

“时间序列，也能说话”： ChatTime赋能AIOps的多模态基础模型探索

ChatTime: A Unified Multimodal Time Series Foundation Model Bridging Numerical and Textual Data -- AAI 2025 Oral

王敬宇 北京邮电大学 网络与交换全国重点实验室

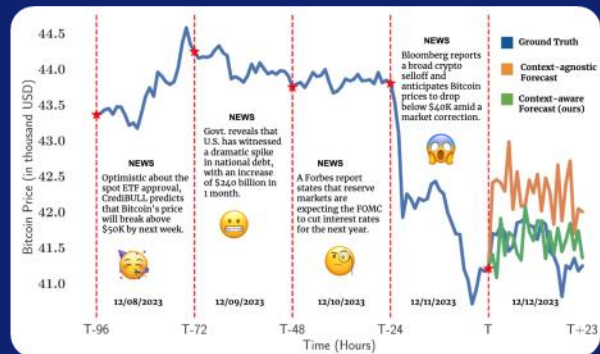
主办单位：中国计算机学会（CCF）

承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学

协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云

研究背景：多模态推理

- 时间序列数据已成为金融、交通、能源，当然也包括AIOps领域的核心资产。
- **单一数值曲线**早已难以支撑日益复杂的预测与诊断需求，更多的要理解曲线背后的文字：日志、告警、工单、邮件。也就是说，这本质上是一个 **“数值+文本”多模态推理任务**。



经济学家综合历史股价序列与宏观经济政策报告（**结合文本信息**）预测未来走势

AIOps典型场景：本质上是 **“数值+文本”** 的多模态推理



传统方法

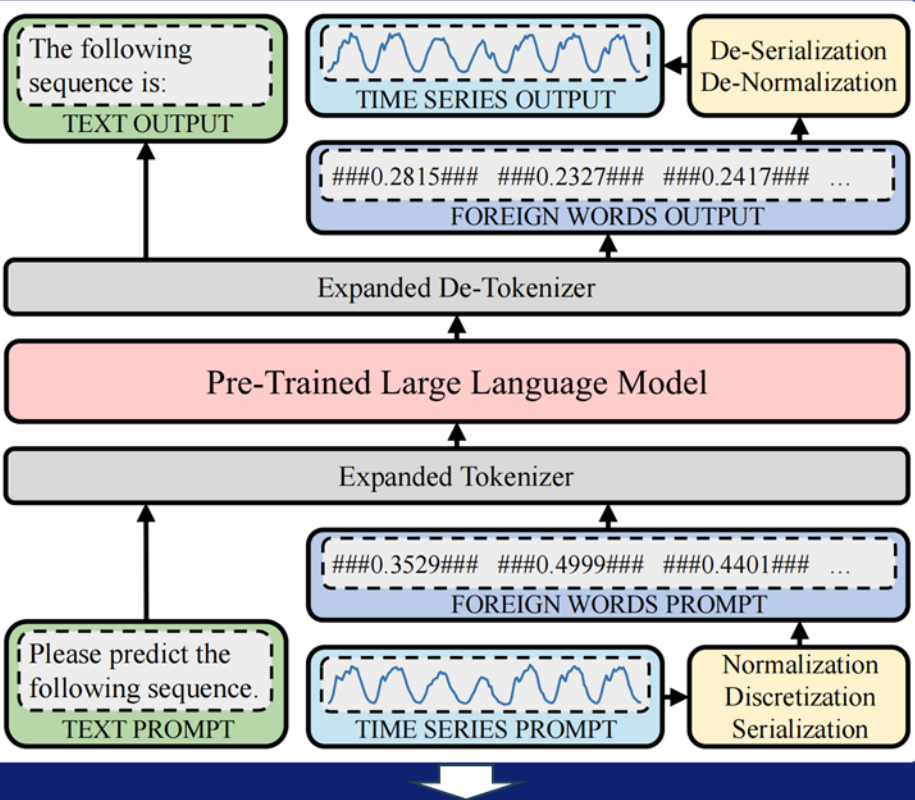
代表方法：Autoformer

- 1.模型泛化弱：传统模型只能在特定数据集上工作，迁移能力差。
- 2.模态单一：主流时序模型无法处理文本。
- 3.无法生成文本：许多模型只能输出数值，无法生成分析报告或问答。

功能点/路线	①时序数值文本化后输入 LLM 代表方法：LLM TIME	②LLM嵌入时序模型 代表方法：Time-LLM	③从零训练时序基础模型 代表方法：Chronos
📄 支持文本-时序双模态输入	✅ 可直接拼接文本与映射序列	⚠️ 需设计融合模块（加适配层）	❌ 主要为数值模态
📄 文本到时间序列的生成能力	✅ 可用Prompt生成离散token序列	✅ 支持时间序列的生成	❌ 无文本理解能力
🔄 时间序列到文本的生成能力	✅ 基于LLM自然支持	❌ 需额外任务解码器	❌ 不支持自然语言生成
🔄 零样本泛化能力（Zero-shot）	⚠️ 文本类任务强，时序弱	❌ 需场景微调，缺乏泛化能力	❌ 通常为单任务预训练
💡 语义推理能力（自然语言）	✅ LLM原生能力保留	⚠️ 被弱化	❌ 完全缺失
💡 时序建模能力（趋势/周期性）	❌ 缺乏依赖理解，需Patch/增强补偿	✅ 可设计融合模块引入时序依赖	✅ 完全基于时序优化结构
🔧 工程部署/计算成本	⚠️ Token映射，利用率不足	⚠️ 中等，需训练融合模块	❌ 成本极高（训练+数据）
🔄 兼容预训练权重利用	✅ 完全兼容（如Llama、GPT）	⚠️ 局部兼容（部分冻结+插入层）	❌ 不使用预训练权重
🔧 多任务适应能力（预测+QA）	✅ 优	⚠️ 多需手工构造适配头	❌ 需从头为每任务建模

能否高效构建出一个统一时序文本生成与理解的多模态基础模型呢？

模型架构： ChatTime三步建模流程



ChatTime 与 LLMTIME 令牌消耗对比

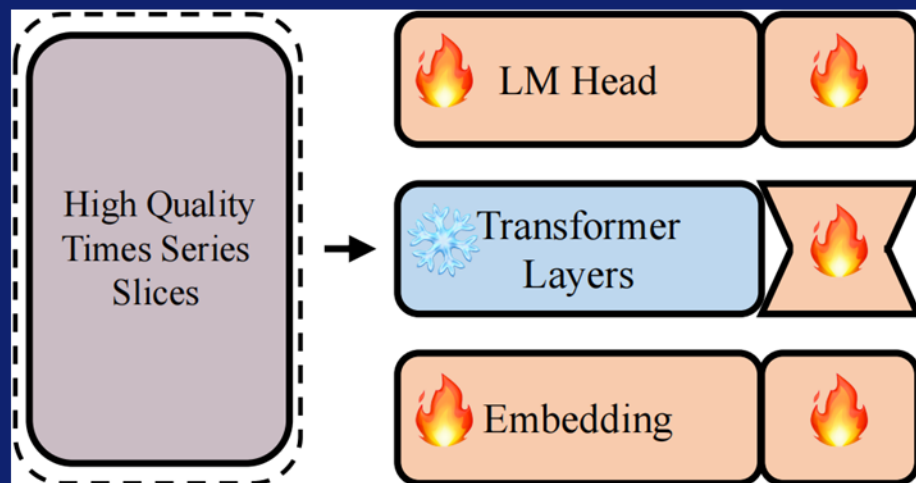
Time Series [0.2835, 0.2285, 0.1587, 0.4001]
GPT (34 tokens) "2 8 3 5, 2 2 8 5, 1 5 8 7, 4 0 0 1" ['2', '-', '8', '-', '3', '-', '5', '-', '-', '2', '-', '2', -, '8', '-', '5', '-', '-', '1', '-', '5', '-', '8', -, '7', '-', '-', '4', '-', '0', '-', '0', '-', '1']
LLaMA (22 tokens) "2835, 2285, 1587, 4001" ['2', '8', '3', '5', ',', '-', '2', '2', '8', '5', ',', -, '1', '5', '8', '7', ',', '-', '4', '0', '0', '1']
ChatTime (7 tokens) "###0.2835### ###0.2285### ###0.1587### ###0.4001###" ['###0.2835###', '-', '###0.2285###', '-', '###0.1587###', '-', '###0.4001###']

- 将时间序列数值建模为“外语”，**提高令牌的利用率**
- 参照扩充词表进行微调，**无需修改架构或从零训练**

- **Min-Max缩放**：将任意范围内的时间值映射至 $[-1, 1]$ ；
- **离散分箱**：把连续值划分成10000个区间，获得“词”；
- **添加特殊符号**：构造类似“语言单词”的 token，极大提升token效率。

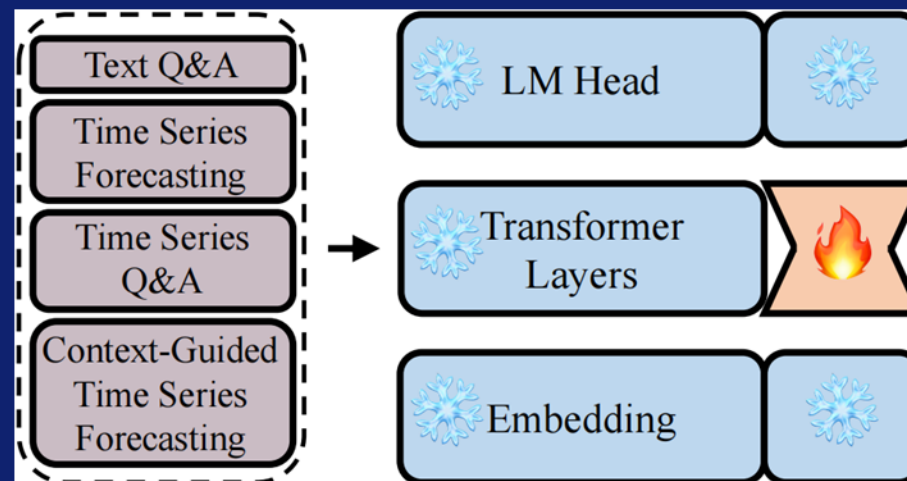
模型训练：两阶段训练策略

持续预训练（增强“看得懂”）



- ✓ 构建高质量持续预训练数据集：
 - 源自 Monash 和 TFB 两大开源仓库的时序切片，覆盖约 100 个子数据集
 - 对 10M 原始切片进行聚类抽样，以优化资源利用并过滤重复数据
- ✓ Llama2-7b采用 LoRA 微调Transformer层、Embedding 层以及 Output 头

指令精调（增强“看得懂” + “做得好”）



- ✓ 构建4类任务多模态精调数据集：
 - 文本问答：保留语言能力
 - 零样本预测：时序推理能力
 - 文本辅助预测：文本→时序
 - 时间序列问答：时序→文本
- ✓ 采用 LoRA 微调 Transformer 层，Embedding 层以及 Output 头固定

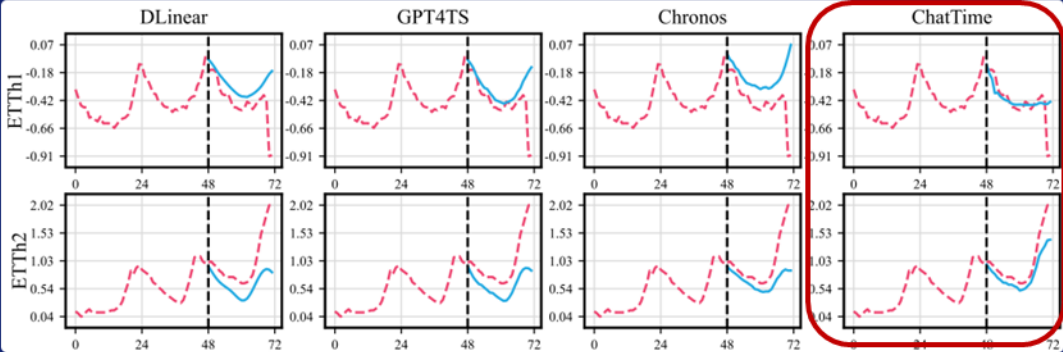
关键实验成果1：零样本预测

Dataset	Hist	Pred	Full-Shot Forecast				Zero-Shot Forecast				
			DLinear	iTransformer	GPT4TS	TimeLLM	TimeGPT	Moirai	TimesFM	Chronos	ChatTime
ETTh1	48	24	0.1462	0.1650	0.1389	0.1467	0.1604	0.1694	0.2021	0.1634	0.1698
	72	24	0.1358	0.1852	0.1469	0.1439	0.1603	0.1796	0.1599	0.1372	0.1403
	96	24	0.1398	0.1964	0.1447	0.1473	0.1577	0.1433	0.1454	0.1374	0.1374
	120	24	0.1371	0.1971	0.1414	0.1513	0.1594	0.1492	0.1502	0.1348	0.1431
ETTh2	48	24	0.2724	0.2937	0.2742	0.2758	0.2874	0.2963	0.3360	0.3128	0.2906
	72	24	0.2756	0.3118	0.2717	0.2972	0.2888	0.3109	0.2880	0.3045	0.3092
	96	24	0.2831	0.3417	0.2900	0.2864	0.2902	0.3139	0.3144	0.3158	0.2917
	120	24	0.2863	0.3299	0.2854	0.3175	0.3026	0.2905	0.3311	0.3150	0.3124
ETTm1	192	96	0.1479	0.1608	0.1384	0.1503	0.1921	0.1608	0.1719	0.1604	0.1442
	288	96	0.1400	0.1813	0.1345	0.1425	0.1715	0.1848	0.1650	0.1452	0.1587
	384	96	0.1428	0.1680	0.1518	0.1452	0.1616	0.1619	0.1584	0.1463	0.1393
	480	96	0.1406	0.2001	0.1472	0.1527	0.1570	0.1703	0.1582	0.1401	0.1802
ETTm2	192	96	0.2793	0.3397	0.2792	0.2918	0.4294	0.4206	0.3405	0.3759	0.3135
	288	96	0.2881	0.3623	0.2918	0.2904	0.3625	0.3882	0.3277	0.3472	0.3340
	384	96	0.2947	0.2880	0.3089	0.3003	0.3389	0.3742	0.3562	0.3589	0.3434
	480	96	0.3014	0.3725	0.2945	0.3054	0.3242	0.3597	0.3679	0.3353	0.4213
Electric	48	24	0.5719	0.5951	0.5008	0.5733	0.5276	0.6617	0.6005	0.6098	0.6083
	72	24	0.5486	0.5619	0.4896	0.4989	0.4953	0.6018	0.5454	0.5914	0.6238
	96	24	0.5536	0.5290	0.4432	0.4816	0.4971	0.5260	0.5276	0.5139	0.4951
	120	24	0.4714	0.5622	0.4540	0.4848	0.5196	0.4963	0.4900	0.5031	0.5101
Exchange	14	7	0.0543	0.0526	0.0533	0.0531	0.0620	0.0784	0.0647	0.0555	0.0540
	21	7	0.0571	0.0547	0.0505	0.0505	0.0599	0.0812	0.0743	0.0635	0.0556
	28	7	0.0595	0.0581	0.0508	0.0511	0.0610	0.0844	0.0652	0.0595	0.0559
	35	7	0.0615	0.0607	0.0493	0.0524	0.0629	0.0677	0.0632	0.0598	0.0558
Traffic	48	24	0.4662	0.5000	0.4557	0.4473	0.4668	0.4887	0.4483	0.4718	0.4220
	72	24	0.4475	0.4443	0.4116	0.4252	0.4635	0.4581	0.4196	0.3725	0.3873
	96	24	0.4438	0.4348	0.4190	0.4064	0.4332	0.4082	0.3714	0.3787	0.4074
	120	24	0.4190	0.4149	0.3416	0.4279	0.4161	0.3539	0.3542	0.3908	0.4125
Weather	288	144	0.0339	0.0367	0.0364	0.0352	0.0331	0.0305	0.0354	0.0343	0.0352
	432	144	0.0366	0.0404	0.0401	0.0395	0.0321	0.0302	0.0298	0.0346	0.0356
	576	144	0.0364	0.0379	0.0399	0.0377	0.0328	0.0331	0.0321	0.0349	0.0284
	720	144	0.0371	0.0395	0.0392	0.0392	0.0323	0.0353	0.0369	0.0335	0.0332
Avg. MAE			0.2409	0.2661	0.2286	0.2390	0.2544	0.2659	0.2541	0.2512	0.2515
Avg. Rank			3.7500	6.9688	3.0000	3.9688	5.5625	6.5000	5.7500	4.8438	4.4688

单模态时间序列 zero-shot 预测精度对比

Method	Zero-Shot Forecast	Missing Support	Training Token	Trainable Parameter
TimesFM	✓	✗	3T	200M
Moirai	✓	✓	150B	300M
TimeGPT	✓	✓	100B	Unknown
MOMENT	✗	✓	100B	300M
Timer	✗	✗	50B	50M
Chronos	✓	✓	25B	700M
ChatTime	✓	✓	1B	350M

时间序列基础模型特性对比



单模态时间序列 zero-shot 预测可视化对比

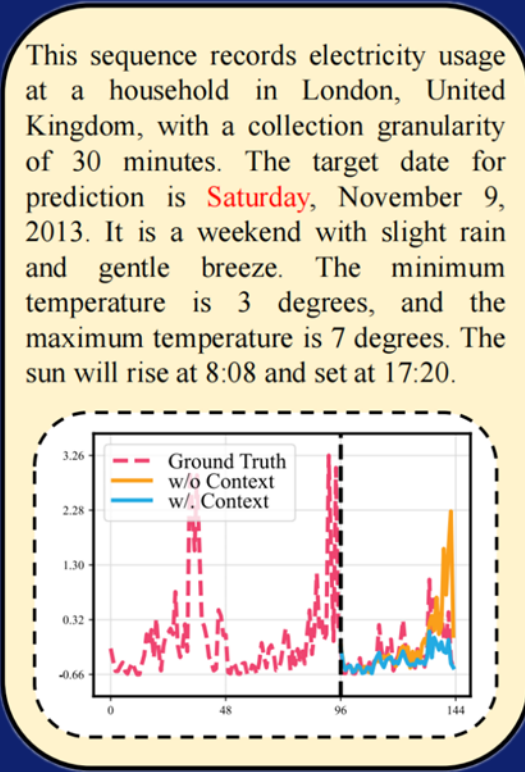
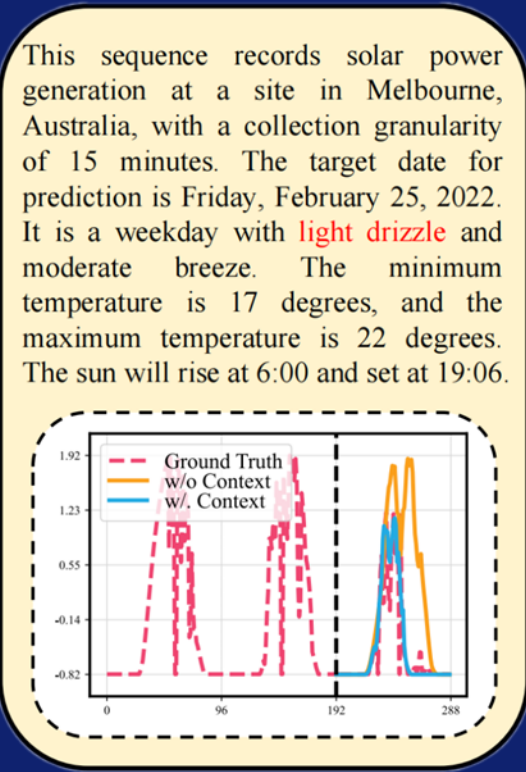
仅需 **4%** 训练数据量，即达先前 SOTA 时间序列基础模型 **99.9%** 的 zero-shot 预测精度

关键实验成果2：文本辅助预测

Dataset	Hist	Pred	Dataset-Specific Forecast				Dataset-Shared Forecast				
			DLinear	GPT4TS	TimeLLM	TGForecaster	Moirai	TimesFM	Chronos	ChatTime-	ChatTime
MSPG	192	96	0.7136	0.7558	0.7697	0.7595	0.8108	0.8362	0.7427	0.7606	0.7346
	288	96	0.7083	0.7464	0.7959	0.7610	0.7849	0.7896	0.7408	0.7606	0.7353
	384	96	0.7014	0.7388	0.7672	0.7638	0.7749	0.7811	0.7352	0.7607	0.7330
	480	96	0.7018	0.7311	0.7632	0.7695	0.7664	0.7667	0.7344	0.7607	0.7292
LEU	96	48	0.6676	0.6697	0.6531	0.6181	0.6228	0.6670	0.6571	0.6496	0.6305
	144	48	0.6495	0.6567	0.6474	0.6355	0.6085	0.6475	0.6597	0.6506	0.6231
	192	48	0.6407	0.6771	0.6329	0.6458	0.6008	0.6490	0.6645	0.6407	0.6111
	240	48	0.6316	0.6383	0.6356	0.6329	0.5968	0.6333	0.6631	0.6377	0.6085
PTF	48	24	0.5204	0.4373	0.4211	0.4411	0.5981	0.4851	0.4813	0.5155	0.4849
	72	24	0.5075	0.4253	0.4031	0.3943	0.5776	0.4258	0.4276	0.4436	0.4307
	96	24	0.4965	0.3921	0.4392	0.3653	0.5179	0.4054	0.4336	0.4172	0.3920
	120	24	0.4796	0.3713	0.3594	0.3594	0.5245	0.3807	0.3902	0.3943	0.3480
Avg. MAE			0.6182	0.6033	0.6073	0.5955	0.6487	0.6223	0.6109	0.6160	0.5884
Avg. Rank			4.7500	5.0833	4.9167	3.9167	5.9167	6.4167	5.5000	5.7500	2.5833

多模态时间序列文本辅助预测精度对比

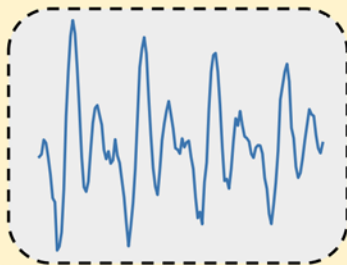
添加文本辅助信息后，预测精度相较于仅使用单模态数值输入时提升 **4.5%**，并优于在每个场景下单独训练的 SOTA 基线 **1.2%**



多模态时间序列文本辅助预测可视化对比

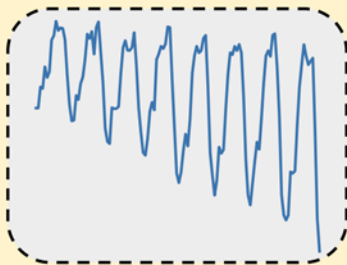
关键实验成果3：时间序列问答

Which option best describes the time series: constant trend, upward trend, or downward trend?



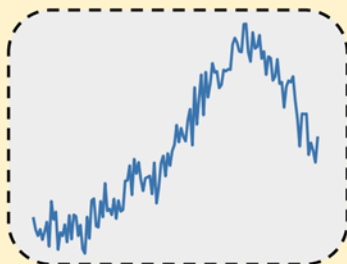
GPT4: upward trend
GPT3.5: constant trend
GLM4: constant trend
Llama3: upward trend
ChatTime: constant trend

Which option best describes the time series: constant volatility, increased volatility, or decreased volatility?



GPT4: increased volatility
GPT3.5: decreased volatility
GLM4: decreased volatility
Llama3: increased volatility
ChatTime: increased volatility

Which option best describes the time series: no seasonality, fixed seasonality, or shifting seasonality?



GPT4: no seasonality
GPT3.5: no seasonality
GLM4: shifting seasonality
Llama3: shifting seasonality
ChatTime: no seasonality

Which option best describes the time series: no outlier, sudden spike, or level shift?



GPT4: level shift
GPT3.5: sudden spike
GLM4: sudden spike
Llama3: level shift
ChatTime: level shift

Feat	Len	GPT4	GPT3.5	GLM4	LLaMA3	ChatTime
Trend	64	0.6532	0.3507	0.7319	0.6799	0.9011
	128	0.7015	0.5846	0.7574	0.5855	0.9068
	256	0.7482	0.5028	0.6377	0.6143	0.8843
	512	0.6346	0.5903	0.6697	0.6753	0.8234
Volatility	64	0.5585	0.5633	0.6797	0.6373	0.7874
	128	0.4979	0.3839	0.4770	0.4756	0.6954
	256	0.4624	0.4894	0.5418	0.5246	0.6228
	512	0.3169	0.3796	0.4549	0.5261	0.5736
Season	64	0.3518	0.3428	0.3366	0.3484	0.6639
	128	0.3515	0.3952	0.3464	0.3958	0.6517
	256	0.5283	0.5089	0.3892	0.4120	0.6463
	512	0.4457	0.4889	0.3892	0.4127	0.6244
Outlier	64	0.7230	0.4325	0.5359	0.7051	0.8773
	128	0.6327	0.5940	0.5298	0.5694	0.9032
	256	0.6795	0.4579	0.5019	0.5073	0.8593
	512	0.6219	0.4996	0.2822	0.4085	0.7478
Avg. Acc		0.5567	0.4728	0.5163	0.5299	0.7605
Avg. Rank		3.0625	4.0625	3.6250	3.2500	1.0000

多模态时间序列问答精度对比

在理解时间序列典型特征
方面表现卓越，优于通用
SOTA 基线 36.6%

多模态时间序列问答可视化对比

- **统一范式：**首次实现时序-文本双向建模
- **高效训练：**利用预训练LLM，单张RTX 4090，无需从零开始
- **开放生态：**创建了时序多模态数据集，填补了该领域空白；所有代码、数据和权重均已完整开源，模型下载量达 **16.5K**，为未来研究提供了宝贵资源，欢迎大家共建、共创！
<https://github.com/ForestsKing/ChatTime>

ChatTime让模型真正 “读懂数据，表达结果”

ChatTime Models

ChatTime: A Unified Multimodal Time Series Foundation Model Brid...

ChengsenWang/ChatTime-1-7B-Base

Time Series Forecasting • Updated Dec 17, 2024 • 104 • 4

ChengsenWang/ChatTime-1-7B-Chat

Time Series Forecasting • Updated Dec 17, 2024 • 16.5k • 2

ChatTime Datasets

ChatTime: A Unified Multimodal Time Series Foundation Model Brid...

ChengsenWang/ChatTime-1-Pretrain-1M

Viewer • Updated Dec 17, 2024 • 1 • 152

ChengsenWang/ChatTime-1-Finetune-100K

Viewer • Updated Dec 17, 2024 • 1 • 100

ChatTime Public

PyTorch implementation of "ChatTime: A Unified Multimodal Time Series Foundation Model Bridging Numerical and Textual Data" (AAAI 2025 [oral])

Jupyter Notebook • 84 • 13

THANKS



感谢王程森博士、戚琦教授、孙海峰、吴金明等作者

联系方式：王敬宇 wangjingyu@bupt.edu.cn

主办单位：中国计算机学会（CCF）

承办单位：中国计算机学会互联网专委会、中国科学院计算机网络信息中心、中国移动研究院、清华大学

协办单位：华为2012实验室、阿里云、中兴通讯、中国移动九天团队、南开大学、西安电子科技大学、清华大学计算机科学与技术系、神州灵云

**ChatTime: A Unified Multimodal Time Series Foundation Model Bridging
Numerical and Textual Data**

Chengsen Wang^{1*} Qi Qi^{1*} Jingyu Wang^{1,2†}
Haifeng Sun¹ Zirui Zhuang¹ Jinming Wu¹ Lei Zhang³ Jianxin Liao¹

¹Beijing University of Posts and Telecommunications, Beijing, China

²Pengcheng Laboratory, Shenzhen, China

³China Unicom Network Communications Corporation Limited, Beijing, China

{cswang, qiqi8266, wangjingyu}@bupt.edu.cn