RHYTHM: Reasoning with Hierarchical Temporal Tokenization for Human Mobility

Haoyu He[†], Haozheng Luo[‡], Yan Chen[‡], Qi R. Wang[†]

[†]Northeastern University, [‡]Northwestern University

Introduction

Background

What is Human Mobility?

Human mobility describes how people move and interact across geographic spaces.

Are Human Movements Random?

No. 93% of trajectories are predictable with underlying periodic patterns.

Key Challenges:

- Mobility data are often **sparse, noisy, and incomplete**.
- Human movement exhibits **strong spatial heterogeneity** and **long-term temporal dependencies**.

From Challenges to Opportunity

Traditional Methods

Cannot capture long-term patterns

Current LLM Approaches

Ignore mobility-specific structures

Our Insight

Combine temporal structure with LLM reasoning

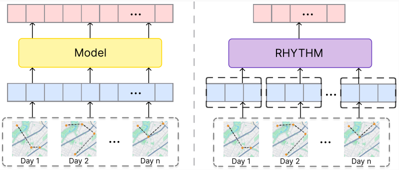


Figure 1: **Motivation for RHYTHM.** Instead of processing entire trajectories as a continuous sequence, RHYTHM segments trajectories into tokens to better capture periodic patterns.

Method

Problem Definition

Given:

- Historical trajectory: $\mathcal{X} = \{x_1, x_2, \ldots, x_T\}$
  where $x_i = (t_i, l_i)$
  - t_i: timestamp
  - $l_i \in \mathcal{L}$: location from finite set $\mathcal{L}$
- Future timestamps $\mathcal{T} = \{t_{T+1}, t_{T+2}, \ldots, t_{T+H}\}$
  - H: prediction horizon

Objective: Predict future locations $\mathcal{Y} = \{l_{T+1}, l_{T+2}, \ldots, l_{T+H}\}$

Goal: Learn mapping $f : (\mathcal{X}, \mathcal{T}) \mapsto \mathcal{Y}$

Structure

RHYTHM (Reasoning with Hierarchical Temporal Tokenization for Human Mobility

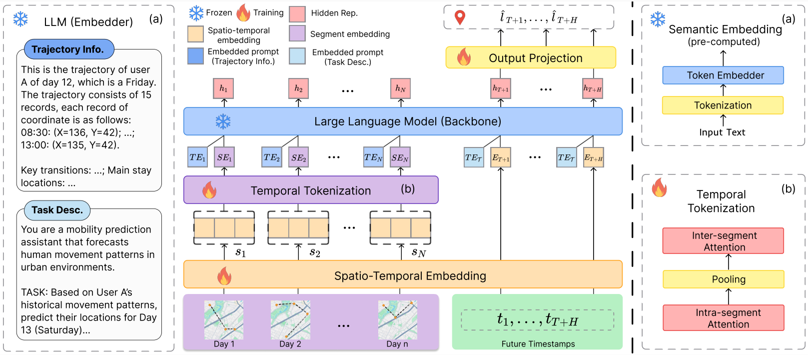


Figure 2: **The proposed architecture of RHYTHM.** Our framework processes historical trajectories through spatio-temporal embedding and temporal tokenization (**b**), capturing local and global dependencies via hierarchical attention. Segment representations are enriched with semantic embeddings from trajectory information, while future timestamps incorporate task description context (**a**). This combined sequence passes through a frozen LLM backbone with output projection to generate location predictions.

Spatio-Temporal Feature Encoding

For each observation (t_i, l_i)

Temporal Embedding:

$$\mathsf{E}_i^{\text{temporal}} = \mathsf{E}^{\text{ToD}}(t_i) \| \mathsf{E}^{\text{DoW}}(t_i),$$

Spatial Embedding:

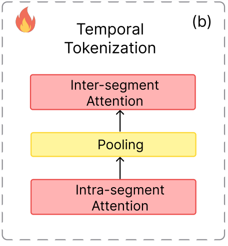
$$\mathsf{E}_i^{\text{spatial}} = \mathsf{E}^{\text{Loc}}(l_i) \| (W_{\text{coord}}[\text{lat}_i, \text{lon}_i]^T + b_{\text{coord}}),$$

Spatio-Temporal Embedding:

$$\mathsf{E}_i = \mathsf{E}_i^{\text{temporal}} + \mathsf{E}_i^{\text{spatial}}.$$

Capture multi-scale patterns through hierarchical attention reducing $T \to N$ segments

$$\begin{aligned}
\mathsf{s}_i &= \{E_{(i-1)L+1}, \ldots, E_{iL}\} \quad \text{(partition)} \\
\tilde{\mathsf{E}}^{(i)} &= \mathrm{Attention}(\mathsf{s}_i) \quad \text{(local patterns)} \\
\mathsf{SE}_i &= \mathrm{Pool}(\tilde{\mathsf{E}}^{(i)}) \quad \text{(compress)} \\
\tilde{\mathsf{SE}}_{1:N} &= \mathrm{Attention}(\mathsf{SE}_{1:N}) \quad \text{(global patterns)}
\end{aligned}$$

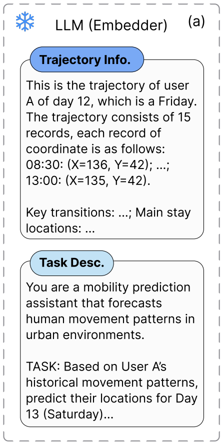


Semantic Context Integration

Semantic Embedding: Enrich tokens with trajectory context using frozen LLM

$$\mathsf{TE}_i = \mathrm{SelectLast}(\mathrm{LLM}(\mathrm{Prompt}(s_i)))$$

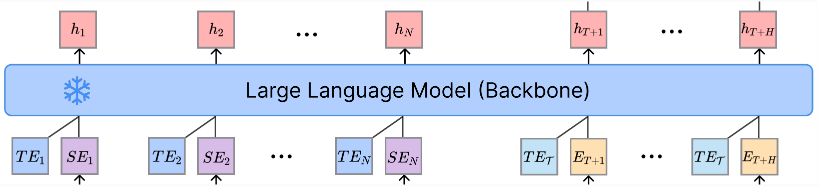
$$\mathsf{CE}_i = \widetilde{\mathsf{SE}}_i + \mathsf{TE}_i$$



Cross-representational Mobility Prediction

Leverage pretrained backbone for location prediction

$$h_i = \mathrm{LLM}(\mathsf{CE}_i)$$
$$P(l_{T+j}|\mathcal{X}, \mathcal{T}) = \mathrm{softmax}(W_o h_{T+j} + b_o)$$



Computational Efficiency

- Offline semantic embedding computation—no runtime LLM inference
- Attention complexity: $\mathcal{O}((T+H)^2) \rightarrow \mathcal{O}((N+H)^2)$ where $N \ll T$
- Frozen LLM backbone: only **12.37%** trainable parameters with **24%** training time reduction!

Experiments

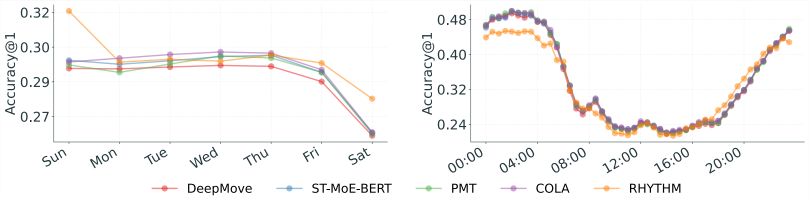
Overall Performance

RHYTHM outperforms all baselines with **2.4%** higher prediction accuracy across most metrics.

Table 1: **Performance of RHYTHM and baselines on the Kumamoto, Sapporo, and Hiroshima datasets.** The evaluation metrics include Accuracy@k for different values of k, with variance $\leq 2\%$. The best results are highlighted in **bold**, and the second-best results are underlined. RHYTHM demonstrates superior performance compared to baselines across most configurations.

	Kumamoto			Sapporo			Hiroshima		
Model	Acc@1	Acc@3	Acc@5	Acc@1	Acc@3	Acc@5	Acc@1	Acc@3	Acc@5
LSTM	0.2652	0.4799	0.5472	0.2310	0.3940	0.4526	0.2129	0.3775	0.4415
DeepMove	0.2779	0.4986	0.5683	0.2825	0.4672	0.5264	0.2804	0.4810	0.5477
PatchTST	0.2751	0.5018	0.5716	0.2703	0.4582	0.5168	0.2752	0.4839	0.5522
iTransformer	0.2609	0.4724	0.5412	0.2696	0.4500	0.5070	0.2804	0.4857	0.5523
TimeLLM	0.2712	0.4848	0.5535	0.2792	0.4746	0.5352	0.2698	0.4753	0.5426
CMHSA	0.2862	0.5182	0.5887	0.2890	**0.4901**	<u>0.5525</u>	0.2874	0.5001	0.5684
PMT	0.2697	0.4475	0.5187	0.2878	<u>0.4896</u>	0.5522	0.2850	0.4982	0.5668
COLA	0.2864	0.5186	0.5896	0.2847	0.4865	0.5497	0.2874	0.5013	0.5708
ST-MoE-BERT	0.2862	0.5155	0.5871	0.2869	0.4856	0.5480	0.2839	0.4925	0.5601
Mobility-LLM	0.2666	0.4793	0.5448	0.2838	0.4703	0.5288	0.2826	0.4856	0.5525
RHYTHM-LLaMA-1B	<u>0.2929</u>	<u>0.5200</u>	0.5835	0.2931	0.4876	0.5502	0.2913	0.5027	0.5753
RHYTHM-Gemma-2B	0.2923	0.5191	<u>0.5932</u>	**0.2943**	0.4896	**0.5545**	**0.2953**	**0.5074**	**0.5798**
RHYTHM-LLaMA-3B	**0.2941**	**0.5205**	**0.5947**	<u>0.2938</u>	0.4875	0.5523	<u>0.2929</u>	<u>0.5032</u>	<u>0.5756</u>

Achieves **5.0%** improvement on challenging scenarios (weekends & peak hours) where baselines struggle.



Efficiency: Reduces training time by **24.6%** compared to best baseline while maintaining superior performance.
Scalability: Performance scales predictably with model size, following established scaling laws.

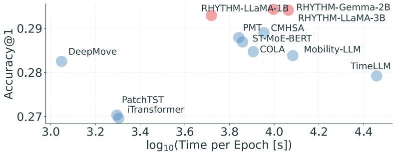


Figure 3: **Training Speed vs. performance of RHYTHM and baseline models on the Sapporo Dataset.**

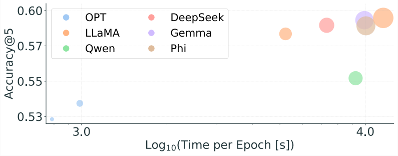


Figure 4: **Efficiency comparison of alternative LLMs, evaluated by the same configuration of Table 4.**

Conclusion

Summary

- Temporal tokenization captures multi-scale periodicity via hierarchical attention
- Prompt-guided approach enhances semantic pattern understanding
- 87.6% parameters frozen → 24.6% computational savings
- 2.4% improvement in prediction accuracy and a 5.0% increase on weekends.

Thank you!