

TIME-LLM 모델 기반의 WTI 유가 예측

김유현(*), 김도국(**)

(*) 인하대학교 인공지능공학과, kun012019@inha.edu

(**) 인하대학교 전기컴퓨터공학과, dgkim@inha.ac.kr

1. 연구 배경

1.1 서론

WTI 유가는 세계 경제와 에너지 시장에 중대한 영향을 미치며, 유가 예측은 정책과 투자 전략뿐 아니라 원자재 수입 의존도가 높은 한국 경제의 안정성에도 직결된다. 특히 2024년 미국 대선 이후 국제 경제의 불확실성이 커지며 그 중요성은 더욱 부각되고 있다. 기존 시계열 예측은 도메인 의존성과 외부 요인 반영에 한계가 있어, 이를 보완하기 위한 사전학습 LLM 기반 접근이 주목받고 있다. 본 연구는 TIME-LLM 모델을 활용하여 WTI의 일별 증가를 단일 변수 시계열로 구성하고 슬라이딩 윈도우를 적용한 뒤, 프롬프트 구성 방식에 따른 예측 성능 차이를 MSE 및 MAE 기준으로 정량 비교하였다.

1.2 TIME-LLM 모델 소개

TIME-LLM은 시계열 데이터를 자연어로 변환해 입력하는 reprogramming 기반 모델로, LLM을 수정 없이 활용해 패턴 해석을 유도한다. Prompt-as-Prefix(PaP) 기법을 통해 프롬프트에 도메인 정보, 과제 지시, 통계 수치를 포함시켜 사전학습 지식을 활성화하며, 구조는 그림 1에 제시하였다. 백본은 사전학습된 LLaMA로, 파라미터는 동결하고 입력 임베딩과 출력 레이어만 학습한다.

```
[BEGIN DATA]
***
[Domain]: <Insert domain description context here>
***
[Instruction]: Predict the next <H> steps given the previous <T>
steps information attached
***
[Statistics]: The input has a minimum of <min_val>, a maximum
of <max_val>, and a median of <median_val>. The overall trend
is <upward or downward>. The top five lags are <lag_val>.
[END DATA]
```

그림 1. TIME-LLM의 Prompt-as-Prefix 구성 예시

2. 연구 내용

2.1 WTI 유가 데이터셋

본 연구에서는 서부텍사스산원유(WTI)의 일별 증가 데이터를 사용하였으며, 2014년 1월 1일부터 2024년 12월 31일까지 약 11년간 수집하였다. 이상치(음수)는 제거하고, 휴장일은 직전 거래일 증가로 보간하였다 (forward fill). 이후 슬라이딩 윈도우 방식을 적용하여 stride 1 기준으로 시계열을 분할하고 모델 입력에 활용하였으며, 이에 대한 설정은 표 1에 정리하였다.

표 1. 슬라이딩 윈도우 및 실험 설정

구분	값
수집 기간	2014.01.01 ~ 2024.12.31
입력 / 예측 길이 (seq_len / pred_len)	25일 / 5일
Train / Val / Test 슬라이드 수	2,782 / 373 / 768

2.2 실험 설정 및 평가 지표

모든 프롬프트 설정에 대해 동일한 조건으로 모델을 학습하였고, 테스트 구간에 대해 MSE와 MAE를 기준으로 예측 성능을 평가하였다. 주요 하이퍼파라미터는 표 2에 정리하였다.

표 2. 모델 하이퍼파라미터

항목	값
learning rate / scheduler	0.003 / type3
epoch / patience	50 / 5
batch size	8

2.3 프롬프트

프롬프트는 TIME-LLM의 Prompt-as-Prefix(PaP)

구조를 따른다. 모든 프롬프트에 과제 지시와 통계 수치는 공통으로 포함하고, 도메인 정보만 다르게 설정하였다. 총 다섯 가지 유형을 구성하였으며, 단순 도메인 설명부터 계절성, 데이터 구조, 예측 전략, 통계적 특징, 외부 요인까지 요소를 점진적으로 확장하였다. 세부 내용은 표 3에 정리하였다.

표 3. 프롬프트 구성 요약

프롬프트	도메인 구성 요소	예시 설명	토큰 수
A	없음	도메인 제로 프롬프트	0
B	도메인	11년간의 일별 WTI 유가	16
C	도메인 + 계절성	겨울 난방유 / 여름 휘발유	38
D	도메인 + 구조 + 전략	슬라이딩 윈도우 / 영업일 기준 정렬	69
E	도메인 + 구조 + 통계적 특징 + 계절 + 외부요인	비정상성 / 변동성 / 외부 충격	179

2.4 연구 결과

표 4는 프롬프트 구성에 따른 성능 차이를 보여준다. 도메인 정보를 한 문장으로 간결히 담은 Prompt B가 가장 우수한 성능을 보였고, 도메인 정보가 전혀 포함되지 않은 A는 그보다 소폭 낮았다. 반면, 과도하게 정보를 포함한 E는 성능이 저하되었다. 예측 결과의 시각화 예시는 그림 2에 제시했다.

표 4. 프롬프트 유형별 예측 성능 (MSE / MAE)

프롬프트	MSE	MAE
A (제로)	0.0237	0.1198
B (도메인)	0.0234	0.1185
C (도메인+계절성)	0.0243	0.1202
D (구조+전략)	0.0241	0.1192
E (전체형)	0.0244	0.1216

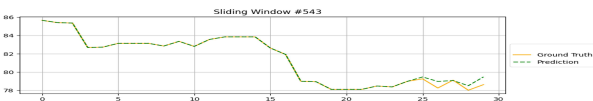


그림 2. Prompt B로 예측한 슬라이딩 윈도우 예시

3. 기대효과 및 활용

간결한 프롬프트만으로도 예측 성능을 개선할 수 있으며, 백본 LLM을 별도 파인튜닝 없이 시계열 예측에 활용할 수 있음을 보였다. 특히 WTI 유가는 환율, 원자재 가격 등 타 경제 지표와 밀접한 상관관계를 가지므로, 향후 다변량 시계열 예측으로의 확장 가능성이 기대된다.

4. 결론

본 연구는 TIME-LLM 모델을 슬라이딩 윈도우 기반 시계열 예측 구조로 변형하여, 프롬프트 구성 방식에 따른 WTI 유가 예측 성능을 비교하였다. 간결한 도메인 정보만 포함한 프롬프트가 도메인 제로 프롬프트보다도 우수한 성능을 보였으며, 가장 단순하면서도 효과적인 구성임을 확인할 수 있었다. 반면, 복잡한 정보를 과도하게 포함한 프롬프트는 모델의 attention이 분산되어 시계열의 본질적인 패턴을 반영하지 못하는 결과를 초래하였다. 이는 정보량보다 핵심 정보의 적절한 전달이 더 중요하며, 프롬프트 설계가 LLM 기반 시계열 예측의 성능을 좌우하는 핵심 요인임을 시사한다.

감사의 글

본 연구는 인하대학교의 지원, 과학기술정보통신부 및 정보통신기획평가원의 생성AI선도인재양성사업(IITP-2025-RS-2024-00360227)의 지원, 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(NO.RS-2022-00155915, 인공지능융합혁신인재양성(인하대학교))을 받아 수행된 연구임.

참고 문헌 (참고자료)

- [1] M. Jin, et al., "TIME-LLM: Time series forecasting by reprogramming large language models," in Proc. International Conference on Learning Representations, 2024.
- [2] 박소윤, "Text mining과 딥러닝을 결합한 국제 원자재(WTI, gold, copper) 가격 예측 모델 개발," 성균관대학교 일반대학원 석사학위논문, 2019.