

Empowering Time Series Analysis with Large Language Models: A Survey

<https://arxiv.org/abs/2402.03182>

0. Introduction

- 대규모 언어 모델(LLMs)의 자연어 처리 분야 성능이 크게 향상됨에도 time series 분석에서는 여러 어려움 존재함
- 시계열 데이터 특성: 방대한 양, 다양한 형태, 비정상성(non-stationarity) 및 concept drift 때문에 모델 적응과 재학습이 부담됨
- 본 서베이 목적: LLM을 시계열 분석에 활용한 기존 방법들을 체계적으로 정리하고 비교함으로써 어떤 접근이 유망한지, 그리고 앞으로의 연구 방향 제시

1. Overview

- LLM 기반 시계열 분석의 일반적 파이프라인 설명됨: 데이터 준비, 프롬프트 설계(prompt design), 미세 조정(fine-tune), 모델 통합(model integration) 등
- 방법 분류(method taxonomy): direct query, tokenization, prompt design, fine-tuning, model integration 다섯 그룹으로 나눔
- 시계열 데이터 종류: 일반 시계열(general), 시공간(spatial-temporal) 시계열 포함됨

2. Challenges

- 시계열의 비정상성(non-stationarity) 및 concept drift이 지속적 모델 적응과 재학습에 장애가 됨
- LLM은 본래 텍스트 데이터를 학습하였기 때문에 수치형 시계열 데이터와의 형식적 차이(modality gap, numerical vs text) 존재함
- 입력 길이, 메모리 요구, 계산 비용 등이 크며, LLM 미세 조정(fine-tuning) 시 과적합 가능성 있음

- 프롬프트(prompt) 설계 및 토큰화(tokenization) 방법 선택이 성능에 민감하게 영향 줌

3. Taxonomy of Methods

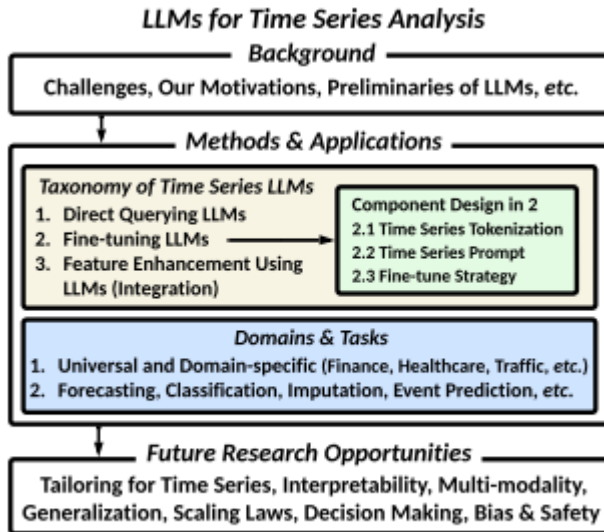


Figure 1: The framework of our survey

- Direct Query: 시계열 데이터를 직접 쿼리 형태로 LLM에 입력하고 결과 얻는 방식
- Tokenization: 수치 시계열을 토큰(token) 혹은 구간(bins)으로 분할하여 LLM이 처리 가능하도록 변환함
- Prompt Design: 작업 설명(prompt), context 정보 포함, few-shot/zero-shot 예시 제공 등의 전략을 통해 LLM 응답 최적화
- Fine-tuning: 사전학습된 LLM을 특정 시계열 태스크(예측, 이상탐지 등)에 맞게 조정 됨
- Model Integration: LLM을 순수한 텍스트 모델이 아니라, 기존의 시계열 모델이나 특화 모듈(GNN, CNN 등)과 결합하는 방식

4. Applications / Case Studies

- 일반 시계열 분석 태스크: 예측, 분류, 이상 탐지 등 다양한 분야에서 LLM 활용됨
- 공간-시간 데이터(spatio-temporal)인 교통, 기상, 센서 네트워크 등 도메인 적용 사례 있음
- 도메인 특화 사례: 금융, 의료 등에서 LLM을 활용한 분석 결과 일부 보고됨

5. Results & Comparisons

Table 1: Taxonomy of time series LLMs. The data type **TS** denotes general time series, **ST** denotes spatial-temporal time series, the prefix **M-** indicates multi-modal inputs. The task entry **Multiple** includes forecasting, classification, imputation and anomaly detection. **Query** denotes direct query the whole LLMs for output, **Token** denotes the design of time series tokenization, **Prompt** indicates the design of textual or parameterized time series prompts, **Fine-tune** indicates if the parameters of LLMs are updated, **Integrate** indicates if LLMs are integrated as part of final model for downstream tasks. Code availability is assessed on January 31st, 2024.

Method	Data Type	Domain	Task	Modeling Strategy					LLM	Code
				Query	Token	Prompt	Fine-tune	Integrate		
Time-LLM [Jin <i>et al.</i> , 2024]	M-TS	General	Forecasting	✗	✓	✓	✓	✗	LLaMA, GPT-2	Yes ^[1]
OFA [Zhou <i>et al.</i> , 2023]	TS	General	Multiple	✗	✓	✗	✓	✗	GPT-2	Yes ^[2]
TEMPO [Cao <i>et al.</i> , 2024]	TS	General	Forecasting	✗	✓	✓	✓	✗	GPT-2	No
TEST [Sun <i>et al.</i> , 2024]	M-TS	General	Forecasting Classification	✗	✓	✓	✗	✓	BERT, GPT-2 ChatGLM, LLaMA2	Yes ^[3]
LLM4TS [Chang <i>et al.</i> , 2023]	TS	General	Forecasting	✗	✓	✗	✓	✗	GPT-2	No
PromptCast [Xue and Salim, 2023]	TS	General	Forecasting	✓	✗	✓	✗	✗	Bart, BERT, etc.	Yes ^[4]
LLMTIME [Gruver <i>et al.</i> , 2023]	TS	General	Forecasting	✓	✓	✗	✗	✗	GPT-3, LLaMA-2	Yes ^[5]
LAMP [Shi <i>et al.</i> , 2023]	TS	General	Event Prediction	✓	✗	✓	✗	✓	GPT-3&3.5, LLaMA-2	Yes ^[6]
[Gunjal and Durrett, 2023]	TS	General	Event Prediction	✓	✗	✓	✗	✗	GPT-3.5, Flan-T5, etc.	No
[Yu <i>et al.</i> , 2023]	M-TS	Finance	Forecasting	✓	✗	✓	✓	✗	GPT-4, Open-LLaMA	No
[Lopez-Lira and Tang, 2023]	M-TS	Finance	Forecasting	✓	✗	✓	✗	✓	ChatGPT	No
[Xie <i>et al.</i> , 2023]	M-TS	Finance	Classification	✓	✗	✓	✗	✗	ChatGPT	No
[Chen <i>et al.</i> , 2023b]	M-TS	Finance	Classification	✗	✗	✓	✗	✓	ChatGPT	Partial ^[7]
METS [Li <i>et al.</i> , 2023]	M-TS	Healthcare	Classification	✓	✗	✓	✗	✓	ClinicalBERT	No
[Jiang <i>et al.</i> , 2023]	M-TS	Healthcare	Classification	✗	✗	✗	✓	✗	NYUTron(BERT)	Yes ^[8]
[Liu <i>et al.</i> , 2023]	M-TS	Healthcare	Forecasting Classification	✓	✗	✓	✓	✗	PaLM	No
AuxMobLCast [Xue <i>et al.</i> , 2022a]	ST	Mobility	Forecasting	✗	✗	✓	✓	✓	BERT, RoBERTa GPT-2, XLNet	Yes ^[9]
LLM-Mob [Wang <i>et al.</i> , 2023]	ST	Mobility	Forecasting	✓	✗	✓	✗	✗	GPT-3.5	Yes ^[10]
ST-LLM [Liu <i>et al.</i> , 2024]	ST	Traffic	Forecasting	✗	✓	✗	✓	✗	LLaMA, GPT-2	No
GATGPT [Chen <i>et al.</i> , 2023a]	ST	Traffic	Imputation	✗	✓	✗	✓	✗	GPT-2	No
LA-GCN [Xu <i>et al.</i> , 2023]	M-ST	Vision	Classification	✗	✓	✗	✗	✓	BERT	Yes ^[11]

[1] <https://github.com/kimmeen/time-llm>

[4] <https://github.com/HaoUNSW/PISA>

[8] <https://github.com/nyuolab/nyuTron>

[2] <https://github.com/DAMO-DI-ML/NeuIPS2023-One-Fits-All>

[5] <https://github.com/ngruver/llmtime>

[9] <https://github.com/cruiseresearchgroup/AuxMobLCast>

[6] <https://github.com/Lampard/lamp>

[10] <https://github.com/dwang233/LLM-Mob>

[7] <https://github.com/ZihanChen1995/ChatGPT-GNN-StockPredict>

[11] <https://github.com/damNull/LAGCN>

[3] <https://openreview.net/forum?id=Tub4nZVb0g> (supplementary material)

- 아직 다양한 실험이 이루어졌으나, 방법들 간 비교는 프롬프트 디자인, 토큰화 방식, 모델 통합 전략 등에 따라 성능 차이 큼
- Direct query 방식은 간단하고 빠르나 정확도 한계 있음
- Fine-tuning 및 모델 통합 방식이 더 높은 성능 제공하나 비용·리소스 요구
- Tokenization 및 prompt 설계가 특히 성능 최적화에 중요한 변수임

6. Insight / Future Directions

- 시계열 + LLM 연구는 modality gap 해소(tokenization, aligning 등) 및 적절한 프롬프트 설계가 핵심됨
- domain generalization, 적은 데이터 환경(few/zero-shot)에서의 안정성 강화 필요
- 계산 효율성, 메모리 최적화, 리소스 제약 환경에서의 활용 가능성 탐구 중요

- interpretability(해석 가능성), 편향과 안전성(bias & safety) 이슈도 중요하게 다루는 방향
- 멀티모달 시계열 데이터, 실시간 처리(real-time), 지속적 학습(continuous adaptation) 등이 향후 연구 주제