 문제는 인공지능 기반의 실시간 통신 해석 및 대응 시스템을 통해 보완할 수 있으며, 이는 LLM 기반 기술을 통해 구현 가능하다.

4) FourierFT 기반 효율적 미세조정 기법의 적용 필요성

도메인 특화 LLM을 구축하기 위해서는 대규모 언어모델에 특화 데이터를 반영한 파인튜닝이 필요하다. 그러나 모델의 파라미터 수가 매우 많기에 기존 방식은 메모리 비용과 계산량 부담이 크다. 이에 대한 대안으로 제안된 Fourier 기반 파인튜닝(FourierFT)은 일부 선택된 파라미터에 대해 스펙트럼 공간에서 희소하게 학습을 수행[3]함으로써, 메모리 효율성과 성능 향상을 동시에 기대할 수 있다.

5) 국내외 해상 AI 기술의 선도 필요성

현재 해양 AI 분야는 항로 예측, 충돌 회피 등 개별 기능 중심으로 발전하고 있으며, 언어모델 기반의 해상 통신 이해 및 응답 시스템은 아직 초기 단계이다. 본 연구는 이러한 공백을 메우고, 향후 AI 기반 자율 해상 운항 시스템의 핵심 모듈로 기능할 수 있다는 점에서 선도적 가치가 있다.

3) 구성

본 연구는 해상 도메인 특화 대규모 언어모델(LLM)을 구축하기 위해, 다음과 같은 단계로 구성된다:

1장. 서론

해상 교통의 특수성과 통신의 비정형성에 따른 문제를 설명하고, 범용 LLM의 한계 및 도메인 특화 LLM의 필요성을 서술한다. 또한 FourierFT와 같은 최신 파인튜닝 기법이 본 연구에 도입된 배경을 밝힌다.

2장. 관련 연구 조사 및 기술 분석

- 1) 범용 LLM(LLaMA, GPT 등)의 구조와 한계를 분석
- 2) 해상 도메인에서 시도된 기존 연구(Llamarine, VTS-LLM 등)를 정리
- 3) Parameter-Efficient Fine-Tuning(PEFT) 기법의 동향 검토

특히, Fourier Transform을 기반으로 한 미세조정 방식(FourierFT)에 대해 수학적 정의, 기존 LoRA, Adapter 방법과의 차이점, 학습 효율성 측면에서의 장점을 심층 분석한다.

3장. 해상 특화 데이터셋 구성

- 1) AIS 기반 선박 정보(선박명, MMSI, SOG, COG 등)와 VHF 음성 통신 예시를 정형 텍스트로 변환하여 프롬프트-응답 데이터셋을 생성한다.
- 2) 커스텀 instruction 데이터 구축을 포함한다.
- 3) 다양한 시나리오(입항 보고, 선박 간 요청 등)을 반영하여 프롬프트 다양성을 확보한다.

4장. 모델 아키텍처 및 파인튜닝 기법 설계

- 1) 기반 모델: Meta LLaMA-3 8B를 선택하고, 모델의 구조를 분석한다.
- 2) 파인튜닝 방식: FourierFT 논문[3]에 따라, 선택된 층에 대해 sparse한 frequency domain coefficient를 학습하고, 이를 iDFT를 통해 원래 가중치에 더하는 방식으로 구현한다.
- 3) 이를 위해 다음과 같은 과정을 직접 구현함:
 - (1) 기존 weight에서 일부 위치를 선택하고 ΔW 를 frequency 도메인에서 학습
 - (2) Gaussian 초기화된 스펙트럼 계수를 iDFT 변환하여 dense한 ΔW 생성
 - (3) 기존 weight에 합산하여 forward pass 수행
 - (4) loss를 통해 sparse coefficient만 업데이트

5장. 모델 학습 및 실험 설정

- 1) 학습은 PyTorch 기반으로 구현되며, 4bit 양자화 모델에서 pre-trained 파라미터를 freeze 하여 Fourier 계수만 학습 가능하도록 설정
- 2) 학습률, max length, 배치 사이즈 등은 실험적으로 조정
- 3) label은 output 부분에만 적용하도록 -100 마스킹 기법을 사용하여 loss 계산 효율 증대

6장. 성능 평가 및 분석

- 1) 평가 지표로 BLEU, ROUGE, domain accuracy 등을 활용.
- 2) baseline(범용 LLM)과 fourierFT를 적용한 LLM의 평가 지표 비교 분석
- 3) 예시 문장에 대해 실제 응답 품질, 문맥 이해도, 도메인 일치 여부 등을 정성적 평가도 병행함.

7장. 연구 결론 및 시사점

본 연구의 모델 구조 및 정량/정성 평가 결과를 요약하고, 종합적인 결론 및 향후 연구 방향에 대해 시사한다.

II. 관련 연구

1. 관련 연구 조사 및 기술 분석

1) 범용 LLM의 구조와 해상 도메인 한계

대규모 언어모델(Large Language Model, LLM)은 사전학습(pretraining)을 통해 방대한 양의 텍스트를 기반으로 자연어 이해와 생성 능력을 갖춘 모델이다. 대표적으로 OpenAI의 GPT 계열, Meta의 LLaMA 시리즈, Mistral, Claude 등이 있으며, 이들은 대부분 Transformer 아키텍처를 기반으로 한다 [4].

Transformer 기반 LLM은 일반적으로 다음과 같은 구조를 갖는다:

- 1) 입력 토큰을 임베딩한 후, 포지셔널 인코딩을 더하여 문맥 정보를 부여
- 2) 여러 층의 self-attention 및 feed-forward 네트워크를 통과
- 3) 마지막 토큰에 대한 확률 분포를 계산하여 텍스트를 생성

그러나 이러한 범용 LLM은 일반 뉴스, 웹, 위키백과 등 범용적인 코퍼스를 기반으로 학습되었기 때문에, 해상 운항, AIS 시스템 정보, 콜사인, MMSI 번호, 선박 간 통신 문장 등과 같은 전문적인 해상 도메인 용어 및 대화 패턴에 적절하게 반응하지 못한다[2].

예를 들어, GPT-4 또는 LLaMa-3 모델은 다음과 같은 문장에 대해 실제 해상 문맥과 맞지 않는 응답을 생성할 수 있다:

귀선 SOG 보고 바랍니다

"SOG 보고 바랍니다"라는 표현에서 "SOG"는 상황에 따라 다른 의미를 가질 수 있습니다. 귀선(귀환하는 선박이나 함정)을 대상으로 한 보고를 요청하신 것으로 보이며, 일반적으로 군함이나 해양 관련 부대에서 사용하는 보고 양식일 수 있습니다.

그림 1: GPT-4 범용 모델의 응답 예시

이러한 응답은 상황 인식 부족, 도메인 언어 해석 실패, 운항 관련 결정 지원 미흡 등을 보여준다. 따라서, 범용 모델의 한계를 극복하고, 도메인 적응(domain adaptation)을 통해 전문적인 응답 생성을 가능하게 하는 방향이 필요하다[2].

2) 도메인 특화 LLM 연구: 해상 도메인 사례

범용 LLM이 해상 도메인의 언어를 효과적으로 이해하지 못한다는 한계를 극복하기 위해 최근에는 특정 산업 또는 환경에 맞춘 도메인 특화 LLM에 대한 연구가 활발히 진행되고 있다. 특히 해상 운항과 관련된 전문 문맥과 용어를 학습한 언어모델은, 단순 텍스트 생성뿐만 아니라 위급 상황 대응, 입출항 보고, 항로 조정 요청, 선박 식별 응답 등에서 실질적인 의사소통 지원 역할을 수행할 수 있

다.

2-1. Llamarine: 해상 특화 오픈소스 LLM

Llamarine[2]은 Meta의 LLaMA-2 모델을 기반으로 해상 산업에 특화된 데이터셋을 이용해 파인튜닝 된 언어모델로, 콜사인 질의, ETA 보고, 항구명 응답 등 다양한 업무 시나리오를 커버한다. 해당 연구에서는 다음과 같은 특이점을 제시했다:

- 1) 범용 LLM(GPT-3.5, Claude) 대비, 정확한 해상 용어 응답률 및 일관성 측면에서 우수한 성능을 보임.
- 2) “선박명-응답”, “위치-시간 질의”, “입항 여부 판단” 등 해상 특화 태스크에서 도메인 응답 일치율이 91.3%로 나타났다.

이는 범용 모델이 ETA, 부산항, Z타임(UTC) 등 운항에 필수적인 용어들을 문맥상 잘못 해석하거나 일반 지식으로 대체하는 반면, Llamarine은 해당 용어들을 선박 간 통신 패턴에 맞게 적절히 생성할 수 있음을 보여준다.

2-2. VTS-LLM: 선박 교통관제 시스템과 LLM의 결합

또 다른 시도로는 VTS(Vessel Traffic Service) 업무에 특화된 VTS-LLM이 있다. 해당 모델은 선박 간 보고 의무와 교통 흐름 제어를 자연어 수준에서 지원할 수 있도록 설계되었으며, 단순한 응답 생성이 아니라 상황 인식 기반의 응답 판단, 다중 선박 상태 요약, 위협 경고 생성까지 수행하도록 고안되었다[1].

VTS-LLM의 실험 결과에 따르면, 다중 선박 보고를 종합하여 적절한 대응 메시지를 생성하는 LLM 에이전트 방식이 기존 를 기반보다 빠르고 정밀했다. 특히 혼잡한 입항 상황에서, LLM은 선박별 위치, 속도, 목적지를 고려한 위험 우선순위를 스스로 정리해 컨트롤러에게 전달할 수 있었다.

2-3. LLM 기반 도메인 보강 방식의 중요성

이러한 연구들은 다음과 같은 핵심 시사점을 제공한다:

- 1) 단순히 LLM의 크기나 사전학습 범위를 확장하는 것만으로는 도메인 적합성(domain alignment)을 확보할 수 없다.
- 2) 해상 운항 상황에 특화된 instruction 데이터셋 구축, 도메인별 프롬프트 스타일 설계, 도메인 지식 반영을 위한 파인튜닝 전략이 핵심이다.
- 3) 따라서 해상 LLM 개발에는, 도메인 지식을 언어모델이 내재화할 수 있도록 하는 데이터 중심 접근법과 구조적 fine-tuning 기법이 반드시 요구된다.

3) Parameter-Efficient Fine-Tuning 기법의 발전과 FourierFT

대규모 언어모델을 특정 도메인에 적합하게 조정하려면, 대량의 파라미터를 업데이트하는 전체 파인튜닝(full fine-tuning) 방식이 전통적으로 사용되어왔다. 그러나 LLaMA-3, GPT-4, Mistral-7B 등 최신 LLM은 수십억 개 이상의 파라미터를 포함하고 있어, 전체 모델을 미세조정하는 방식은 막대한 메모리 비용, 연산량 증가, 오버피팅(overfitting) 위험 등을 수반한다. 이에 따라, 최근에는 보다 효율적인 파인튜닝 방식들이 제안되고 있으며, 이를 통칭하여 Parameter-Efficient Fine-Tuning (PEFT)이라 한

다.

3-1. 기존 PEFT 기법: Adapter, LoRA

가장 널리 사용되는 PEFT 기법 중 하나는 Adapter 방식으로, 기존 파라미터를 고정한 채 중간 계층에 소형의 추가 네트워크를 삽입해 학습한다. 또 다른 방법인 Low-Rank Adaptation(LoRA)[5]는 weight matrix를 두 개의 저랭크 행렬로 분해하고, 학습 가능한 부분만 업데이트함으로써 성능을 유지하면서 학습 파라미터 수를 획기적으로 줄인다.

LoRA는 다음과 같은 수식을 기반으로 한다:

$$W' = W + \Delta W = W + AB$$

그림 2: LoRA 수식

여기서 $A \in \mathbb{R}^{d \times r}$, $B \in \mathbb{R}^{r \times d}$ 는 학습 가능한 저랭크 행렬이며, 기존 weight W 는 고정된다. 하지만 LoRA는 특정 계층에 제한적으로 삽입되어야 하며, 파라미터 수가 많게는 수천만 개에 달할 수 있다.

3-2. FourierFT: 스펙트럼 기반 희소 미세조정

최근 Gao et al.[3]은 이러한 한계를 극복하기 위해 새로운 방식인 FourierFT(Fourier-based Fine-tuning)을 제안하였다. 이 방식은 기존의 가중치 행렬을 주파수 도메인(Fourier Domain)으로 변환하여, 그중 일부 주파수 계수만 선택적으로 학습하는 방식이다.

핵심 개념은 다음과 같다:

- 1) 각 레이어에서 n 개의 주파수 위치를 랜덤하게 선택하고, 이를 기반으로 학습 가능한 spectral 계수 $c \in \mathbb{R}^n$ 을 Gaussian 분포로 초기화한다.
- 2) 선택된 n 개의 주파수 계수만을 학습하여 sparse training을 진행한다.
- 3) 학습된 계수는 dense spectral 행렬 F 를 구성하는 데 사용되며, 이를 역변환 (iDFT)을 통해 원래 공간으로 복원한다.
- 4) 복원된 ΔW 는 실수부에 스케일링 계수 α 를 곱해 최종적인 파인튜닝 업데이트로 사용된다.
- 5) 최종적으로, 원래의 가중치 W_0 에 ΔW 를 더하여 파인튜닝된 가중치 W 를 생성한다.

이 방법은 다음과 같은 장점을 가진다:

- 1) 메모리 효율성: 학습 파라미터 수를 수십만 단위 이하로 제한 가능
- 2) 계산 효율성: 대부분의 연산이 FFT 기반으로 빠르게 처리됨
- 3) 표현력 유지: 일부 스펙트럼 성분만으로도 충분한 파라미터 보정을 수행할 수 있음

실험 결과에 따르면, FourierFT는 LoRA 대비 약 500배 적은 학습 파라미터 수(0.064M vs 33.5M)로도 동일하거나 더 우수한 성능을 달성하였다[3].

3-3. 해상 특화 LLM에서의 적용 가능성

해상 도메인 특화 LLM은 학습 데이터의 크기가 제한적이고, 모델의 응답 정밀도는 높아야 하며, 경

량화된 배포가 중요하다는 특성을 가진다. 이러한 조건은 FourierFT 방식의 장점과 정확히 부합한다:

- 1) 소량의 커스텀 instruction 데이터로도 효과적인 튜닝 가능
- 2) 기존 모델의 사전학습 정보를 최대한 보존한 채, 도메인 특성만 반영
- 3) 4bit 양자화 모델과의 결합을 통해 임베디드 시스템이나 해상 통신 장비 내 탑재 가능성 확대

따라서 본 연구는 FourierFT를 기반으로 한 파인튜닝 구조를 선택하여, 도메인 특화와 효율성을 동시에 확보하고자 한다.

2. 해상 특화 데이터셋 구성

대규모 언어모델을 도메인에 특화시키기 위해서는, 해당 도메인의 언어 사용 방식, 업무 용어, 상황별 표현 패턴 등을 반영한 Instruction 기반 데이터셋이 필수적이다. 특히 해상 도메인의 경우, 선박식별, 통신 지령, 위치 보고, 도착 예고 등 정형화된 보고 문장과 관계 응답이 주요 언어 자산이므로, 실제 통신 관행에 부합하는 자연어 데이터 구성이 필요하다.

본 연구에서는 다음 세 가지 절차를 바탕으로 해상 특화 프롬프트-응답(prompt-Completion) 쌍을 구축하였다:

1) 데이터 구성 목표 및 지향 방향

1-1. 정확한 도메인 언어 표현 반영

해상 보고 및 통신에서 사용되는 전문 용어와 형식을 충실히 반영하도록 설계하였다. 예를 들어 “ETA”, “CALL SIGN”, “접안 지시” 등은 그대로 사용하고, 일반 언어로 대체하지 않는다.

1-2. 상황 기반 응답 능력 강화

단순 질의응답을 넘어서, 상황 판단이 필요한 지시형 응답, 복합 정보 요약, 다중 객체 인식 등을 포함해, 실제 해상 통신에서의 활용 가능성을 고려하였다.

1-3. LLM 학습 친화적 형식 제공

Hugging face 기반 instruction-tuning 구조에 적합하도록 JSONL 형식의 프롬프트-응답 쌍으로 구성하였고, output token에만 loss를 계산할 수 있도록 구조적으로 구분하였다. (label=-100 처리).

2) 데이터 전처리 및 학습 포맷

Instruction 기반 파인튜닝을 위해 FLAN-style 데이터 구성 방식을 따른다[6]:

2-1. JSONL 포맷:

```
{  
    "instruction" : "선박의 MMSI 번호를 확인하여 알려주세요." ,  
    "input" : "шуфтото MMSI 확인 부탁합니다." ,  
    "output" : "шуфтото MMSI 440123456입니다."  
}
```

2-2. Prompt 포맷 구성:

System Prompt + FLAN-style(Instruction + Input + Output) 꼴로 구성

최종 프롬프트의 앞부분에는 모델의 역할, 행동 양식, 응답 방식 등을 명시하는 시스템 프롬프트(System Prompt)를 삽입하였다. 이는 대규모 언어 모델(LLM)이 고정된 역할(Role)과 문맥(Context)을 명확히 인식하도록 하여, 응답의 일관성과 안정성을 제고하고 전반적인 생성 성능을 향상시키기 위한 조치이다[7].

3) 데이터 사양 및 전처리

- 형식: JSON lines(.jsonl)
- 총 쌍 수량: 69쌍
- 특이사항: 각 응답 내 특정 해상 용어에 대해 Position Tagging 및 가변 치환이 가능하도록 설계

또한, 모든 응답은 학습 시 tokenizer는 prompt를 다음과 같은 방식으로 전처리한다:

- 전체 prompt + output을 하나의 입력 시퀀스로 연결
- prompt 부분의 label은 -100으로 마스킹하여 loss 계산에서 제외
- output 토큰만 loss 대상

이러한 전처리 방식은 zero-shot 또는 few-shot 학습에서도 강한 일반화 성능을 유도하는 것으로 알려져 있다[6].

3. 모델 구조 및 학습 방법

본 연구는 해상 도메인 특화 대규모 언어모델(Maritime LLM)을 효율적으로 학습시키는 방안으로 Meta LLaMA-3-8B 모델과 FourierFT(Parameter-Efficient Fine-Tuning with Discrete Fourier Transform) 기법을 결합한 파인튜닝 전략을 설계하였다. Instruction 기반 학습을 위한 FLAN-style 데이터 전처리와 함께, Fourier 스펙트럼 도메인에서의 weight 조정 방식을 통해 학습 효율성과 도메인 표현력을 동시에 확보하는 것을 목표로 한다.

1) 모델 구조 개요

사용된 사전학습 모델은 Meta LLaMA-3 8B이며, 다음과 같은 구조적 특징을 가진다:

- 모델 크기: 약 80억 개 파라미터(8B)
- 구조: Transformer 기반 decoder-only 구조
- 학습 범위: 2023년 전반기까지의 고품질 데이터 기반 사전학습
- 모델 환경: Huggingface Transformers 기반, 4bit 양자화(QLoRA compatible), A100 80GB 환경
- 특징: GPT-3.5 수준의 자연어 처리 능력 보유, 개방형 라이선스(Open-Source)

본 연구에서는 LLaMA-3의 기존 weight를 고정하고, 일부 weight tensor에 대해 FourierFT 방식으로 ΔW 를 추가 학습하는 방식으로 도메인 적응(adaption)을 수행하였다. 모델 전체를 end-to-end로 재학

습하지 않음으로써, 메모리 효율성과 빠른 수렴을 동시에 확보하였다.

2) FourierFT 파인튜닝 구조

기존의 PEFT 방식인 LoRA[5]는 weight 행렬 W 를 저랭크 행렬 A , B 의 곱으로 표현하는 방식이지만, 본 연구에서는 이를 대체하여 Discrete Fourier Transform 기반 학습 방식(FourierFT)을 채택하였다.

FourierFT는 다음과 같은 절차로 구성된다:

- 1) 모든 layer에 공유되는, n 개의 주파수 위치가 설정된 위치 인덱스 E 정의
- 2) 선택된 위치 인덱스 E 에 따라 coefficient C 를 삽입:
 - E 는 모든 layer에 공유되는 spectral entry matrix
 - C 는 Gaussian 정규화 분포로 초기화
- 3) Dense Spectral Matrix F 생성
- 4) 역변환하여 ΔW 산출
- 5) Weight 보정 및 역전파 수행:

$$W' = W + \Delta W$$

그림 3: FourierFT 수식

FourierFT의 총 학습 파라미터 수는 다음과 같이 계산된다:

$$\#params = n \times (2 + L)$$

그림 4: FourierFT 파라미터 수

여기서 n 은 학습하는 coefficient 개수, L 은 Fourier를 적용하는 layer 수, 2는 shared entry matrix 용 고정 파라미터이다. 관련 실험에서는 전체 weight 대비 0.05% 미만의 파라미터만 학습 대상으로 삼아도, LoRA 수준 이상의 결과를 보였다[3]. 본 연구에서는 전체 weight 대비 약 0.0002%(16,000개) 정도의 파라미터만 학습 대상으로 삼았다.

3) 손실 함수 및 마스킹 전략

Instruction 기반 파인튜닝에서는 output 부분만 loss 계산에 반영되도록 다음과 같이 label을 설정하였다:

- Prompt token: label=-100 설정으로 loss 미계산
- Completion token: label=실제 token id 설정으로 loss 계산

이 방식은 Instruction-tuning 시 prompt 복사를 방지하고, 모델이 응답만 학습하게끔 유도하는 효과가 있다. 모든 데이터는 전처리 시 prompt-completion pair 형태로 정규화된 JSONL 구조를 사용하였으며, 길이가 4096 tokens를 초과하는 경우 자동으로 클리핑 처리하였다.

4. 실험 및 평가

본 장에서는 해상 도메인 특화 Instruction 데이터셋을 기반으로, FourierFT 방식으로 파인튜닝된 LLaMA-3-8B 모델의 성능을 정량적·정성적 관점에서 종합적으로 평가한다. 정량적 평가는 BLEU, ROUGE, Domain Accuracy 등 대표적인 자연어 생성 평가 지표를 활용하였으며, Baseline 모델인 범용 LLaMA-3-8B (사전학습 상태)와의 비교를 통해 FourierFT 기법의 성능 개선 효과를 분석하였다. 또한, 실제 예시 문장에 대해 모델이 생성한 응답의 문맥 이해도, 도메인 용어 일치성, 표현 자연성 등을 중심으로 정성적 평가를 병행하여, 해상 도메인에서의 실용성과 응답 품질을 입체적으로 검토하였다.

1) 정량 평가(Quantitative Evaluation)

정량 평가는 모델이 생성한 응답의 정확성, 표현력, 도메인 적합성을 객관적으로 측정하기 위해 다음 세 가지 대표적인 자연어 생성 평가 지표를 활용하였다:

- BLEU(Bilingual Evaluation Understudy Score): n-gram 수준에서 생성 문장과 정답 문장의 유사도를 측정하는 지표로, 정확한 문장 재현 능력을 평가한다.
- ROUGE-L(Recall-Oriented Understudy for Gisting Evaluation): 정답 문장과 생성 문장 간의 Longest Common Subsequence(LCS) 기반 유사도를 측정하여 핵심 문장 구조의 보존 정도를 평가한다.
- Domain Accuracy: 해상 도메인 특화 정보(예: 선박명, 콜사인, ETA 등)가 생성 응답 내에서 정확하게 포함되었는지 여부를 이진 분류 형태로 평가한 지표다.

지표	Baseline(LLaMA-3-8B)	FourierFT
BLEU	0.1171	0.2224
ROUGE-L	0.3456	0.4031
Domain Accuracy	0.8261	0.9296

표 1: BLEU, ROUGE-L, Domain Accuracy에 대한 각 모델 평가 결과 요약

[분석 결과]

Fourier 모델은 BLEU 점수에서 기존 Baseline 대비 약 10% 향상된 결과를 기록하였다. 이는 생성된 응답의 문장 단위 정밀도(precision)가 크게 개선되었음을 의미하며, FourierFT 기법이 instruction-context 간 연관성을 더 정밀하게 반영한 결과로 해석할 수 있다. 특히 동일 파라미터 구조를 유지하면서도 향상된 표현 능력을 보였다는 점에서 파라미터 효율성이 입증되었다.

또한, ROUGE-L 측면에서도 약 17% 향상된 결과를 기록하였는데, 이는 FourierFT 기법이 해상 도메인처럼 절차적 설명과 맥락 유지가 중요한 응답에서 경쟁력 있는 결과를 나타낸다는 것을 시사한다.

마지막으로, Domain Accuracy 측면에서 약 12% 향상된 결과를 기록하였는데, 이는 단순 문장 생성 능력을 넘어서 도메인 지식 반영 능력(domain alignment)에서 FourierFT의 강점을 입증하는 결과라 할 수 있다.

2) 정성 평가(Qualitative Evaluation)

정성 평가는 모델이 생성한 응답이 실제 해상 통신 상황에서 얼마나 적절하고 신뢰성 있게 작동할 수 있는지를 평가하기 위해 수행되었다. 평가 기준은 다음 세 가지 항목으로 설정하였다:

- 문맥 이해도: 입력된 instruction과 input에 대한 상황 맥락을 올바르게 해석했는가.
- 도메인 일치성: 해상 용어(예: 선박명, MMSI, ETA 등)를 정확하고 적절하게 사용했는가.
- 표현의 자연스러움 및 응답 일관성: 실제 통신 상황에 어울리는 표현 방식과 구조로 응답이 이루어졌는가.

2-1. 선박 MMSI 질의

Input: 쉬프트오토 MMSI 확인 부탁합니다.

Model output: 쉬프트오토 MMSI 440123456입니다.

[분석 결과]

모델은 사용자 입력에서 요구한 정보가 ‘쉬프트오토’라는 선박의 MMSI를 질의하는 것임을 정확히 파악하였다. 모델은 해상 용어(MMSI)를 적절하게 사용하였으며, 문장 구조를 간결하고 자연스럽게 구성하였다.

2-2. ETA 보고 질의

Input: 쉬프트오토 ETA 확인 부탁합니다.

Model output: 쉬프트오토 ETA 2024-10-05 8:00입니다.

[분석 결과]

모델은 사용자 입력에서 요구한 정보가 ‘쉬프트오토’라는 선박의 ETA를 질의하는 것임을 정확히 파악하였다. 모델은 해상 용어(ETA)를 적절하게 사용하였으며, 문장 구조를 간결하고 자연스럽게 구성하였다.

2-3. 선박 속도 변경 질의

Input: 귀선의 속도 5.0노트까지 변경 바랍니다.

Model output: 쉬프트오토 SOG 5.0 변경 완료했습니다.

[분석 결과]

모델은 사용자 입력에서 요구한 정보가 ‘쉬프트오토’라는 선박의 속도(SOG)를 변경하는 것임을 정확히 파악하였다. 모델은 해상 용어(SOG)를 적절하게 사용하였으며, 문장 구조를 간결하고 자연스럽게 구성하였다.

정성 평가 결과, FourierFT 모델은 다음과 같은 강점을 보였다:

- 도메인 정보에 대한 정확한 해석과 반영 능력
- 실제 해상 관제 및 선박 간 통신에서 사용되는 표현과 구조의 높은 자연성
- 상황 이해도를 바탕으로 한 응답 재구성 및 응급 상황 대응 가능성

이러한 특성은 모델이 단순한 질문 응답을 넘어, 현실적인 해상 운용 환경에서의 의사소통 보조 역할을 수행할 수 있고, 향후 자율운항 선박 시스템 개발의 핵심 기반이 될 수 있음을 시사한다.

III. 결론

1. 연구 요약 및 시사점

본 연구는 해상 도메인에 특화된 자연어 이해 및 응답 생성을 목표로, FourierFT 기법을 활용한 LLaMA-3-8B 파인튜닝 모델을 구축하고, 이에 대한 정량적·정성적 평가를 수행하였다. 모델 학습을 위해 System Prompt + FLAN-style 형식의 데이터셋을 설계하여, 모델이 사용자 지시와 문맥 정보를 효과적으로 해석하고 응답할 수 있도록 유도하였다.

정량 평가 결과, FourierFT를 적용한 모델은 기존 범용 LLaMA-3-8B 대비 BLEU, ROUGE-L, Domain Accuracy 모든 항목에서 유의미한 성능 향상을 보였고, 정성 평가 또한 모델이 질의 상황에 맞는 표현 구조와 문맥적 흐름을 비교적 자연스럽게 생성한다는 점에서 긍정적인 결과를 보였다. 이는 FourierFT 기법이 도메인 적응성과 파라미터 효율성을 동시에 확보할 수 있는 효과적인 방법임을 시사한다.

종합적으로, 본 연구는 FourierFT 기반 파인튜닝이 해상 도메인 특화 LLM 개발에 있어 실질적인 효과를 갖는 접근 방식임을 실험적으로 입증하였으며, 향후 해상 교신 자동화, 자율 운항 선박, 스마트 항만 서비스 등 다양한 응용 분야에서 활용될 수 있는 기반 기술로 발전할 수 있음을 확인하였다.

참고문헌

- [1] J. L. Vazquez, T. Semmelrock, et al. Using LLM-Generated Draft Replies to Support Human Experts in Responding to Stakeholder Inquiries in Maritime Industry. arXiv preprint arXiv:2412.12732, 2024.
- [2] K. H. Wu, B. C. Yen, and H. Y. Lo. Llamarine: Open-source Maritime Industry-specific Large Language Model. arXiv preprint arXiv:2503.00203, 2025.
- [3] M. Liu, S. Lin, C. Zhang, et al. Parameter-Efficient Fine-Tuning with Discrete Fourier Transform. arXiv preprint arXiv:2405.03003, 2024.
- [4] Vaswani et al. Attention Is All You Need. NeurIPS, 2017.
- [5] E. Hu et al. LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint, arXiv:2106.09685, 2021.
- [6] Wei, J., Bosma et al. Finetuned Language Models Are Zero-Shot Learners. arxiv preprint arXiv:2109.01652, 2022.
- [7] Lechen Zhang et al. SPRING: Improving Large Language Model Performance by System Prompt Optimization. arxiv preprint arxiv: 2410.14826, 2024.

부록: System Prompt

본 연구에서는 다음과 같은 System Prompt를 학습 및 추론 전 단계에서 공통적으로 적용하였다. 해당 프롬프트는 모델에게 해상 도메인 특화 역할(Role) 및 상황, 참고 데이터셋 등을 명시적으로 부여 하며, FLAN-style 구조의 Instruction + Input + Answer 앞에 선행되도록 구성하였다. 이는 도메인 일관성과 응답 품질 향상에 핵심적인 요소로 작용하였다.

```
<<System>>

상황: 당신은 항해 중인 선박으로, 상대 선박의 공식 통신 요청에 응답해야 합니다. 모든 응답은 해상 통신 규약에 따라 정확하고 간결하게 전달되어야 합니다.

역할: 당신은 응답 선박입니다. 입력된 요청(Input)에 대해, 정적 데이터셋에 기반한 정보를 해상 도메인 규칙에 따라 응답해야 합니다.

응답 규칙:
1. 모든 요청은 즉시 수행 가능한 상황으로 간주합니다.
2. 응답은 반드시 1문장 이내로 작성하며, 간결하고 명확하게 전달해야 합니다.
3. 한 요청(input)에 대해 하나의 응답(output)만 생성하며, 중복/반복 응답은 금지됩니다.
4. 부연 설명, 해석, 추가 문장, 표현 변형은 허용되지 않습니다.
5. 조사의 사용은 최소화하며, 명령형 또는 보고체 표현을 간단히 사용합니다.
6. 모든 응답은 반드시 아래의 정적 데이터셋을 기반으로 해야 하며, 요청된 항목이 존재하지 않을 경우, “없습니다” 등 부정 응답을 명확히 포함해야 합니다.

정적 데이터셋:
- 선명: 쉬프트오토
- 쿨싸인: 쉬프트오토토
- MMSI: 440123456
- 위도: 126.6
- 경도: 34.6
- SOG: -2
- COG: 180
- ETD: 2024-10-06 06:00
- ETB: 2024-10-05 14:00
- ETA: 2024-10-05 08:00
- 화물 종류: 없음
- 화물량: 0
- 승무원 수: 6
- 여객 수: 6
- 위치: 거제조선소
- 현재 작업: 급유
- 엔진 위치: 좌현
- 엔진 출력: 전속
- 레더 각도: 10도
- 속도: 5 knots
- 침로: 9도
- 채널 번호: 16
- 투묘 방향: 우현
- 선체 통과 방향: 좌현
- 선종: 연구선
- 기국: 대한민국
- 타요: 0
- 타효: 2.0

아래의 하나의 요청(Input)에 대해, 단 하나의 응답(Answer)만을 생성한 후 종료하세요.
```

그림 5: System Prompt