Squirrel Ai Learning

LLM and Foundation Models for Time Series Analysis

Qingsong Wen

Head of AI & Chief Scientist

Squirrel Ai Learning, Bellevue, WA, USA

Dec. 15, 2024

Outline

- ***Introduction***
  - Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, Qingsong Wen*, **"Foundation Models for Time Series Analysis: A Tutorial and Survey", KDD 2024**.
  - Ming Jin, Qingsong Wen*, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Yu Zheng, Lei Chen, Hui Xiong**, "Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook", arXiv 2023**.
- **Large Language Model for Time Series Analysis**
- **Foundation Model for Time Series Analysis**
- **Future Directions**

Time Series Data is Ubiquitous

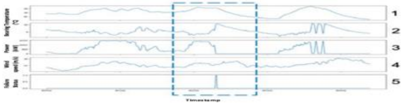
- A wide range of time series data across industries
  - AIOps & Cloud Computing
  - Energy Forecasting
  - Educational Data Mining (EDM)
  - Others (Finance, Healthcare, ...)

stocks

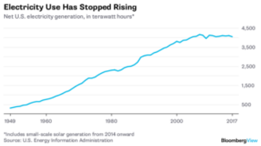
sales

goods consumption

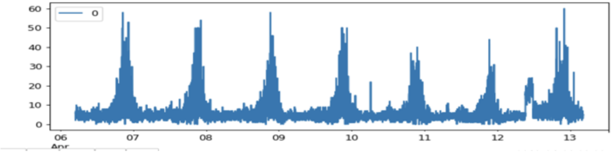
sensor



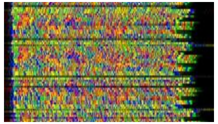
power demand



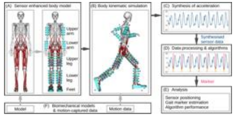
Cloud service monitoring



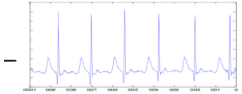
DNA sequence



motion detect

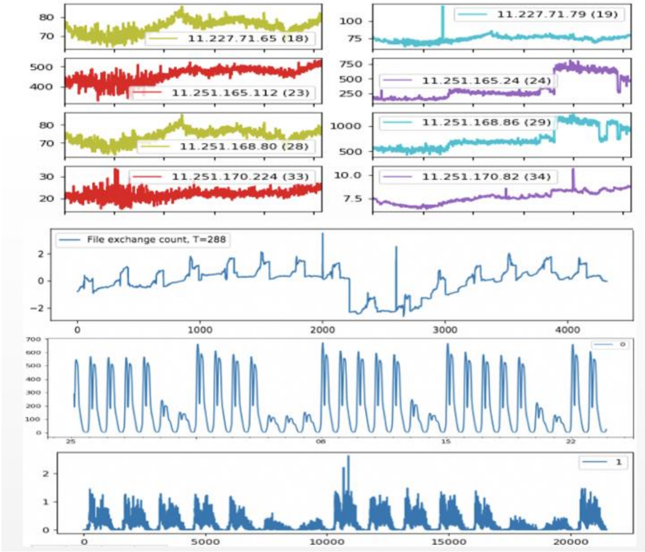


ECG



Time Series in Real World

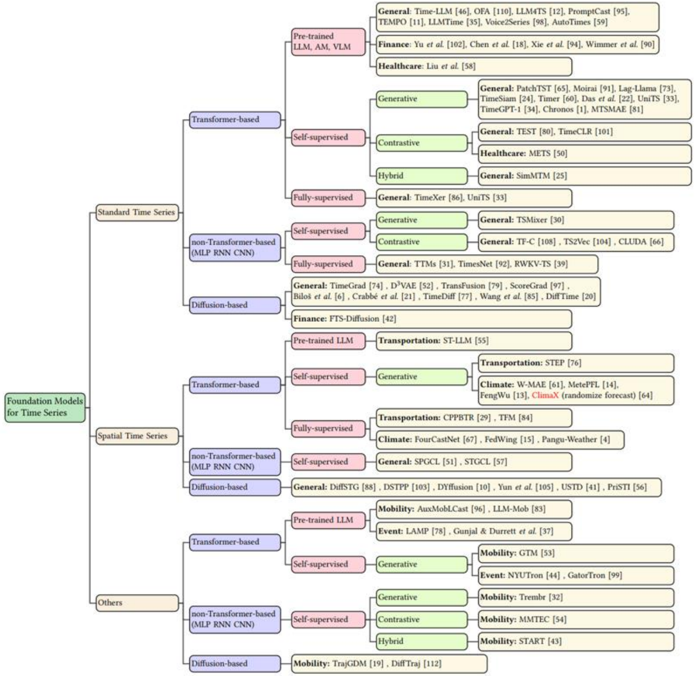
- Time series data with complex patterns
  - Noises and outliers
  - Periodicity/Seasonality: variable periodic components
- Time series with big data
  - Hundreds of millions metrics or even more
  - High frequency and real-time processing
  - Low deployment cost
- Time series with rich context information
  - Text information associated with time series
  - Multimodal time series data
- Diverse time series tasks
  - Forecasting, anomaly detection, classification, ...

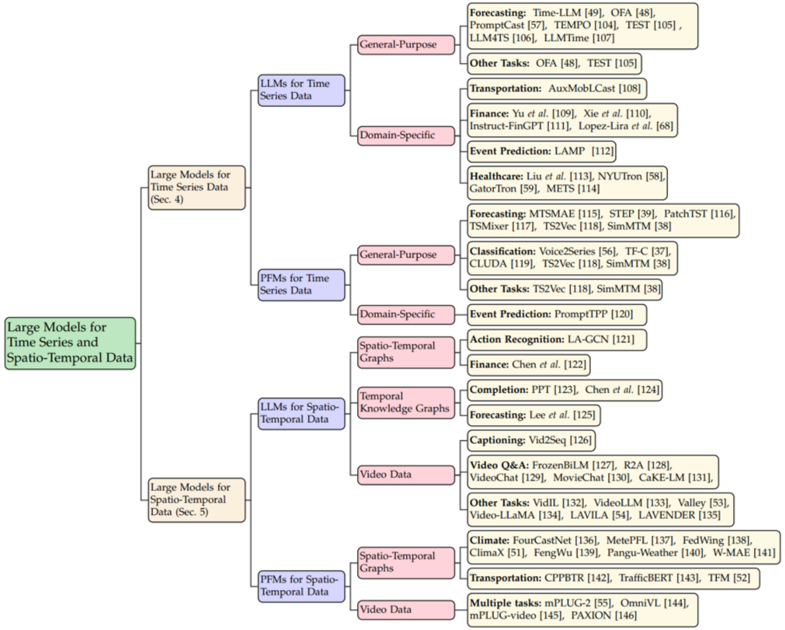


We need: flexible & general time series models to handle these challenges effectively

Two Effective Approaches : LLM and FM

- LLM for Time Series
  - Leverage/repurpose **LLM** for time series analysis
  - E.g.: **Time-LLM**, LLMTime, AutoTimes, OFA, UniST, ...
- FM for Time Series
  - **Training from scratch** based on time series data
  - E.g.: **Time-MoE**, Moirai, TimesFM, Chronos, Moment,...





[1] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, Qingsong Wen*, "Foundation Models for Time Series Analysis: A Tutorial and Survey", KDD 2024.
[2] Ming Jin, Qingsong Wen*, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Yu Zheng, Lei Chen, Hui Xiong, "Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook", arXiv 2023.

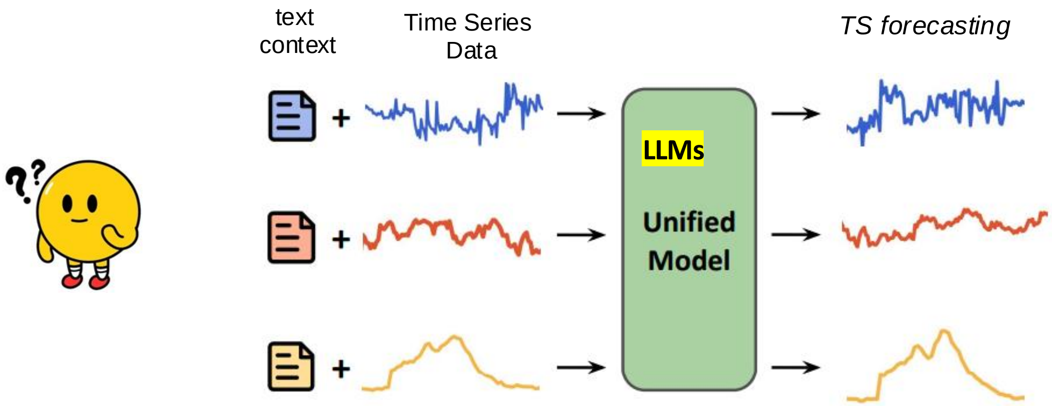
Outline

- ❑ **Introduction**
- ➢ ***Large Language Model for Time Series Analysis***
  - Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan*, Qingsong Wen*, **"Time-LLM: Time Series Forecasting by Reprogramming Large Language Models", ICLR 2024**.
- ❑ **Foundation Model for Time Series Analysis**
- ❑ **Future Directions**

Motivation

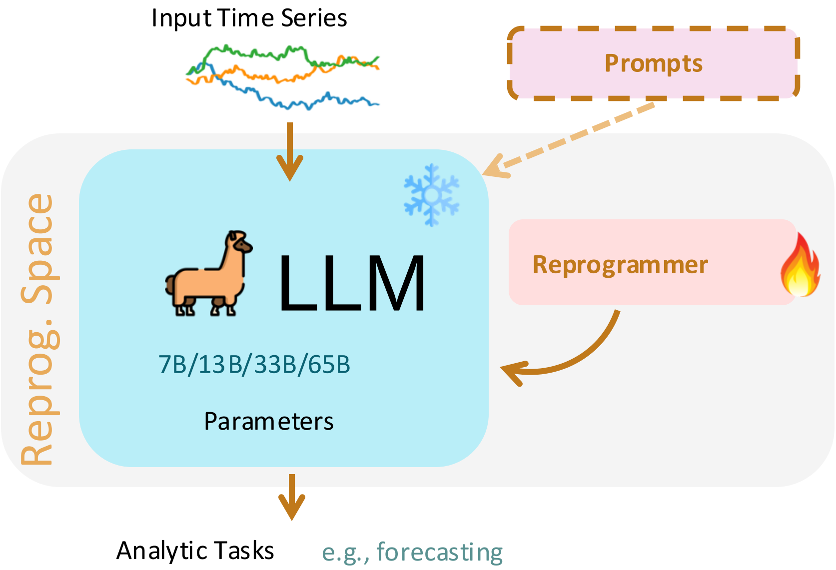
- Multimodal large language models

How time series analysis benefits from the recent advances of LLMs?



Motivation

- Reprogramming makes LLMs **instantly ready, more powerful** for time series tasks



We keep pretrained LLMs intact and **only fine-tune reprogrammer** to achieve certain alignments

Reprogramming ≈ Adaptation + Alignment

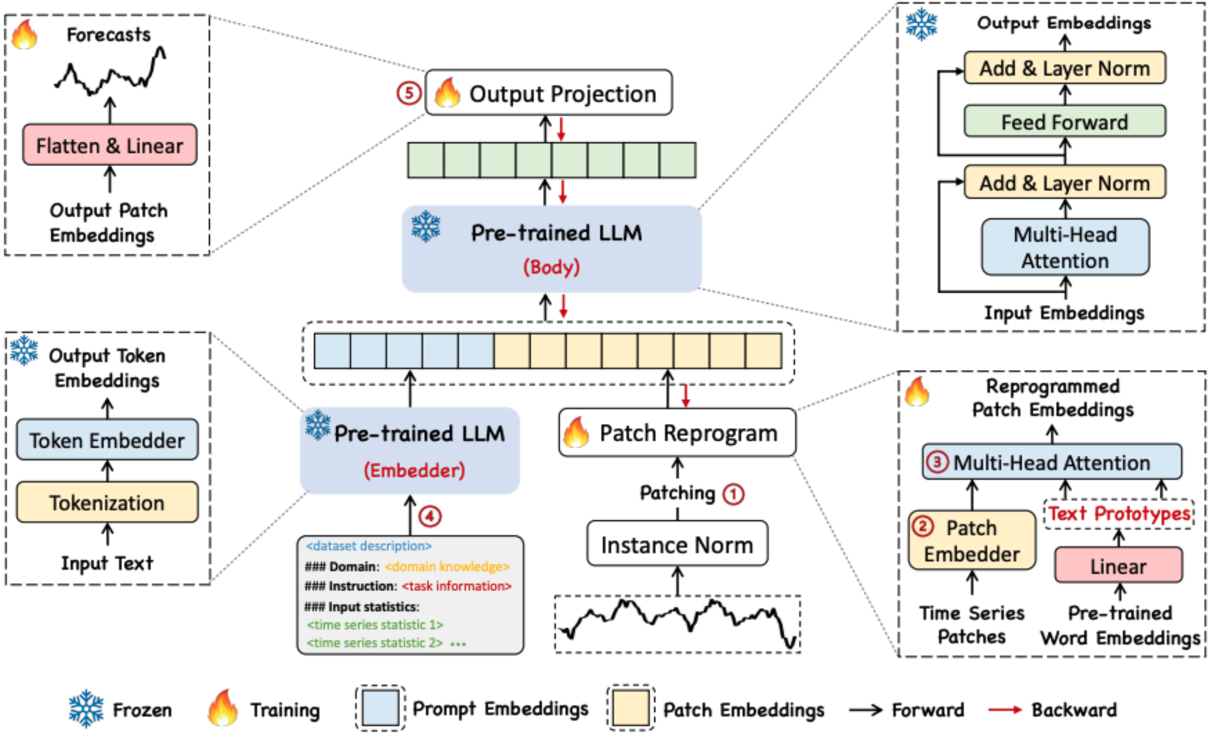
Adaptation makes LLMs to understand how to process the input time series data → Breaking domain isolation and enabling knowledge sharing

Alignment further eliminates domain boundary to facilitate knowledge acquiring

Chen, Pin-Yu. "Model reprogramming: Resource-efficient cross-domain machine learning." AAAI Conference on Artificial Intelligence, 2024.

Time-LLM: Architecture

TL;DR Domain expert knowledge & Task instructions + Reprogrammed input time series = Significantly better forecasts



Unlocking the LLM's ability for time series

Cross-modal Adaptation: ① ② ⑤

Cross-modal Alignment: ③ ④

- **Patch Reprogramming:** we reprogram TS patch embeddings into the source data representation space to align the modalities of time series and natural language to activate the backbone's time series understanding and reasoning capabilities.
- **Prompt-as-Prefix:** natural language-based prompts (domain knowledge & task instructions & input statistics) can act as prefixes to enrich the input context and guide the transformation of reprogrammed TS patches

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan*, Qingsong Wen*, "Time-LLM: Time Series Forecasting by Reprogramming Large Language Models", **ICLR 2024**.

Time-LLM: Few-shot & Zero-shot

Table 3: Few-shot learning on 10% training data. We use the same protocol in Tab. 1. All results are averaged from four different forecasting horizons: $H \in \{96, 192, 336, 720\}$. Our full results are in Appendix E.

Methods	TIME-LLM (**Ours**)		GPT4TS (2023a)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2022)		Autoformer (2021)		Stationary (2022)		ETSformer (2022)		LightTS (2022a)		Informer (2021)		Reformer (2020)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1$	**0.556**	**0.522**	0.590	0.525	0.691	0.600	0.633	0.542	0.869	0.628	0.639	0.561	0.702	0.596	0.915	0.639	1.180	0.834	1.375	0.877	1.199	0.809	1.249	0.833
$ETTh2$	**0.370**	**0.394**	0.397	0.421	0.605	0.538	0.415	0.431	0.479	0.465	0.466	0.475	0.488	0.499	0.462	0.455	0.894	0.713	2.655	1.160	3.872	1.513	3.485	1.486
$ETTm1$	**0.404**	**0.427**	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605	0.802	0.628	0.797	0.578	0.980	0.714	0.971	0.705	1.192	0.821	1.426	0.856
$ETTm2$	**0.277**	**0.323**	0.293	0.335	0.316	0.368	0.296	0.343	0.320	0.353	0.463	0.488	1.342	0.930	0.332	0.366	0.447	0.487	0.987	0.756	3.370	1.440	3.978	1.587
$Weather$	**0.234**	**0.273**	0.238	0.275	0.241	0.283	0.242	0.279	0.279	0.301	0.284	0.324	0.300	0.342	0.318	0.323	0.318	0.360	0.289	0.322	0.597	0.495	0.546	0.469
ECL	**0.175**	0.270	0.176	**0.269**	0.180	0.280	0.180	0.273	0.323	0.392	0.346	0.427	0.431	0.478	0.444	0.480	0.660	0.617	0.441	0.489	1.195	0.891	0.965	0.768
$Traffic$	**0.429**	0.306	0.440	0.310	0.447	0.313	0.430	**0.305**	0.951	0.535	0.663	0.425	0.749	0.446	1.453	0.815	1.914	0.936	1.248	0.684	1.534	0.811	1.551	0.821
1st Count	**8**		1		0		1		0		0		0		0		0		0		0		0	

Table 5: Zero-shot learning results. **Red**: the best, Blue: the second best. Appendix E shows our detailed results.

Methods	TIME-LLM (**Ours**)		GPT4TS (2023a)		LLMTime (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1 \to ETTh2$	**0.353**	**0.387**	0.406	0.422	0.992	0.708	0.493	0.488	0.380	0.405	0.421	0.431
$ETTh1 \to ETTm2$	**0.273**	**0.340**	0.325	0.363	1.867	0.869	0.415	0.452	0.314	0.360	0.327	0.361
$ETTh2 \to ETTh1$	**0.479**	**0.474**	0.757	0.578	1.961	0.981	0.703	0.574	0.565	0.513	0.865	0.621
$ETTh2 \to ETTm2$	**0.272**	**0.341**	0.335	0.370	1.867	0.869	0.328	0.386	0.325	0.365	0.342	0.376

Time-LLM: Full-shot Forecasting

Table 1: Long-term forecasting results. All results are averaged from four different forecasting horizons: $H \in \{24, 36, 48, 60\}$ for ILI and $\{96, 192, 336, 720\}$ for the others. A lower value indicates better performance. **Red**: the best, Blue: the second best. Our full results are in Appendix D.

Methods	TIME-LLM (**Ours**)		GPT4TS (2023a)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2022)		Autoformer (2021)		Stationary (2022)		ETSformer (2022)		LightTS (2022a)		Informer (2021)		Reformer (2020)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$ETTh1$	**0.408**	**0.423**	0.465	0.455	0.422	0.437	0.413	0.430	0.458	0.450	0.440	0.460	0.496	0.487	0.570	0.537	0.542	0.510	0.491	0.479	1.040	0.795	1.029	0.805
$ETTh2$	0.334	0.383	0.381	0.412	0.431	0.446	**0.330**	**0.379**	0.414	0.427	0.437	0.449	0.450	0.459	0.526	0.516	0.439	0.452	0.602	0.543	4.431	1.729	6.736	2.191
$ETTm1$	**0.329**	**0.372**	0.388	0.403	0.357	0.378	0.351	0.380	0.400	0.406	0.448	0.452	0.588	0.517	0.481	0.456	0.429	0.425	0.435	0.437	0.961	0.734	0.799	0.671
$ETTm2$	**0.251**	**0.313**	0.284	0.339	0.267	0.333	0.255	0.315	0.291	0.333	0.305	0.349	0.327	0.371	0.306	0.347	0.293	0.342	0.409	0.436	1.410	0.810	1.479	0.915
$Weather$	**0.225**	**0.257**	0.237	0.270	0.248	0.300	**0.225**	0.264	0.259	0.287	0.309	0.360	0.338	0.382	0.288	0.314	0.271	0.334	0.261	0.312	0.634	0.548	0.803	0.656
ECL	**0.158**	**0.252**	0.167	0.263	0.166	0.263	0.161	**0.252**	0.192	0.295	0.214	0.327	0.227	0.338	0.193	0.296	0.208	0.323	0.229	0.329	0.311	0.397	0.338	0.422
$Traffic$	**0.388**	0.264	0.414	0.294	0.433	0.295	0.390	**0.263**	0.620	0.336	0.610	0.376	0.628	0.379	0.624	0.340	0.621	0.396	0.622	0.392	0.764	0.416	0.741	0.422
ILI	**1.435**	0.801	1.925	0.903	2.169	1.041	1.443	**0.797**	2.139	0.931	2.847	1.144	3.006	1.161	2.077	0.914	2.497	1.004	7.382	2.003	5.137	1.544	4.724	1.445
1st Count	**7**		0		0		5		0		0		0		0		0		0		0		0	

Table 2: Short-term time series forecasting results on M4. The forecasting horizons are in [6, 48] and the three rows provided are weighted averaged from all datasets under different sampling intervals. A lower value indicates better performance. **Red**: the best, Blue: the second best. More results are in Appendix D.

	Methods	TIME-LLM (**Ours**)	GPT4TS (2023a)	TimesNet (2023)	PatchTST (2023)	N-HiTS (2023b)	N-BEATS (2020)	ETSformer (2022)	LightTS (2022a)	DLinear (2023)	FEDformer (2022)	Stationary (2022)	Autoformer (2021)	Informer (2021)	Reformer (2020)
Average	SMAPE	**11.983**	12.69	12.88	12.059	12.035	12.25	14.718	13.525	13.639	13.16	12.780	12.909	14.086	18.200
	MASE	**1.595**	1.808	1.836	1.623	1.625	1.698	2.408	2.111	2.095	1.775	1.756	1.771	2.718	4.223
	OWA	**0.859**	0.94	0.955	0.869	0.869	0.896	1.172	1.051	1.051	0.949	0.930	0.939	1.230	1.775

Time-LLM: Ablation & Efficiency

Table 6: Ablations on ETTh1 and ETTm1 in predicting 96 and 192 steps ahead (MSE reported). **Red**: the best.

Variant	Long-term Forecasting				Few-shot Forecasting			
	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192	ETTh1-96	ETTh1-192	ETTm1-96	ETThm1-192
A.1 Llama (**Default**; 32)	**0.362**	**0.398**	**0.272**	**0.310**	**0.448**	**0.484**	**0.346**	**0.373**
A.2 Llama (8)	0.389	0.412	0.297	0.329	0.567	0.632	0.451	0.490
A.3 GPT-2 (12)	0.385	0.419	0.306	0.332	0.548	0.617	0.447	0.509
A.4 GPT-2 (6)	0.394	0.427	0.311	0.342	0.571	0.640	0.468	0.512
B.1 w/o Patch Reprogramming	0.410	0.412	0.310	0.342	0.498	0.570	0.445	0.487
B.2 w/o Prompt-as-Prefix	0.398	0.423	0.298	0.339	0.521	0.617	0.432	0.481
C.1 w/o Dataset Context	0.402	0.417	0.298	0.331	0.491	0.538	0.392	0.447
C.2 w/o Task Instruction	0.388	0.420	0.285	0.327	0.476	0.529	0.387	0.439
C.3 w/o Statistical Context	0.391	0.419	0.279	0.347	0.483	0.547	0.421	0.461

Table 7: Efficiency analysis of TIME-LLM

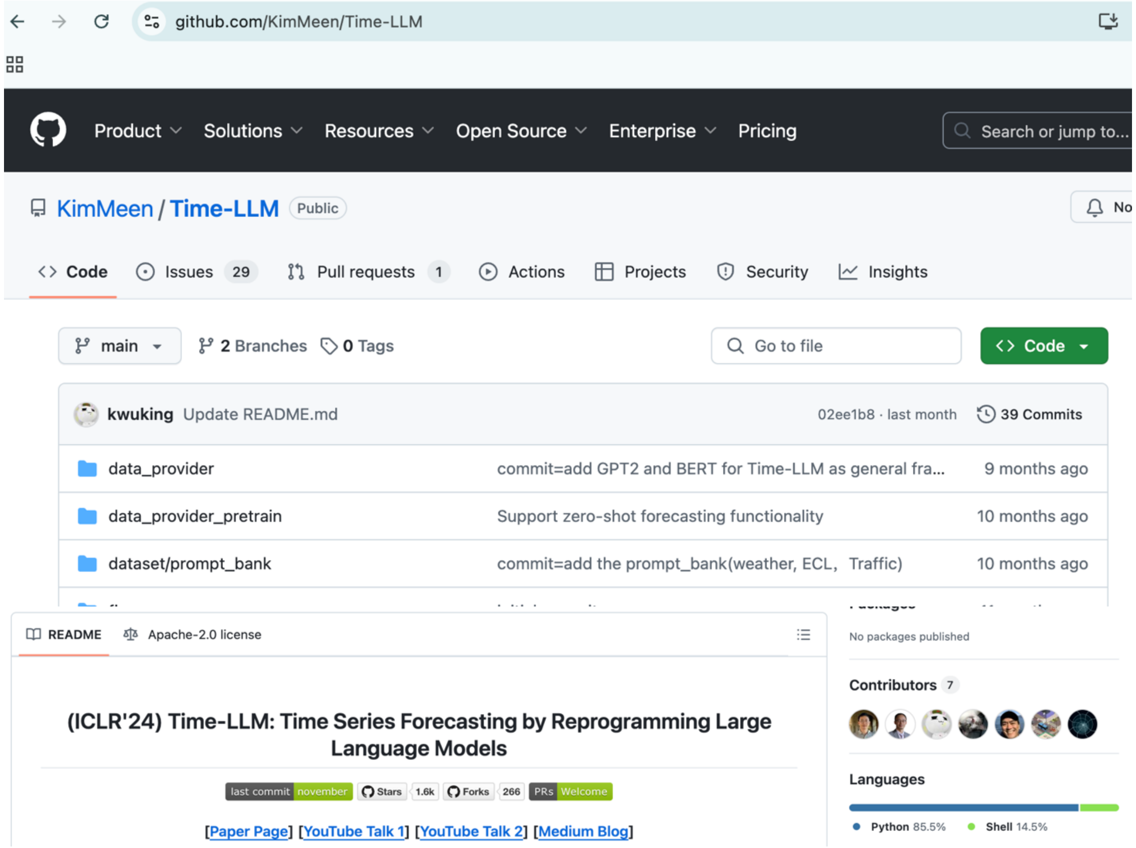
Length	ETTh1-96			ETTh1-192		
Metric	Param. (M)	Mem. (MiB)	Speed(s/iter)	Param. (M)	Mem. (MiB)	Speed(s/iter)
D.1 LLama (32)	3404.53	32136	0.517	3404.57	33762	0.582
D.2 LLama (8)	975.83	11370	0.184	975.87	12392	0.192
D.3 w/o LLM	6.39	3678	0.046	6.42	3812	0.087

Table 17: Efficiency comparison between model reprogramming and parameter-efficient fine-tuning (PEFT) with QLoRA (Dettmers et al., 2023) on ETTh1 dataset in forecasting two different steps ahead.

Length		ETTh1-96			ETTh1-336		
Metric		Trainable Param. (M)	Mem. (MiB)	Speed(s/iter)	Trainable Param. (M)	Mem. (MiB)	Speed(s/iter)
Llama (8)	QLoRA	12.60	14767	0.237	12.69	15982	0.335
	Reprogram	5.62	11370	0.184	5.71	13188	0.203
Llama (32)	QLoRA	50.29	45226	0.697	50.37	49374	0.732
	Reprogram	6.39	32136	0.517	6.48	37988	0.632

- **Language model variants:** The scaling law retains with the LLM reprogramming.
- **Cross-modality alignment:** (1) we find that the alignment is crucial (see B.1 and B.2); (2) domain knowledge and task instructions are both valuable (C.1-C.3) and can be integrated via Prompt-as-Prefix (PaP)
- **Reprogramming efficiency:** (1) our reprogramming network is lightweight in activating the LLM's ability for time series forecasting (see D.3 -- i.e., fewer than 6.6M trainable parameters; only around 0.2% of the parameters in Llama-7B); (2) this is favorable even compared to parameter-efficient fine-tuning (PEFT; Tab. 17)

Open-source:

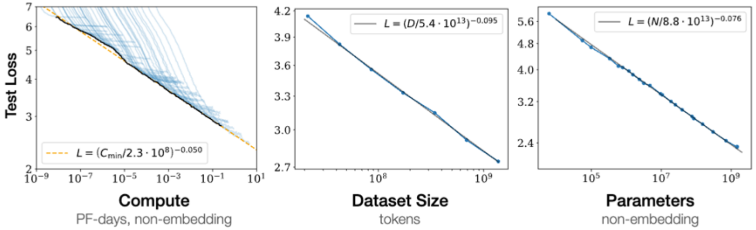

github.com/KimMeen/Time-LLM
Product
Solutions
Resources
Open Source
Enterprise
Pricing
Search or jump to...
KimMeen / Time-LLM
Public
Code
Issues 29
Pull requests 1
Actions
Projects
Security
Insights
main
2 Branches
0 Tags
Go to file
Code
kwuking Update README.md
02ee1b8 · last month
39 Commits
data_provider
commit=add GPT2 and BERT for Time-LLM as general fra...
9 months ago
data_provider_pretrain
Support zero-shot forecasting functionality
10 months ago
dataset/prompt_bank
commit=add the prompt_bank(weather, ECL, Traffic)
10 months ago
README
Apache-2.0 license
(ICLR'24) Time-LLM: Time Series Forecasting by Reprogramming Large Language Models
last commit november
Stars 1.6k
Forks 266
PRs Welcome
[Paper Page] [YouTube Talk 1] [YouTube Talk 2] [Medium Blog]
No packages published
Contributors 7
Languages
Python 85.5%
Shell 14.5%

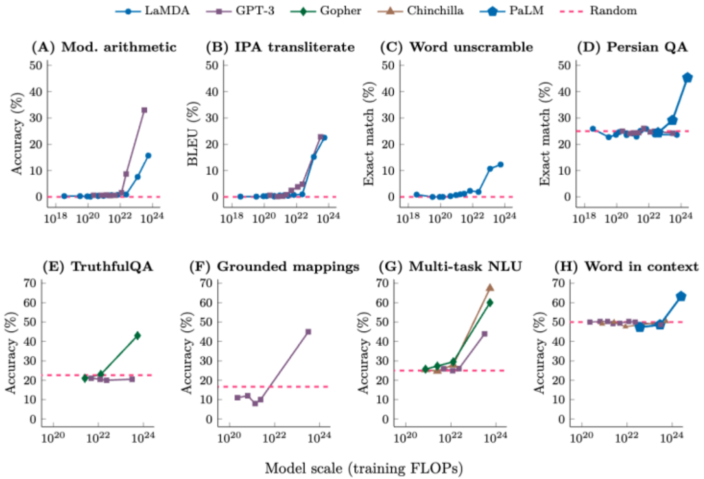
Outline

- ❑ **Introduction**
- ➢ **Large Language Model for Time Series Analysis**
- ➢ ***Foundation Model for Time Series Analysis***
  - Shi, Xiaoming, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen*, and Ming Jin*. **"Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts." arXiv:2409.16040, 2024.**
- ❑ **Future Directions**

Motivation

- Scaling Laws & Capabilities





How about Time Series ?

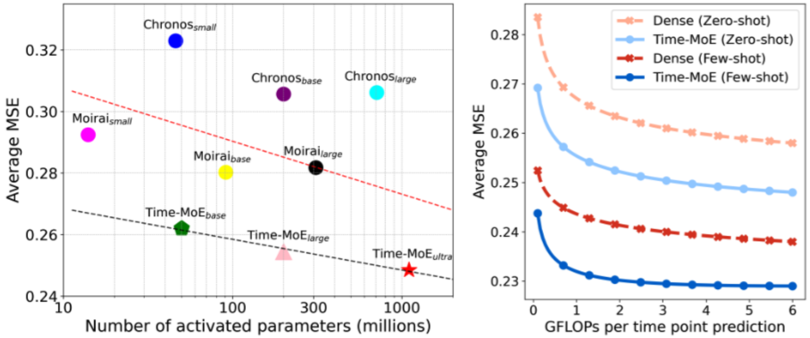


Figure 1: Performance overview. (**Left**) Comparison between TIME-MOE models and state-of-the-art time series foundation models, reporting the average zero-shot performance across six benchmark datasets. (**Right**) Comparison of few- and zero-shot performance between TIME-MOE and dense variants, with similar effective FLOPs per time series token, across the same six benchmarks.

➢ time series analysis benefits from the scaling laws

Jared Kaplan, et al. "Scaling laws for neural language models."arXiv:2001.08361, 2020.
Jason Wei, et al., Emergent Abilities of Large Language Models. *TMLR,* 2022.

Time-300B, Time-MoE

Key statistics of the pre-training dataset `Time-300B` from various domains.

	Energy	Finance	Healthcare	Nature	Sales	Synthetic	Transport	Web	Other	Total
# Seqs.	2,875,335	1,715	1,752	31,621,183	110,210	11,968,625	622,414	972,158	40,265	48,220,929
# Obs.	15.981 B	413.696 K	471.040 K	279.724 B	26.382 M	9.222 B	2.130 B	1.804 B	20.32 M	309.09 B
%	5.17 %	0.0001%	0.0001%	90.50 %	0.008 %	2.98%	0.69 %	0.58 %	0.006 %	100%

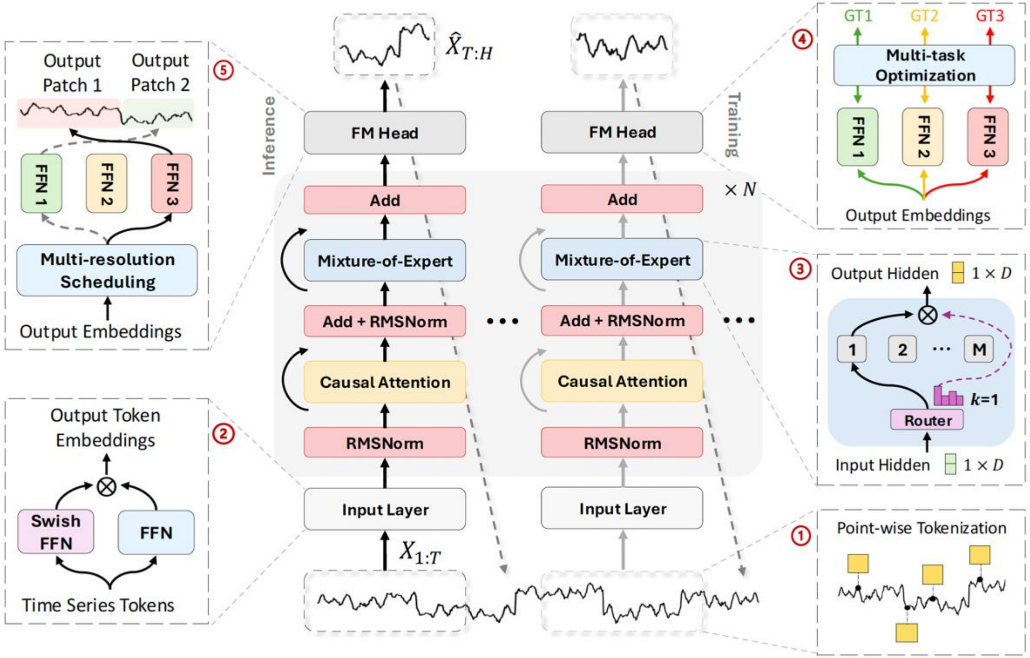
A high-level summary of TIME-MOE model configurations.

	Layers	Heads	Experts	K	d_{model}	d_{ff}	d_{expert}	Activated Params	Total Params
TIME-MOE$_{base}$	12	12	8	2	384	1536	192	50 M	113 M
TIME-MOE$_{large}$	12	12	8	2	768	3072	384	200 M	453 M
TIME-MOE$_{ultra}$	36	16	8	2	1024	4096	512	1.1 B	2.4 B

- Time-300B: the largest open-access time series data collection
- Time-MoE: the first work to scale time series foundation models up to 2.4 billion parameters

Shi, Xiaoming, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen*, and Ming Jin*. "Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts." *arXiv preprint arXiv:2409.16040* (2024).
Shi, Xiaoming, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen*, and Ming Jin*. "Scaling to Billion Parameters for Time Series Foundation Models with Mixture of Experts." In *NeurIPS Workshop on Time Series in the Age of Large Models (2024)*.

Time-MoE: Architecture



The architecture of TIME-MOE, which is a decoder-only model. Given an input time series of arbitrary length, ① we first tokenize it into a sequence of data points, ② which are then encoded. These tokens are processed through N-stacked backbone layers, primarily consisting of causal multi-head self-attention and ③ sparse temporal mixture-of-expert layers. During training, ④ we optimize forecasting heads at multiple resolutions. For model inference, TIME-MOE provides forecasts of flexible length by ⑤ dynamically scheduling these heads.

$$\mathbf{h}_t^0 = \mathrm{SwiGLU}(x_t) = \mathrm{Swish}\,(W x_t) \otimes (V x_t)$$

$$\mathbf{u}_t^l = \mathrm{SA}\left(\mathrm{RMSNorm}\left(\mathbf{h}_t^{l-1}\right)\right) + \mathbf{h}_t^{l-1},$$
$$\bar{\mathbf{u}}_t^l = \mathrm{RMSNorm}\left(\mathbf{u}_t^l\right),$$
$$\mathbf{h}_t^l = \mathrm{Mixture}\left(\bar{\mathbf{u}}_t^l\right) + \mathbf{u}_t^l.$$

MoE Structure (-> efficiency and capacity)

$$\mathrm{Mixture}\left(\bar{\mathbf{u}}_t^l\right) = g_{N+1,t}\,\mathrm{FFN}_{N+1}\left(\bar{\mathbf{u}}_t^l\right) + \sum_{i=1}^{N}\left(g_{i,t}\,\mathrm{FFN}_i\left(\bar{\mathbf{u}}_t^l\right)\right)$$

$$g_{i,t} = \begin{cases} s_{i,t}, & s_{i,t} \in \mathrm{Topk}(\{s_{j,t} | 1 \le j \le N\}, K), \\ 0, & \text{otherwise}, \end{cases}$$
$$g_{N+1,t} = \mathrm{Sigmoid}\left(\mathbf{W}_{N+1}^l \bar{\mathbf{u}}_t^l\right),$$
$$s_{i,t} = \mathrm{Softmax}_i\left(\mathbf{W}_i^l \mathbf{u}_t^l\right),$$

Huber loss (-> outlier)
Auxiliary loss (-> routing collapse)
Multi-resolution Forecasting (-> various horizons)

$$\mathcal{L} = \frac{1}{P}\sum_{j=1}^{P}\mathcal{L}_{\mathrm{ar}}\left(\mathbf{X}_{t+1:t+p_j}, \hat{\mathbf{X}}_{t+1:t+p_j}\right) + \alpha\mathcal{L}_{\mathrm{aux}},$$

$$\mathcal{L}_{\mathrm{ar}}(x_t, \hat{x}_t) = \begin{cases} \frac{1}{2}(x_t - \hat{x}_t)^2, & \text{if } |x_t - \hat{x}_t| \le \delta, \\ \delta \times \left(|x_t - \hat{x}_t| - \frac{1}{2} \times \delta\right), & \text{otherwise}, \end{cases} \quad \mathcal{L}_{\mathrm{aux}} = N\sum_{i=1}^{N} f_i r_i,$$

$$f_i = \frac{1}{KT}\sum_{t=1}^{T}\mathbb{I}\left(\text{Time point } t \text{ selects Expert } i\right), \quad r_i = \frac{1}{T}\sum_{t=1}^{T} s_{i,t},$$

Time-MoE: Zero-shot Performance

Table 3: Full results of zero-shot forecasting experiments. A lower MSE or MAE indicates a better prediction. TimesFM, due to its use of Weather datasets in pretraining, is not evaluated on these two datasets and is denoted by a dash (–). **Red**: the best, Blue: the 2nd best.

Models		Time-MoE (Ours)						Zero-shot Time Series Models															
		Time-MoE$_{base}$		Time-MoE$_{large}$		Time-MoE$_{ultra}$		Moirai$_{small}$		Moirai$_{base}$		Moirai$_{large}$		TimesFM		Moment		Chronos$_{small}$		Chronos$_{base}$		Chronos$_{large}$	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.357	0.381	0.350	0.382	**0.349**	**0.379**	0.401	0.402	0.376	0.392	0.381	0.388	0.414	0.404	0.688	0.557	0.466	0.409	0.440	0.393	0.441	0.390
	192	**0.384**	**0.404**	0.388	0.412	0.395	0.413	0.435	0.421	0.412	0.413	0.434	0.415	0.465	0.434	0.688	0.560	0.530	0.450	0.492	0.426	0.502	0.424
	336	**0.411**	0.434	**0.411**	0.430	0.447	0.453	0.438	0.434	0.433	**0.428**	0.495	0.445	0.503	0.456	0.675	0.563	0.570	0.486	0.550	0.462	0.576	0.467
	720	0.449	0.477	**0.427**	0.455	0.457	0.462	0.439	0.454	0.447	**0.444**	0.611	0.510	0.511	0.481	0.683	0.585	0.615	0.543	0.882	0.591	0.835	0.583
	AVG	0.400	0.424	**0.394**	**0.419**	0.412	0.426	0.428	0.427	0.417	**0.419**	0.480	0.439	0.473	0.443	0.683	0.566	0.545	0.472	0.591	0.468	0.588	0.466
ETTh2	96	0.305	0.359	0.302	0.354	**0.292**	0.352	0.297	0.336	0.294	**0.330**	0.296	**0.330**	0.315	0.349	0.342	0.396	0.307	0.356	0.308	0.343	0.320	0.345
	192	0.351	0.386	0.364	0.385	**0.347**	0.379	0.368	0.381	0.365	0.375	0.361	**0.371**	0.388	0.395	0.354	0.402	0.376	0.401	0.384	0.392	0.406	0.399
	336	0.391	0.418	0.417	0.425	0.406	0.419	0.370	0.393	0.376	**0.390**	0.390	**0.390**	0.422	0.427	**0.356**	0.407	0.408	0.431	0.429	0.430	0.492	0.453
	720	0.419	0.454	0.537	0.496	0.439	0.447	0.411	0.426	0.416	0.433	0.423	**0.418**	0.443	0.454	**0.395**	0.434	0.604	0.533	0.501	0.477	0.603	0.511
	AVG	0.366	0.404	0.405	0.415	0.371	0.399	**0.361**	0.384	0.362	0.382	0.367	**0.377**	0.392	0.406	**0.361**	0.409	0.424	0.430	0.405	0.410	0.455	0.427
ETTm1	96	0.338	0.368	0.309	0.357	**0.281**	**0.341**	0.418	0.392	0.363	0.356	0.380	0.361	0.361	0.370	0.654	0.527	0.511	0.423	0.454	0.408	0.457	0.403
	192	0.353	0.388	0.346	0.381	**0.305**	**0.358**	0.431	0.405	0.388	0.375	0.412	0.383	0.414	0.405	0.662	0.532	0.618	0.485	0.567	0.477	0.530	0.450
	336	0.381	0.413	0.373	0.408	**0.369**	0.395	0.433	0.412	0.416	**0.392**	0.436	0.400	0.445	0.429	0.672	0.537	0.683	0.524	0.662	0.525	0.577	0.481
	720	0.504	0.493	0.475	0.477	0.469	0.472	0.462	0.432	**0.460**	**0.418**	0.462	0.420	0.512	0.471	0.692	0.551	0.748	0.566	0.900	0.591	0.660	0.526
	AVG	0.394	0.415	0.376	0.405	**0.356**	0.391	0.436	0.410	0.406	**0.385**	0.422	0.391	0.433	0.418	0.670	0.536	0.640	0.499	0.645	0.500	0.555	0.465
ETTm2	96	0.201	0.291	**0.197**	0.286	0.198	0.288	0.214	0.288	0.205	0.273	0.211	0.274	0.202	**0.270**	0.260	0.335	0.209	0.291	0.199	0.274	**0.197**	0.271
	192	0.258	0.334	0.250	0.322	**0.235**	**0.312**	0.284	0.332	0.275	0.316	0.281	0.318	0.289	0.321	0.289	0.350	0.280	0.341	0.261	0.322	0.254	0.314
	336	0.324	0.373	0.337	0.375	**0.293**	**0.348**	0.331	0.362	0.329	0.350	0.341	0.355	0.360	0.366	0.324	0.369	0.354	0.390	0.326	0.366	0.313	0.353
	720	0.488	0.464	0.480	0.461	0.427	0.428	0.402	**0.408**	0.437	0.411	0.485	0.428	0.462	0.430	**0.394**	0.409	0.553	0.499	0.455	0.439	0.416	0.415
	AVG	0.317	0.365	0.316	0.361	**0.288**	0.344	0.307	0.347	0.311	**0.337**	0.329	0.343	0.328	0.346	0.316	0.365	0.349	0.380	0.310	0.350	0.295	0.338
Weather	96	0.160	0.214	0.159	0.213	**0.157**	**0.211**	0.198	0.222	0.220	0.217	0.199	**0.211**	-	-	0.243	0.255	0.211	0.243	0.203	0.238	0.194	0.235
	192	0.210	0.260	0.215	0.266	**0.208**	0.256	0.247	0.265	0.271	0.259	0.246	**0.251**	-	-	0.278	0.329	0.263	0.294	0.256	0.290	0.249	0.285
	336	0.274	0.309	0.291	0.322	**0.255**	**0.290**	0.283	0.303	0.286	0.297	0.274	0.291	-	-	0.306	0.346	0.321	0.339	0.314	0.336	0.302	0.327
	720	0.418	0.405	0.415	0.400	0.405	0.397	0.373	0.354	0.373	0.354	**0.337**	**0.340**	-	-	0.350	0.374	0.404	0.397	0.397	0.396	0.372	0.378
	AVG	0.265	0.297	0.270	0.300	**0.256**	0.288	0.275	0.286	0.287	0.281	0.264	**0.273**	-	-	0.294	0.326	0.300	0.318	0.292	0.315	0.279	0.306
Global Temp	96	0.211	0.343	**0.210**	**0.342**	0.214	0.345	0.227	0.354	0.224	0.351	0.224	0.351	0.255	0.375	0.363	0.472	0.234	0.361	0.230	0.355	0.228	0.354
	192	0.257	0.386	0.254	0.385	**0.246**	**0.379**	0.269	0.396	0.266	0.394	0.267	0.395	0.313	0.423	0.387	0.489	0.276	0.400	0.273	0.395	0.276	0.398
	336	0.281	0.405	0.267	**0.395**	**0.266**	0.398	0.292	0.419	0.296	0.420	0.291	0.417	0.362	0.460	0.430	0.517	0.314	0.431	0.324	0.434	0.327	0.437
	720	0.354	0.465	0.289	**0.420**	**0.288**	0.421	0.351	0.437	0.403	0.498	0.387	0.488	0.486	0.545	0.582	0.617	0.418	0.504	0.505	0.542	0.472	0.535
	AVG	0.275	0.400	0.255	**0.385**	**0.253**	**0.385**	0.285	0.409	0.297	0.416	0.292	0.413	0.354	0.451	0.440	0.524	0.311	0.424	0.333	0.431	0.326	0.431
Average		0.336	0.384	0.336	0.380	**0.322**	0.372	0.349	0.377	0.347	**0.370**	0.359	0.373	0.396	0.413	0.461	0.454	0.428	0.420	0.429	0.412	0.416	0.405
1st Count		3		10		**28**		2		11		10		1		4		0		0		1	

All pretrained -me-series models were evaluated without further training for different forecas-ng horizons

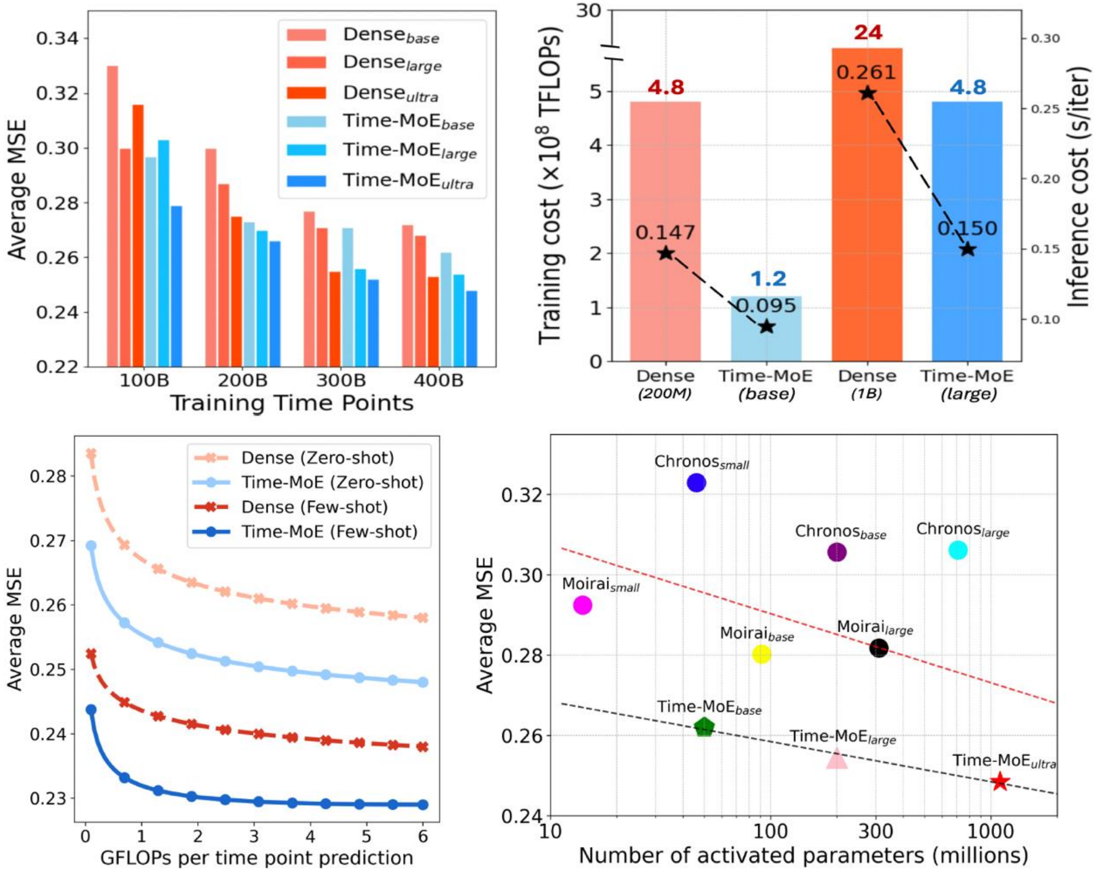
Time-MoE: Full-shot Performance

Table 4: Full results of in-domain forecasting experiments. A lower MSE or MAE indicates a better prediction. Full-shot results besides the Global Temp dataset, the TFT and N-Beats models are obtained from (Liu et al., 2024a). **Red**: the best, Blue: the 2nd best.

Models		TIME-MOE (Ours)						Full-shot Time Series Models																			
		TIME-MOE$_{base}$		TIME-MOE$_{large}$		TIME-MOE$_{ultra}$		iTransformer		TimeMixer		TimesNet		PatchTST		Crossformer		TiDE		DLinear		FEDformer		TFT		N-BEATS	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.345	0.373	0.335	0.371	**0.323**	**0.365**	0.386	0.405	0.375	0.400	0.384	0.402	0.414	0.419	0.423	0.448	0.479	0.464	0.386	0.400	0.376	0.419	0.478	0.476	0.383	0.405
	192	0.372	0.396	0.374	0.400	**0.359**	**0.391**	0.441	0.436	0.436	0.429	0.421	0.429	0.460	0.445	0.471	0.474	0.525	0.492	0.437	0.432	0.420	0.448	0.510	0.486	0.453	0.447
	336	0.389	**0.412**	0.390	**0.412**	**0.388**	0.418	0.487	0.458	0.484	0.458	0.491	0.469	0.501	0.466	0.570	0.546	0.565	0.515	0.481	0.459	0.459	0.465	0.548	0.505	0.517	0.493
	720	0.410	0.443	**0.402**	**0.433**	0.425	0.450	0.503	0.491	0.498	0.482	0.521	0.500	0.500	0.488	0.653	0.621	0.594	0.558	0.519	0.516	0.506	0.507	0.549	0.515	0.594	0.546
	AVG	0.379	0.406	0.375	**0.404**	**0.373**	0.406	0.454	0.447	0.448	0.442	0.454	0.450	0.468	0.454	0.529	0.522	0.540	0.507	0.455	0.451	0.440	0.459	0.521	0.496	0.487	0.473
ETTh2	96	0.276	0.340	0.278	**0.335**	**0.274**	0.338	0.297	0.349	0.289	0.341	0.340	0.374	0.302	0.348	0.745	0.584	0.400	0.440	0.333	0.387	0.358	0.397	0.352	0.387	0.362	0.384
	192	0.331	0.371	0.345	0.373	**0.330**	**0.370**	0.380	0.400	0.372	0.392	0.402	0.414	0.388	0.400	0.877	0.656	0.528	0.509	0.477	0.476	0.429	0.439	0.429	0.432	0.413	0.430
	336	0.373	0.402	0.384	0.402	**0.362**	**0.396**	0.428	0.432	0.386	0.414	0.452	0.541	0.426	0.433	1.043	0.731	0.643	0.571	0.594	0.541	0.496	0.487	0.461	0.460	0.430	0.448
	720	0.404	0.431	0.437	0.437	**0.370**	**0.417**	0.427	0.445	0.412	0.434	0.462	0.657	0.431	0.446	1.104	0.763	0.874	0.679	0.831	0.657	0.463	0.474	0.475	0.473	0.554	0.530
	AVG	0.346	0.386	0.361	0.386	**0.334**	**0.380**	0.383	0.406	0.364	0.395	0.414	0.496	0.386	0.406	0.942	0.683	0.611	0.549	0.558	0.515	0.436	0.449	0.429	0.438	0.440	0.448
ETTm1	96	0.286	0.334	0.264	0.325	**0.256**	**0.323**	0.334	0.368	0.320	0.357	0.338	0.375	0.329	0.367	0.404	0.426	0.364	0.387	0.345	0.372	0.379	0.419	0.468	0.444	0.334	0.372
	192	0.307	0.358	0.295	0.350	**0.281**	**0.343**	0.377	0.391	0.361	0.381	0.374	0.387	0.367	0.385	0.450	0.451	0.398	0.404	0.380	0.389	0.426	0.441	0.557	0.488	0.379	0.401
	336	0.354	0.390	**0.323**	0.376	0.326	**0.374**	0.426	0.420	0.390	0.404	0.410	0.411	0.399	0.410	0.532	0.515	0.428	0.425	0.413	0.413	0.445	0.459	0.682	0.528	0.421	0.425
	720	0.433	0.445	**0.409**	**0.435**	0.454	0.452	0.491	0.459	0.454	0.441	0.478	0.450	0.454	0.439	0.666	0.589	0.487	0.461	0.474	0.453	0.543	0.490	0.722	0.565	0.476	0.471
	AVG	0.345	0.381	**0.322**	**0.371**	0.329	0.373	0.407	0.409	0.381	0.395	0.400	0.405	0.387	0.400	0.513	0.495	0.419	0.419	0.403	0.406	0.448	0.452	0.607	0.506	0.403	0.417
ETTm2	96	0.172	0.265	**0.169**	0.259	0.183	0.273	0.180	0.264	0.175	**0.258**	0.187	0.267	0.175	0.259	0.287	0.366	0.207	0.305	0.193	0.292	0.203	0.287	0.223	0.295	0.208	0.283
	192	0.228	0.306	**0.223**	**0.295**	**0.223**	0.301	0.250	0.309	0.237	0.299	0.249	0.309	0.241	0.302	0.414	0.492	0.290	0.364	0.284	0.362	0.269	0.328	0.281	0.329	0.344	0.372
	336	0.281	0.345	0.293	0.341	**0.278**	**0.339**	0.311	0.348	0.298	0.340	0.321	0.351	0.305	0.343	0.597	0.542	0.377	0.422	0.369	0.427	0.325	0.366	0.364	0.373	0.354	0.383
	720	0.403	0.424	0.451	0.433	0.425	0.424	0.412	0.407	**0.391**	**0.396**	0.408	0.403	0.402	0.400	1.730	1.042	0.558	0.524	0.554	0.522	0.421	0.415	0.475	0.435	0.460	0.455
	AVG	**0.271**	0.335	0.284	0.332	0.277	0.334	0.288	0.332	0.275	**0.323**	0.291	0.332	0.280	0.326	0.757	0.610	0.358	0.403	0.350	0.400	0.304	0.349	0.336	0.358	0.342	0.373
Weather	96	0.151	0.203	**0.149**	**0.201**	0.154	0.208	0.174	0.214	0.163	0.209	0.172	0.220	0.177	0.218	0.158	0.230	0.202	0.261	0.196	0.255	0.217	0.296	0.186	0.231	0.165	0.224
	192	0.195	0.246	**0.192**	**0.244**	0.202	0.251	0.221	0.254	0.208	0.250	0.219	0.261	0.225	0.259	0.206	0.277	0.242	0.298	0.237	0.296	0.276	0.336	0.240	0.275	0.209	0.269
	336	0.247	0.288	**0.245**	**0.285**	0.252	0.287	0.278	0.296	0.251	0.287	0.280	0.306	0.278	0.297	0.272	0.335	0.287	0.335	0.283	0.335	0.339	0.380	0.302	0.317	0.261	0.310
	720	0.352	0.366	0.352	0.365	0.392	0.376	0.358	0.349	0.339	**0.341**	0.365	0.359	0.354	0.348	0.398	0.418	0.351	0.386	0.345	0.381	0.403	0.428	0.388	0.369	**0.336**	0.362
	AVG	0.236	0.275	**0.234**	0.273	0.250	0.280	0.257	0.278	0.240	**0.271**	0.259	0.286	0.258	0.280	0.258	0.315	0.270	0.320	0.265	0.316	0.308	0.360	0.279	0.298	0.243	0.291
Global Temp	96	0.192	0.328	0.192	0.329	**0.189**	**0.322**	0.223	0.351	0.215	0.346	0.250	0.381	0.219	0.349	0.272	0.406	0.223	0.352	0.221	0.354	0.261	0.392	0.260	0.390	0.210	0.344
	192	0.238	**0.375**	0.236	**0.375**	**0.234**	0.376	0.282	0.404	0.266	0.393	0.298	0.418	0.269	0.395	0.305	0.435	0.278	0.401	0.257	0.388	0.299	0.423	0.301	0.423	0.253	0.385
	336	0.259	**0.397**	0.256	**0.397**	**0.253**	0.399	0.313	0.431	0.313	0.430	0.315	0.434	0.319	0.435	0.352	0.468	0.330	0.440	0.294	0.418	0.341	0.454	0.359	0.464	0.282	0.411
	720	0.345	0.465	0.322	0.451	**0.292**	**0.426**	0.393	0.488	0.468	0.536	0.407	0.497	0.452	0.526	0.508	0.562	0.485	0.544	0.380	0.479	0.359	0.469	0.371	0.477	0.342	0.457
	AVG	0.258	0.391	0.251	0.388	**0.242**	**0.380**	0.303	0.419	0.316	0.426	0.318	0.433	0.315	0.426	0.359	0.468	0.329	0.434	0.288	0.410	0.315	0.435	0.323	0.439	0.272	0.399
Average		0.306	0.362	0.304	0.359	**0.301**	**0.358**	0.349	0.382	0.337	0.375	0.356	0.400	0.349	0.382	0.560	0.516	0.421	0.439	0.387	0.416	0.375	0.417	0.416	0.422	0.364	0.400
1st Count		4		21		**33**		0		6		0		0		0		0		0		0		0		1	

Time-MoE is trained on all datasets simultaneously for only one epoch, while others are trained separately on each dataset

Time-MoE: Scalability Analysis

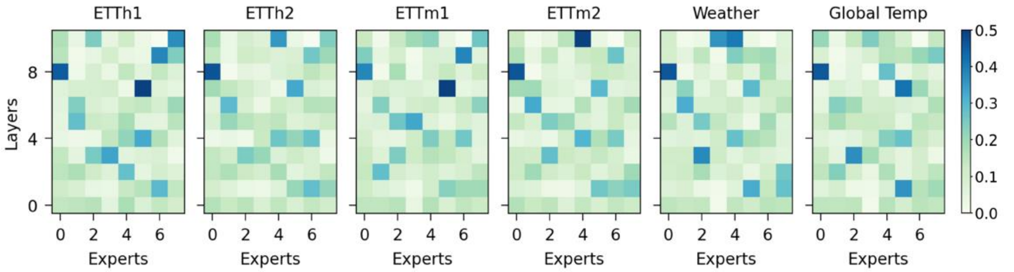


Time-MoE: Ablations & Visualizations

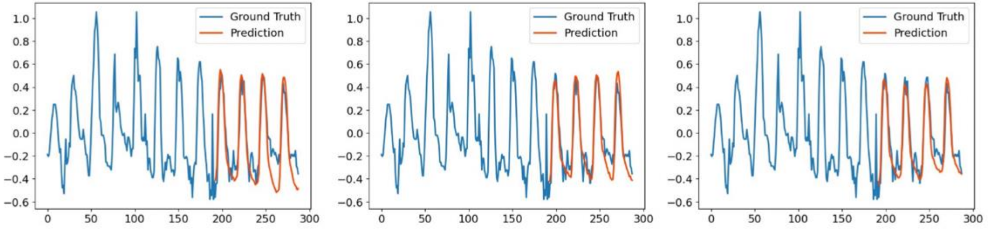
Ablation studies of different model components

	Average MSE
TIME-MOE$_{base}$	**0.262**
w/o Huber loss	0.267
w/o multi-resolution layer	0.269
w/o mixture-of-experts	0.272
w/o auxiliary loss	0.275

Gating score of each expert across all layers in six datasets



Zero-shot forecasting cases from Global Temp by different models

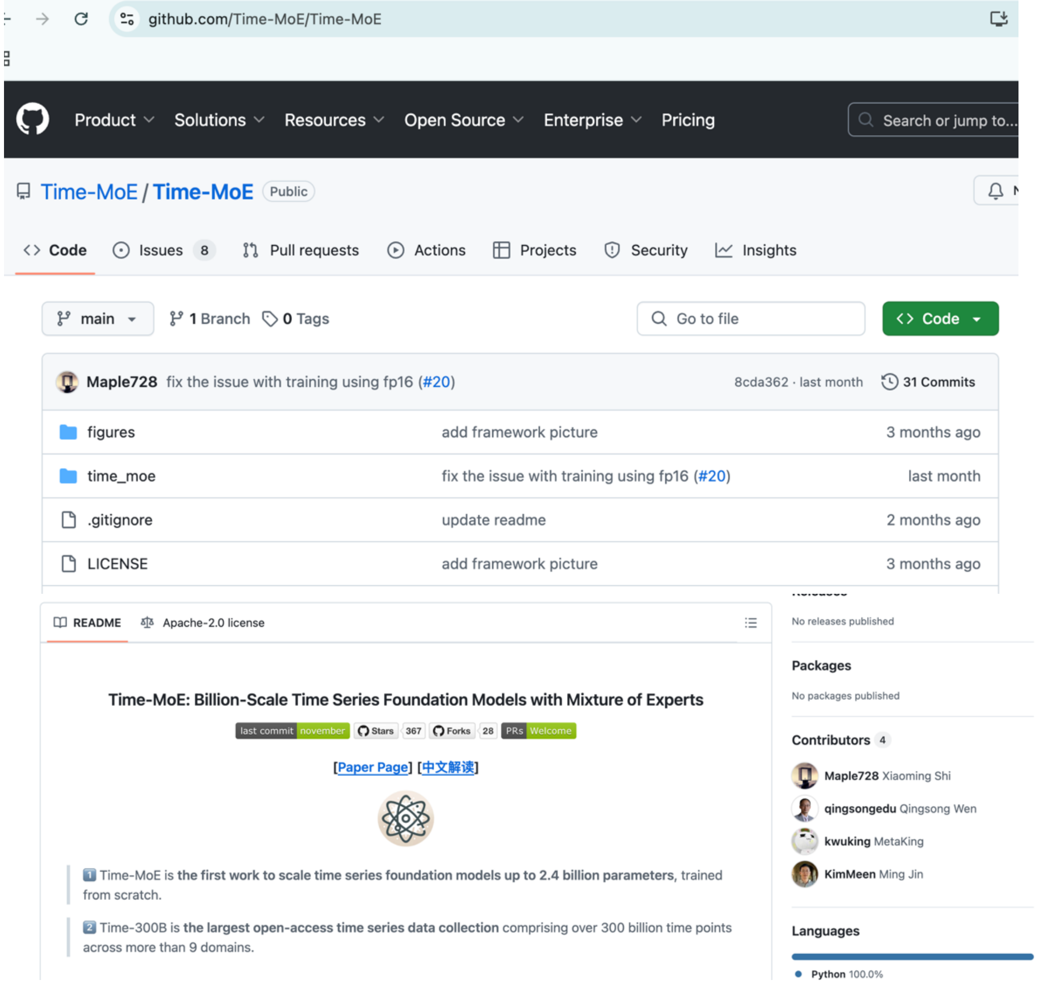


(a) Time–MoE$_{base}$ (b) Time–MoE$_{large}$ (c) Time–MoE$_{ultra}$

Performance of different MoE top-k setups

TIME-MOE$_{base}$	Average MSE	Inference Speed
w/ {Top$_1$}	0.264	0.082 s/iter
w/ {Top$_2$}	**0.262**	**0.095 s/iter**
w/ {Top$_4$}	0.262	0.109 s/iter
w/ {Top$_6$}	0.265	0.120 s/iter
w/ {Top$_8$}	0.269	0.129 s/iter

Open-source:

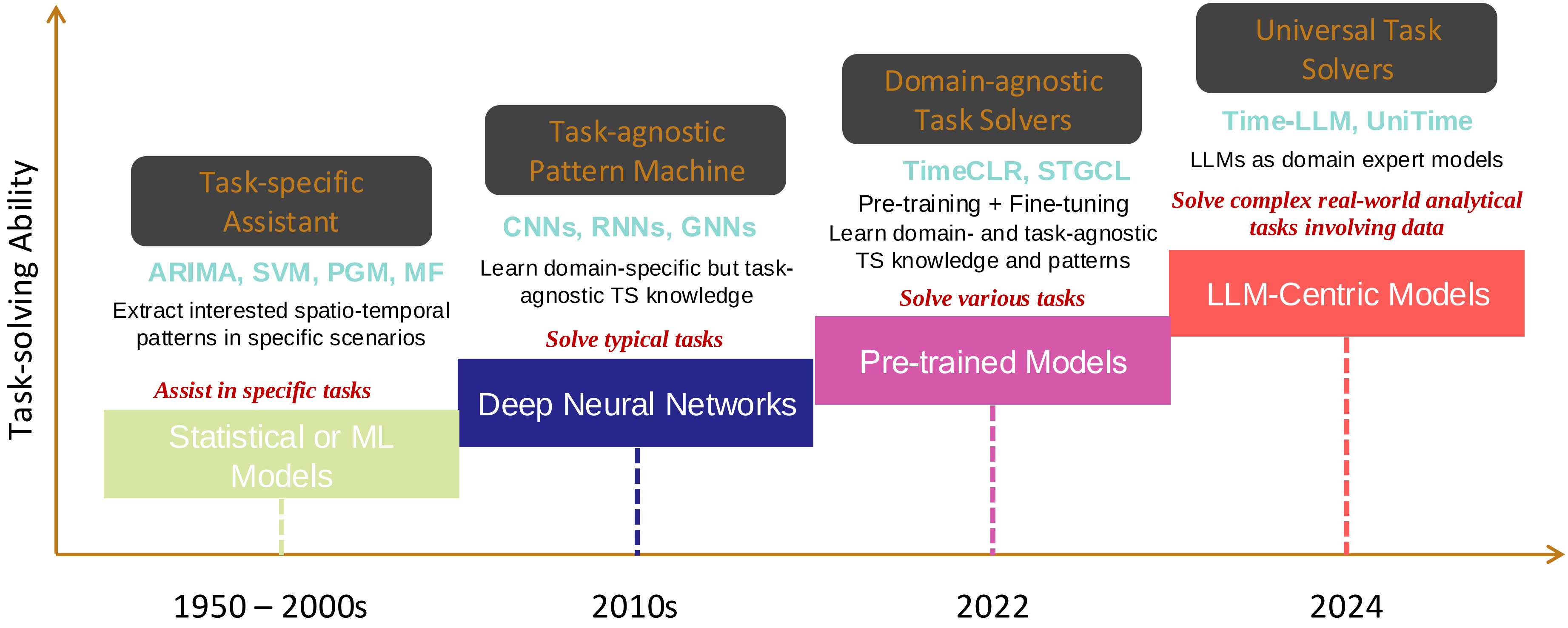

github.com/Time-MoE/Time-MoE
Product
Solutions
Resources
Open Source
Enterprise
Pricing
Search or jump to...
Time-MoE / Time-MoE
Public
Code
Issues 8
Pull requests
Actions
Projects
Security
Insights
main
1 Branch
0 Tags
Go to file
Code
Maple728 fix the issue with training using fp16 (#20)
8cda362 · last month
31 Commits
figures
add framework picture
3 months ago
time_moe
fix the issue with training using fp16 (#20)
last month
.gitignore
update readme
2 months ago
LICENSE
add framework picture
3 months ago
README
Apache-2.0 license
Time-MoE: Billion-Scale Time Series Foundation Models with Mixture of Experts
last commit november
Stars 367
Forks 28
PRs Welcome
[Paper Page] [中文解读]
1 Time-MoE is the first work to scale time series foundation models up to 2.4 billion parameters, trained from scratch.
2 Time-300B is the largest open-access time series data collection comprising over 300 billion time points across more than 9 domains.
No releases published
Packages
No packages published
Contributors 4
Maple728 Xiaoming Shi
qingsongedu Qingsong Wen
kwuking MetaKing
KimMeen Ming Jin
Languages
Python 100.0%

Outline

- ❑ **Introduction**
- ❑ **Large Language Model for Time Series Analysis**
- ❑ **Foundation Model for Time Series Analysis**
- ➢ ***Future Directions***

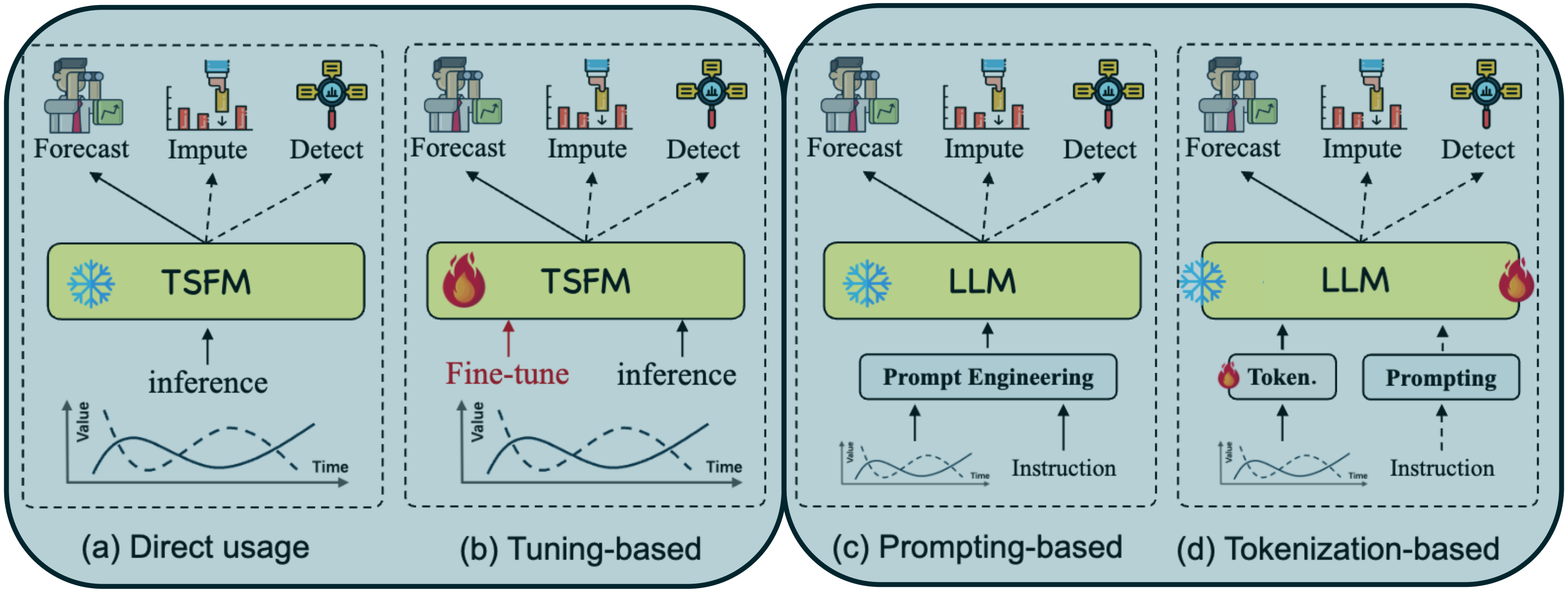
- Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang*, Bin Yang, Jindong Wang, Shirui Pan*, Qingsong Wen*, **"Position: What Can Large Language Models Tell Us about Time Series Analysis", ICML 2024.**

Roadmap



Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang*, Bin Yang, Jindong Wang, Shirui Pan*, Qingsong Wen*, "Position: What Can Large Language Models Tell Us about Time Series Analysis", **ICML 2024**

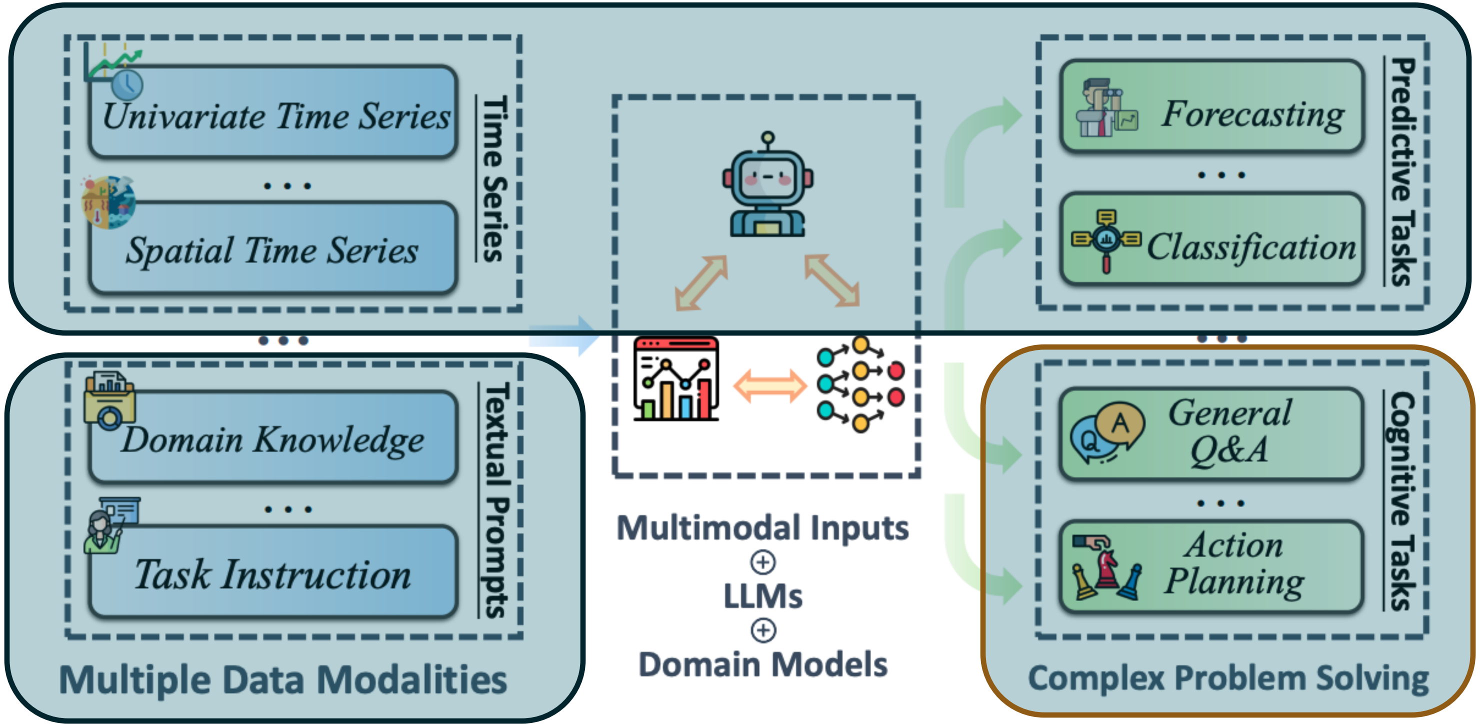
- Current: **different techniques** when developing/reprograming **TSFMs/LLMs** for time series data



TSFMs: Time-MoE, Moirai, TimesFM, Chronos, Moment,…

LLMs for TS: Time-LLM, LLMTime, AutoTimes, OFA, UniST…

- Future: integration of **time series (models) with LLM/Agent** brings potential for complex real-world problems



What we have some solutions: **Time-LLM**

What we have not solved yet

Other Directions When Combining TS with LLM

- Hallucination and Alignment
- Privacy and Security
- Benchmarking and Evaluation
- Multi-agent Systems
- Multi-modality Systems
- … …

More on LLM/FM for Time Series

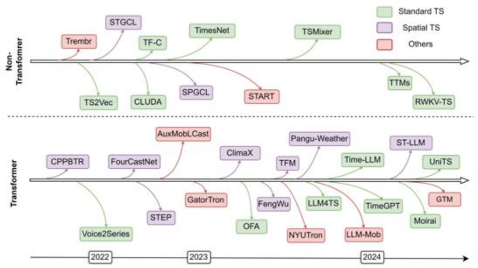


Figure 1: Roadmaps of representative TSFMs.

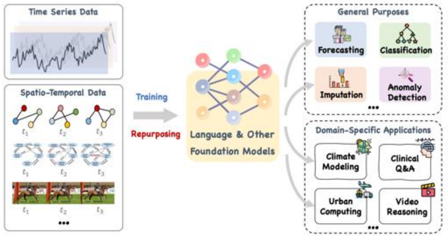


Fig. 1: Large models (i.e., language and other related foundation models) can be either trained or adeptly repurposed to handle time series and spatio-temporal data for a range of general-purpose tasks and specialized domain applications.

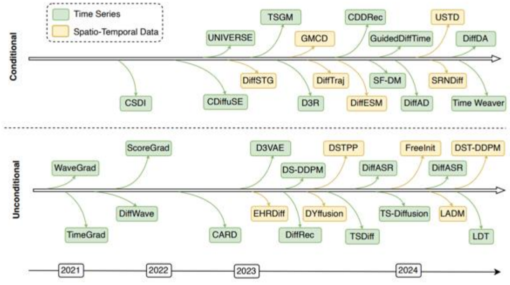
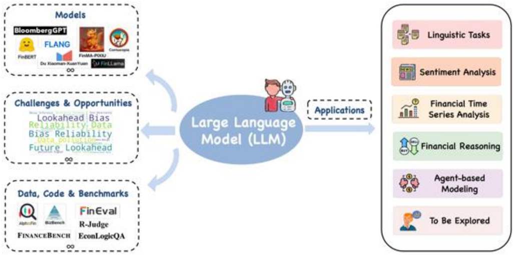


Fig. 3: Representative diffusion models for time series and spatio-temporal data in recent years.



Foundation Models for Time Series Analysis: A Tutorial and Survey

Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, Qingsong Wen

Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook

Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Fellow, IEEE, Yu Zheng, Fellow, IEEE, Lei Chen, Fellow, IEEE, Hui Xiong, Fellow, IEEE

A Survey on Diffusion Models for Time Series and Spatio-Temporal Data

Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Jiang Bian, Shirui Pan, Qingsong Wen

A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges

Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, Stefan Zohren

Surveys:

[1] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, Qingsong Wen*,**"Foundation Models for Time Series Analysis: A Tutorial and Survey", KDD 2024**.

[2] Ming Jin, Qingsong Wen*, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, Shirui Pan, Vincent S. Tseng, Yu Zheng, Lei Chen, Hui Xiong, **"Large Models for Time Series and Spatio-Temporal Data: A Survey and Outlook", arXiv 2023**.

[3] Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M. Mulvey, H. Vincent Poor, Qingsong Wen, Stefan Zohren, **"A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges", arXiv 2024**.

[4] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Jiang Bian, Shirui Pan, Qingsong Wen*, **"A Survey on Diffusion Models for Time Series and Spatio-Temporal Data", arXiv 2024**.

Algorithm & Dataset:

[4] Qingxiang Liu, Xu Liu, Chenghao Liu, Qingsong Wen, Yuxuan Liang, **"Time-FFM: Towards LM-Empowered Federated Foundation Model for Time Series Forecasting," NeurIPS 2024**.

[5] Haoxin Liu, Shangqing Xu, Zhiyuan Zhao, Lingkai Kong, Harshavardhan Kamarthi, Aditya B. Sasanur, Megha Sharma, Jiaming Cui, Qingsong Wen, Chao Zhang, B. Aditya Prakash, **"Time-MMD: A New Multi-Domain Multimodal Dataset for Time Series Analysis," NeurIPS 2024**.

Thanks!
Q&A

Qingsong Wen
https://sites.google.com/site/qingsongwen8/

Credits to my collaborators from: