



UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction

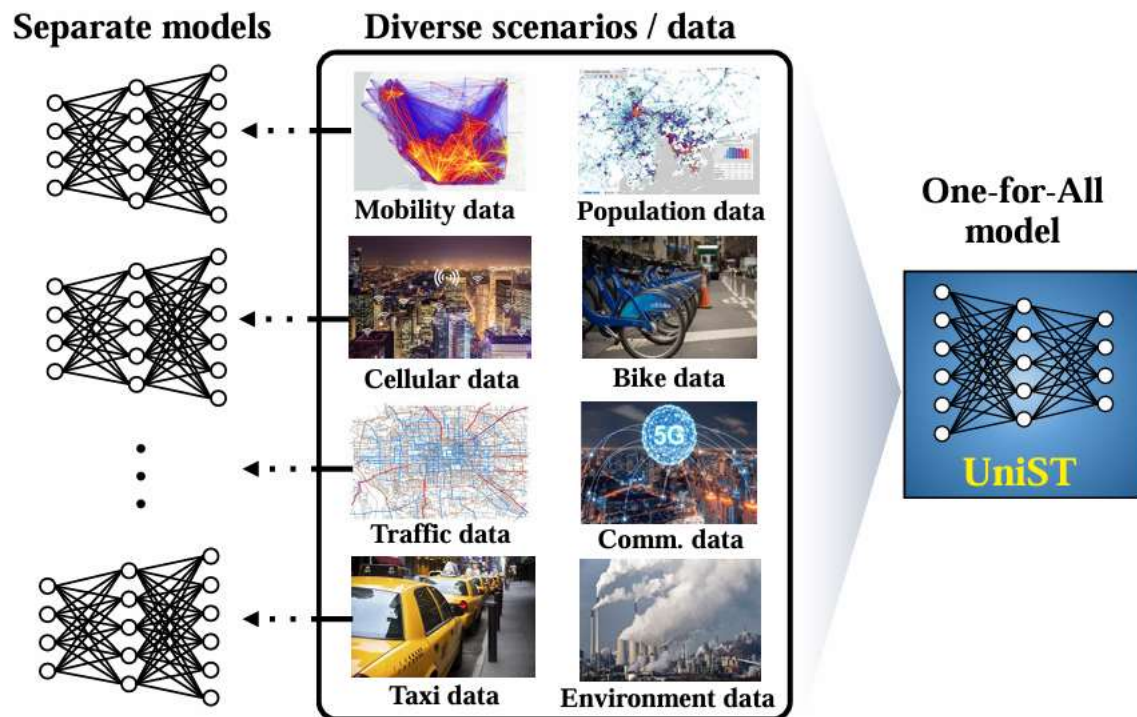


2025/02/16

I 引言

问题背景:

- 不同领域和城市的数字化水平不同→数据短缺
- 现有预测方法通常是针对特定时空场景定制的，需特定的模型设计和大量训练数据，缺乏通用性



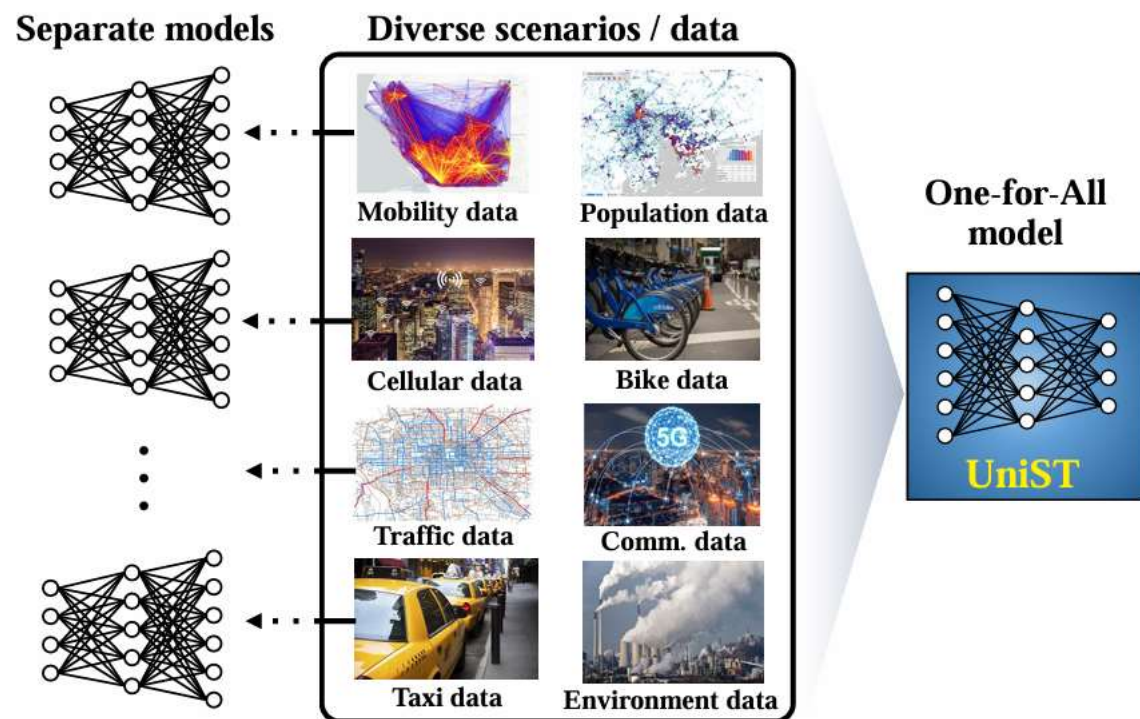
构建通用时空预测模型的挑战:

- 时空数据集具有不同的格式
- 不同时空场景的数据分布差异大

I 引言

通用时空预测模型的两个基本能力：

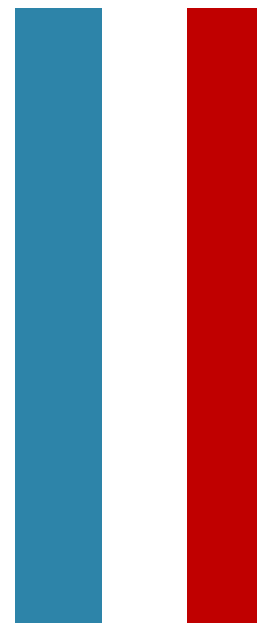
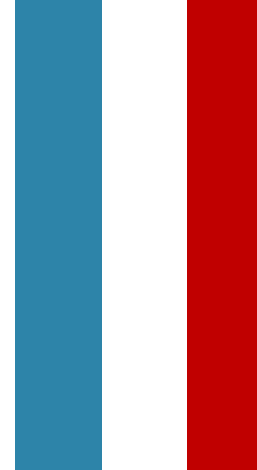
- 能够利用来自不同城市场景的数据进行训练
- 在不同的时空场景中表现出强大的泛化能力，尤其在训练数据有限或者完全缺失的场景中



UniST:

- 具有不同场景时空数据的可扩展性
- 通过有效的预训练，捕捉复杂的时空关系
- 利用时空提示对不同场景的底层共享模式进行对齐

II 方法

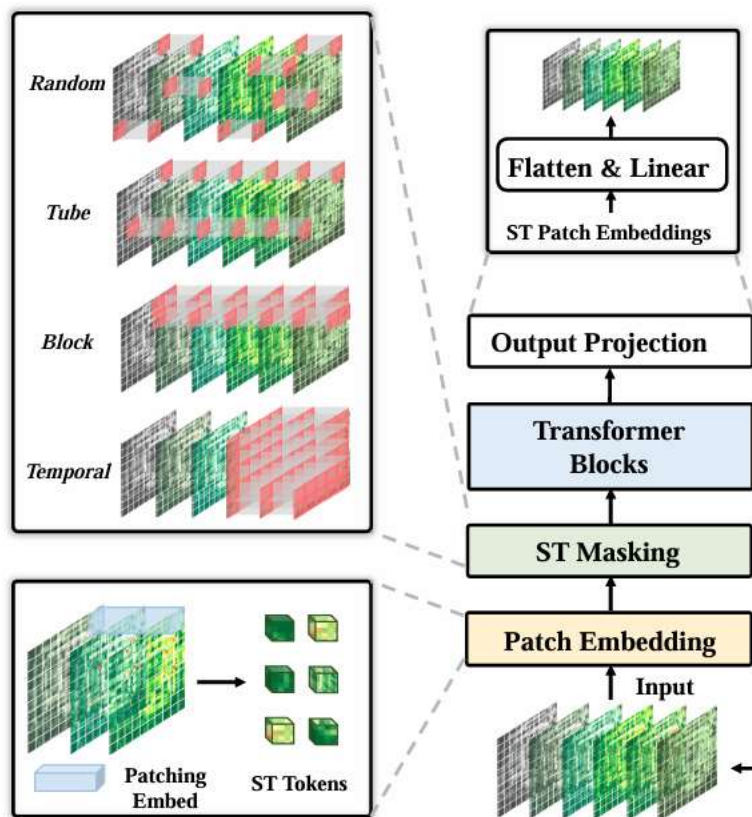


II 模型框架

1. 时空预训练

利用来自不同领域和城市的大量时空数据进行预训练

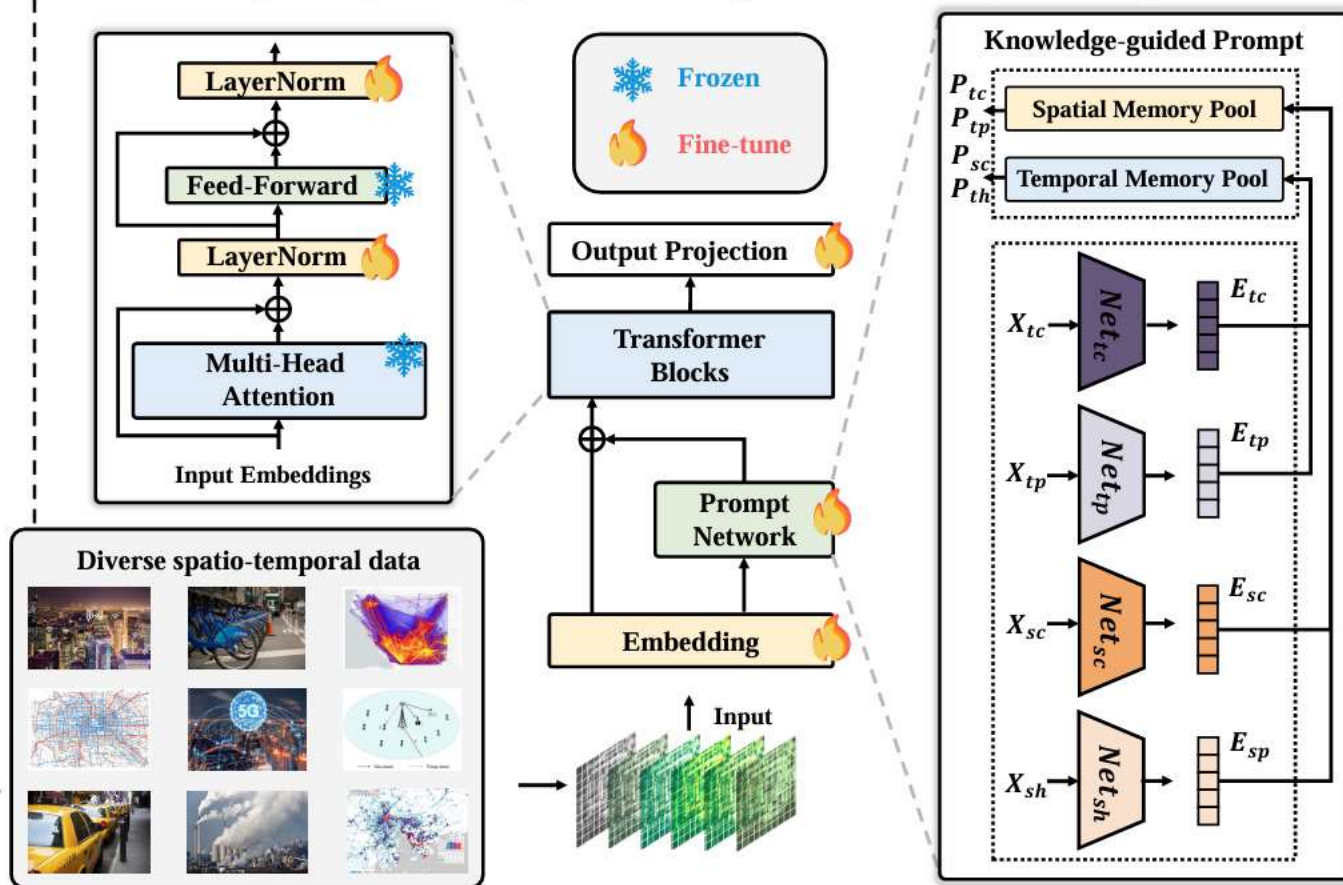
Stage 1 - Spatio-Temporal Pre-Training



2. 时空知识引导的提示学习

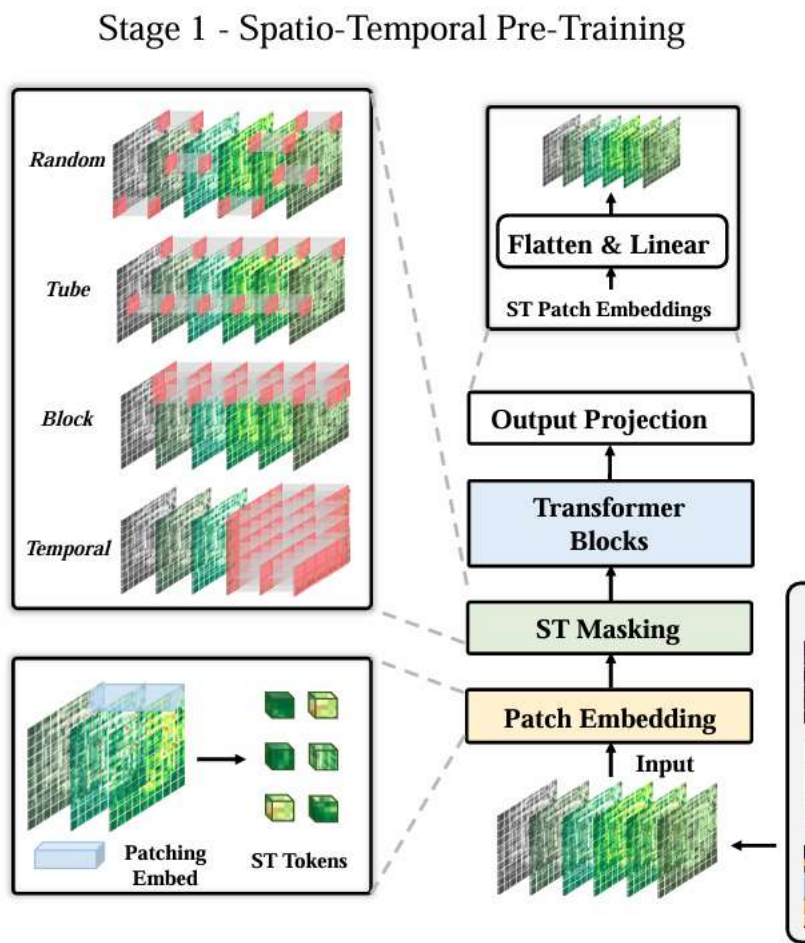
引入一个上下文学习的提示网络，提示的生成由时空领域知识(空间层次和时间周期性)自适应地生成

Stage 2 - Spatio-Temporal Knowledge-Guided Prompt Learning



II 模型框架

Stage 1 -Spatio-Temporal Pre-Training



基础模型： 基于transformer的编码器-解码器架构

时空Patching：

1. 数据拆分为通道独立的三维张量: $X \in \mathbb{R}^{L \times C \times H \times W} \rightarrow X_c \in \mathbb{R}^{L \times H \times W}$
2. 分块为更小的patches: $L \times H \times W \rightarrow l \times h \times w$
3. 三维→一维: $E_x = \text{CONV}_{3d}(X)$

位置编码：

选择正弦和余弦函数分别用于空间和时间维度实现位置编码

ST Masking

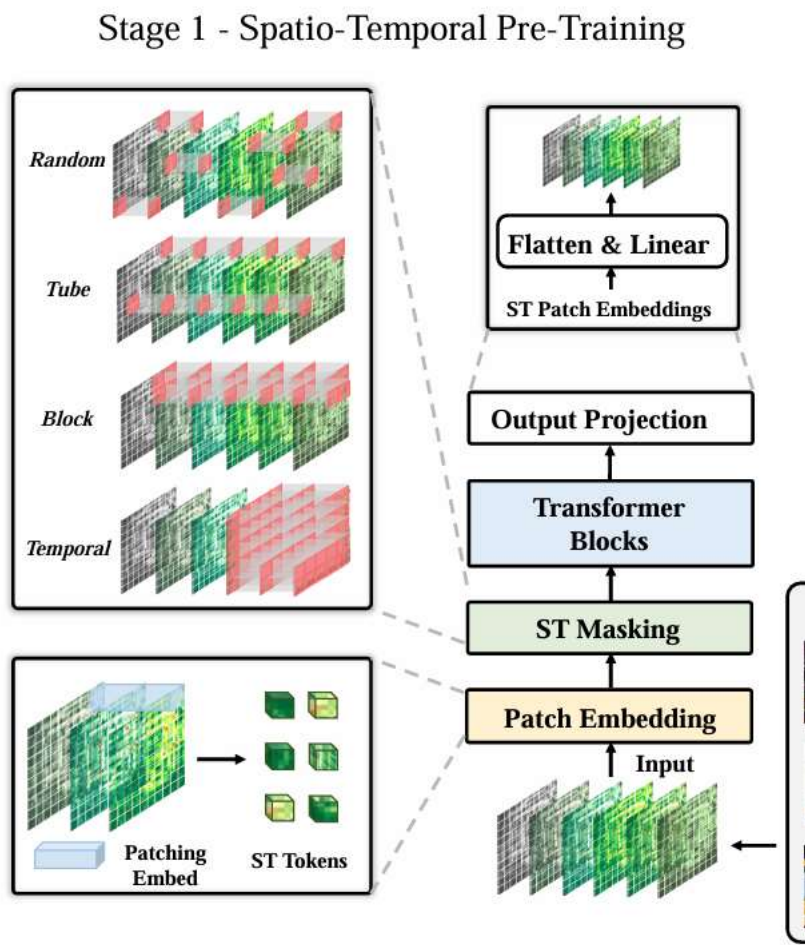
Encoder-Decoder结构：

Encoder处理未mask的块，Decoder利用encoder的输出和被mask的块重构数据

$$E_{enc} = \text{ENCODER}(E_x), Y_{dec} = \text{DECODER}(E_{enc}, E_{mask}),$$

II 模型框架

Stage 1 -Spatio-Temporal Pre-Training



时空自监督预训练

设掩码比例为 r ，引入四种不同的掩码策略：

随机掩码： $M \sim \mathbf{U}[0, 1]$, $E_x = E_x[M < 1 - r]$, $M \in \mathbb{R}^{L' \times H' \times W'}$.

管状掩码： $M \sim \mathbf{U}[0, 1]$, $E_x = E_x[:, M < 1 - r]$, $M \in \mathbb{R}^{H' \times W'}$.

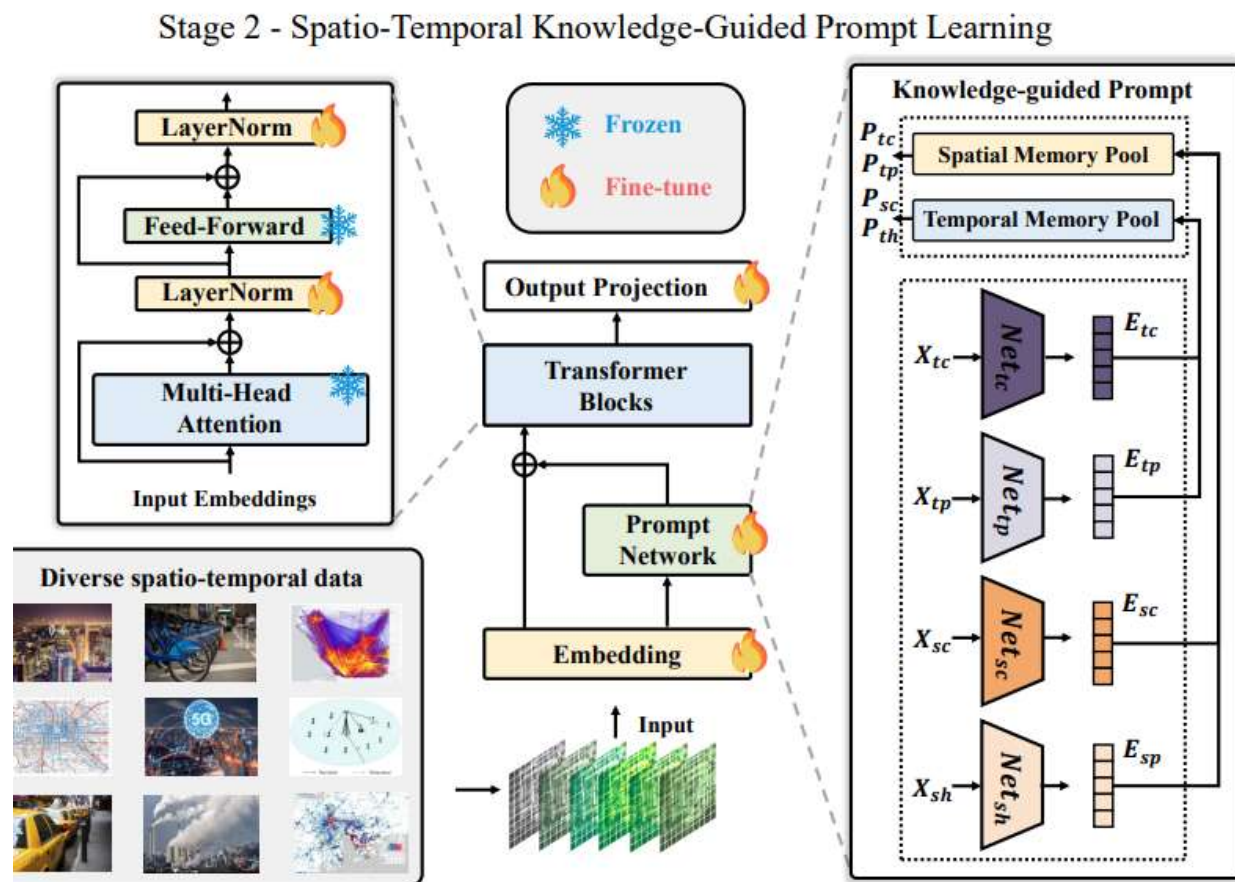
块掩码： $M \sim \mathbf{UNIFORM}(1, 2)$, $E_x = E_x[:, \frac{M-1}{2}H' : \frac{M}{2}H', \frac{M-1}{2}W' : \frac{M}{2}W']$.

时间掩码： $M = \text{CONCAT}([\mathbf{1}_{(1-r)L' \times H' \times W'}, \mathbf{0}_{rL' \times H' \times W'}])$, $E_x = E_x[M = 1]$.

掩码方式	掩盖范围	现实场景	目的
随机掩码	离散时空点	个别传感器短时故障	学习细粒度的时空模式
管状掩码	某些空间单元的长时间数据	个别传感器长期故障	提高空间外推能力
块掩码	整块空间区域的长时间数据	范围区域传感器长期故障/无传感器	增强模型的空间迁移能力
时间掩码	未来数据	预测未来数据	学习时间依赖关系

II 模型框架

Stage 2 - Spatio-Temporal Knowledge-Guided Prompt Learning



如何定义“相似性”模式以识别并对齐不同数据集中共享的时空模式？

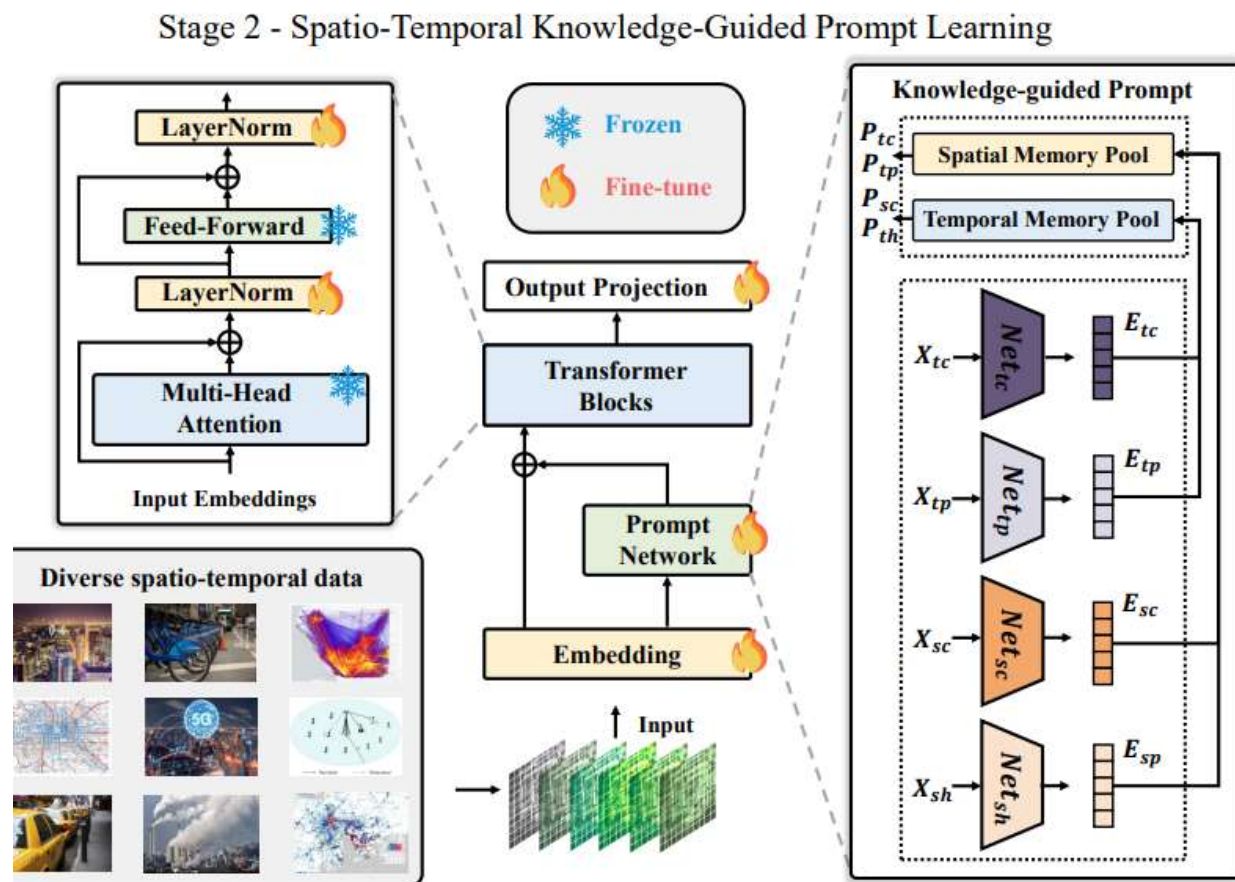
属性	解释
空间接近性sc	相近的空间单元会相互影响，具有更强的相关性或相似性
空间层次性sh	空间数据在不同尺度上的层次化结构。影响时空动态，需对城市结构进行多层次感知
时间接近性tc	时间上接近的事件或状态之间通常具有更强的相关性或依赖性
时间周期性tp	在时间序列数据中，某些模式以固定的时间间隔重复出现的特性

通过四个网络对输入数据分别在空间、时间维度上建模：

$$\begin{aligned}
 E_{sc} &= \text{CONV}_{2D}[3](X_s), \\
 E_{sh} &= \{\text{CONV}_{2D}[2^i + 1](X_s)\}, i \in \{2, 3, 4\}, \\
 E_{tc} &= \text{ATTENTION}(X_c), \\
 E_{tp} &= \text{ATTENTION}(X_p).
 \end{aligned}$$

II 模型框架

Stage 2 - Spatio-Temporal Knowledge-Guided Prompt Learning



如何利用时空知识引导提示的生成?

提出空间记忆池(Temporal Memory Pool)和时间记忆池(Spatial Memory Pool)

在提示学习的过程中, 记忆池会被优化, 用来存储有关时空领域知识的有价值信息

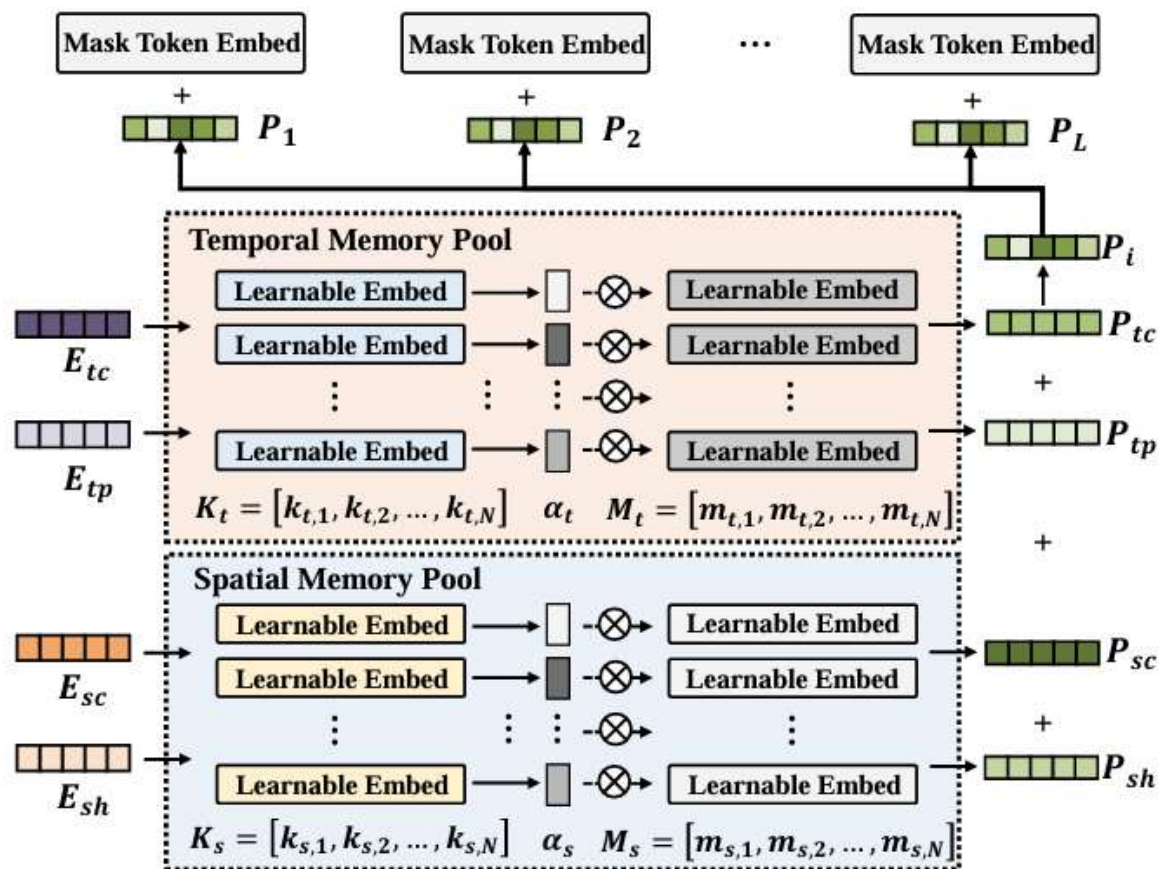
空间记忆池和时间记忆池定义如下:

$$KM_s = \{(k_{s,0}, m_{s,0}), (k_{s,1}, m_{s,1}), \dots, (k_{s,N-1}, m_{s,N-1})\},$$

$$KM_t = \{(k_{t,0}, m_{t,0}), (k_{t,1}, m_{t,1}), \dots, (k_{t,N-1}, m_{t,N-1})\},$$

II 模型框架

Stage 2 - Spatio-Temporal Knowledge-Guided Prompt Learning



提示的生成

空间记忆池和时间记忆池定义如下:

$$KM_s = \{(k_{s,0}, m_{s,0}), (k_{s,1}, m_{s,1}), \dots, (k_{s,N-1}, m_{s,N-1})\},$$

$$KM_t = \{(k_{t,0}, m_{t,0}), (k_{t,1}, m_{t,1}), \dots, (k_{t,N-1}, m_{t,N-1})\},$$

E_{sc} : 空间接近性; E_{sh} : 空间层次性

E_{tc} : 时间接近性; E_{tp} : 时间周期性

过程: 表示 E 与记忆池中的键 k 相似度计算 \rightarrow 权重 \rightarrow 从记忆池中提取信息 \rightarrow 生成提示 P

$$\alpha_{sc} = [k_{s,0}; k_{s,1}; \dots, k_{s,N-1}] E_{sc}^T, \quad P_{sc} = \sum_i \alpha_{sc,i} m_{s,i},$$

$$\alpha_{sh} = [k_{s,0}; k_{s,1}; \dots, k_{s,N-1}] E_{sh}^T, \quad P_{sh} = \sum_i \alpha_{sh,i} m_{s,i},$$

$$\alpha_{tc} = [k_{t,0}; k_{t,1}; \dots, k_{t,N-1}] E_{tc}^T, \quad P_{tc} = \sum_i \alpha_{tc,i} m_{t,i},$$

$$\alpha_{tp} = [k_{t,0}; k_{t,1}; \dots, k_{t,N-1}] E_{tp}^T, \quad P_{tp} = \sum_i \alpha_{tp,i} m_{t,i},$$

III 实验部分

III-A 实验设置

➤ 数据集

Dataset	Domain	City	Temporal Duration	Temporal interval	Spatial partition	Dataset	Min value	Max value	Mean value	Standard deviation
TaxiBJ	Taxi GPS	Beijing, China	20130601-20131030	Half an hour	32 × 32	TaxiBJ	0.0	1285	107	133
			20140301-20140630			Cellular	0.0	2992532	75258	149505
			20150301-20150630			TaxiNYC-1	0.0	1517	32	94
			20151101-20160410			TaxiNYC-2	0.0	1283	37	102
Cellular	Cellular usage	Nanjing, China	20201111-20210531	Half an hour	16 * 20	BikeNYC-1	0.0	266	9.2	18.1
TaxiNYC-1	Taxi OD	New York City, USA	20160101-20160229	Half an hour	16 * 12	BikeNYC-2	0.0	299	4.4	14.6
TaxiNYC-2	Taxi OD	New York City, USA	20150101-20150301	Half an hour	20 * 10	TDrive	0.0	2681	123	229
BikeNYC-1	Bike usage	New York City, USA	20160801-20160929	One hour	16 * 8	Crowd	0.0	593118	21656	40825
BikeNYC-2	Bike usage	New York City, USA	20160701-20160829	Half an hour	10 * 20	TrafficCS	0.0	22.25	6.22	4.79
TDrive	Taxi trajectory	New York City, USA	20150201-20160602	One hour	32 × 32	TrafficWH	0.0	22.35	6.22	4.68
Crowd	Crowd flow	Nanjing, China	20201111-20210531	Half an hour	16 * 20	TrafficCD	0.0	22.25	7.33	4.36
TrafficCS	Traffic speed	Changsha, China	20220305-20220405	Five minutes	28 × 28	TrafficJN	0.0	25.04	5.72	4.71
TrafficWH	Traffic speed	Wuhan, China	20220305-20220405	Five minutes	30 × 28	TrafficNJ	0.0	24.82	5.38	4.73
TrafficCD	Traffic speed	Chengdu, China	20220305-20220405	Five minutes	28 × 26	TrafficSH	0.0	21.83	7.92	3.86
TrafficJN	Traffic speed	Jinan, China	20220305-20220405	Five minutes	32 × 18	TrafficSZ	0.0	22.12	5.11	4.75
TrafficNJ	Traffic speed	Nanjing, China	20220305-20220405	Five minutes	32 × 24	TrafficGZ	0.0	25.16	5.26	4.79
TrafficSH	Traffic speed	Shanghai, China	20220127-20220227	Five minutes	28 × 32	TrafficGY	0.0	28.89	5.95	7.03
TrafficSZ	Traffic speed	Shenzhen, China	20220305-20220405	Five minutes	24 × 18	TrafficTJ	0.0	25.24	6.32	5.05
TrafficGZ	Traffic speed	Guangzhou, China	20220305-20220405	Five minutes	32 × 26	TrafficHZ	0.0	29.50	3.81	4.38
TrafficGY	Traffic speed	Guiyang, China	20220305-20220405	Five minutes	26 × 28	TrafficZZ	0.0	23.26	6.67	4.32
TrafficTJ	Traffic speed	Tianjin, China	20220305-20220405	Five minutes	24 × 30	TrafficBJ	0.0	22.82	6.30	4.22
TrafficHZ	Traffic speed	Hangzhou, China	20220305-20220405	Five minutes	28 × 24					
TrafficZZ	Traffic speed	Zhengzhou, China	20220305-20220405	Five minutes	26 × 26					
TrafficBJ	Traffic speed	Beijing, China	20220305-20220405	Five minutes	30 × 32					

- 基线模型：启发式方法、城市时空预测模型、视频预测模型、多变量时序预测模型、元学习方法
- 指标：平均绝对误差（MAE）、均方根误差（RMSE）

III-B 短期预测

输入步长: 6

预测步长: 6

基线模型:

在对应的数据集
上进行单独训练,
进行预测评估

UniST:

直接在所有数据
集上进行预测评
估

Model	TaxiBJ		Crowd		Cellular		BikeNYC		TrafficJN		TDrive		TrafficSH	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	53.77	29.82	17.80	6.79	72.94	27.57	11.41	3.43	1.38	0.690	150.2	74.5	1.24	0.771
ARIMA	56.70	39.53	21.87	10.23	81.31	40.22	12.37	3.86	1.20	0.651	211.3	108.5	1.17	0.769
STResNet	45.17	30.87	5.355	3.382	24.30	14.32	8.20	4.98	0.964	0.556	220.1	117.4	1.00	0.723
ACFM	37.77	21.59	4.17	2.34	22.79	12.00	<u>3.93</u>	1.67	0.920	0.559	98.1	51.9	0.833	0.566
STID	<u>27.36</u>	<u>14.01</u>	3.85	1.63	18.77	8.24	4.06	1.54	0.880	0.495	47.4	23.3	<u>0.742</u>	<u>0.469</u>
STNorm	29.37	15.71	4.44	2.09	19.77	8.19	4.45	1.66	0.961	0.532	54.3	47.9	0.871	0.579
STGSP	45.04	28.28	7.93	4.56	39.99	21.40	5.00	1.69	0.882	0.490	94.6	47.8	1.02	0.749
MC-STL	29.14	15.83	4.75	2.39	21.22	10.26	4.08	2.05	1.19	0.833	54.2	28.1	1.00	0.720
PromptST	27.44	14.54	<u>3.52</u>	<u>1.54</u>	<u>15.74</u>	<u>7.20</u>	4.36	<u>1.57</u>	0.953	0.490	47.5	22.8	0.811	0.523
MAU	38.14	20.13	4.94	2.35	39.09	18.73	5.22	2.06	1.28	0.697	48.8	22.1	1.37	0.991
PredRNN	27.50	14.29	5.13	2.36	24.15	10.44	5.00	1.74	<u>0.852</u>	<u>0.463</u>	54.9	25.2	0.748	0.469
MIM	28.62	14.77	5.66	2.27	21.38	9.37	4.40	1.62	1.17	0.650	51.4	<u>22.7</u>	0.760	0.505
SimVP	32.66	17.67	3.91	1.96	16.48	8.23	4.11	1.67	0.969	0.556	<u>46.8</u>	22.9	0.814	0.569
TAU	33.90	19.37	4.09	2.11	17.94	8.91	4.30	1.83	0.993	0.566	51.6	28.1	0.820	0.557
PatchTST	42.74	22.23	10.25	3.62	43.40	15.74	5.27	1.65	1.25	0.616	106.4	51.3	1.10	0.663
iTransformer	36.97	19.14	9.40	3.40	37.01	13.93	7.74	2.53	1.11	0.570	86.3	42.6	1.04	0.655
PatchTST(one-for-all)	43.66	23.16	13.51	5.00	56.80	20.56	9.97	3.05	1.30	0.645	127.0	59.26	1.13	0.679
UniST(one-for-all)	26.84	13.95	3.00	1.38	14.29	6.50	3.50	1.27	0.843	0.430	44.97	19.67	0.665	0.405

- 基线模型: 启发式方法、城市时空预测模型、视频预测模型、多变量时序预测模型
- 指标: 平均绝对误差 (MAE)、均方根误差 (RMSE)

III-C 长期预测

输入步长: 64

预测步长: 64

基线模型:

在对应的数据集上进行单独训练, 进行预测评估

UniST:

直接在所有数据集上进行预测评估

Model	TaxiNYC		Crowd		BikeNYC	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
HA	61.03	21.33	19.57	8.49	11.00	3.66
ARIMA	68.0	28.66	21.34	8.93	11.59	3.98
STResNet	29.54	14.46	8.75	5.58	7.15	3.87
ACFM	32.91	13.72	6.16	3.35	4.56	1.86
STID	24.74	11.01	<u>4.91</u>	<u>2.63</u>	4.78	2.24
STNorm	31.81	11.99	9.62	4.30	6.45	2.18
STGSP	28.65	10.38	17.03	8.21	4.71	<u>1.54</u>
MC-STL	29.29	17.36	9.01	6.32	4.97	2.61
MAU	26.28	9.07	20.13	8.49	6.18	2.13
PredRNN	21.17	<u>7.31</u>	19.70	10.66	5.86	1.97
MIM	63.36	29.83	15.70	8.81	7.58	2.81
SimVP	<u>20.18</u>	9.78	5.50	3.13	4.10	1.71
TAU	24.97	10.93	5.31	2.81	<u>3.89</u>	1.73
PatchTST	30.64	17.49	5.25	2.83	5.27	1.65
iTransformer	33.81	11.48	6.94	2.63	6.00	2.02
PatchTST(one-for-all)	34.50	10.63	6.39	2.92	6.02	1.83
UniST (one-for-all)	19.83	6.71	4.25	2.26	3.56	1.31

- 基线模型: 启发式方法、城市时空预测模型、视频预测模型、多变量时序预测模型
- 指标: 平均绝对误差 (MAE)、均方根误差 (RMSE)

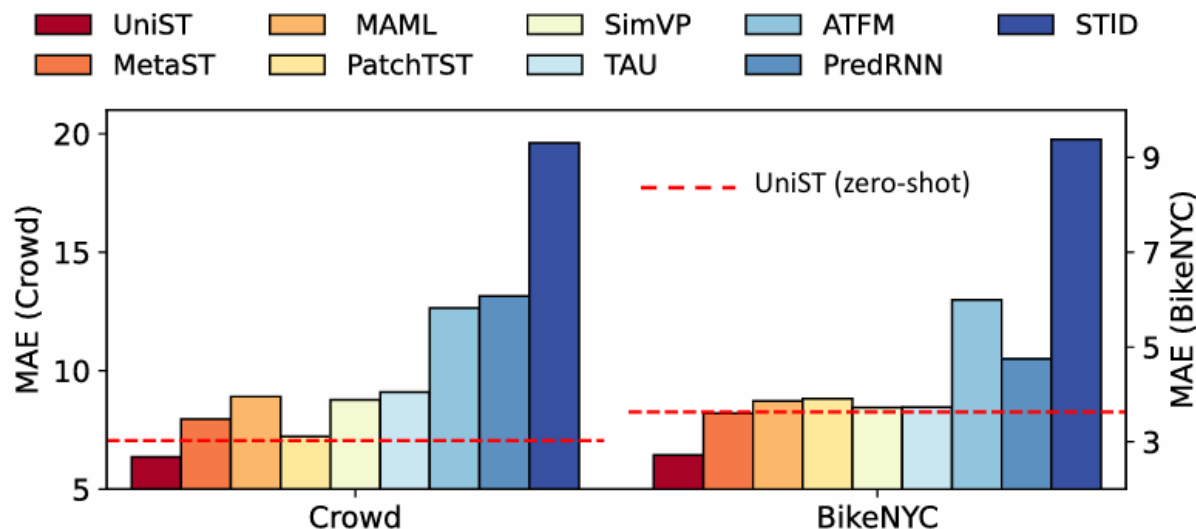
III-D 少样本预测、零样本预测

数据集：训练集、验证集、测试集

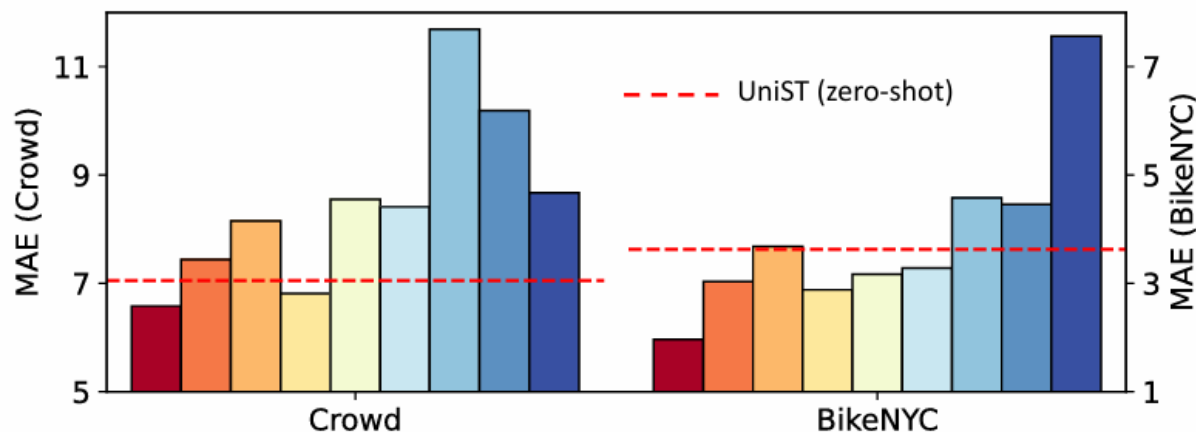
少样本预测：
分别选择了训练数据的1%、5%、10%
进行训练

选择较好的基线模型进行少样本评估
元学习基线：MAML、MetaST

零样本预测



(a) Performance on Few-shot (1%) and Zero-shot Scenarios



(b) Performance on Few-shot (5%) and Zero-shot Scenarios

III-E 消融实验

研究时间和空间属性是否有助于
提高整体性能

c: 时间接近性

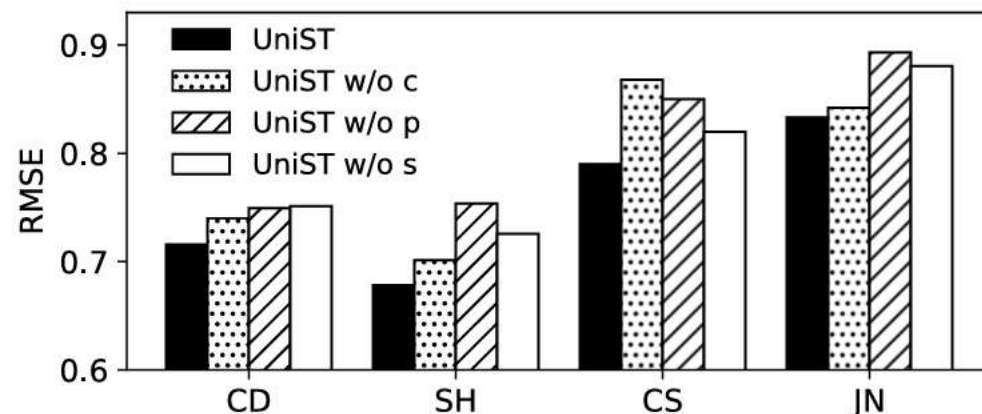
p: 时间周期性

s: 空间接近性和空间层次性

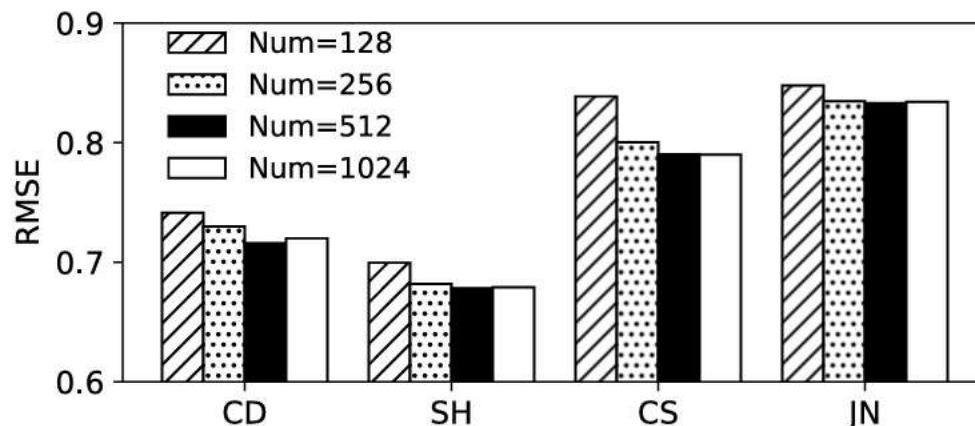
研究记忆池中embedding的数量

对模型性能的影响

数量为512时相对最优



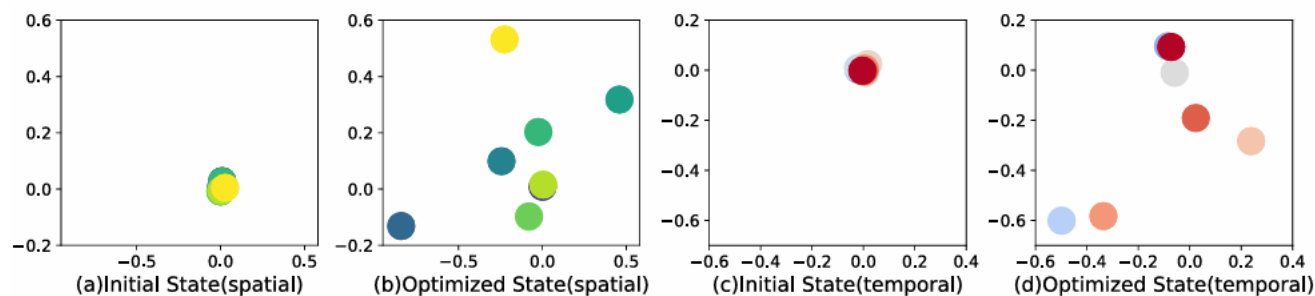
(a) Spatial and Temporal Properties



(b) Number of Embeddings

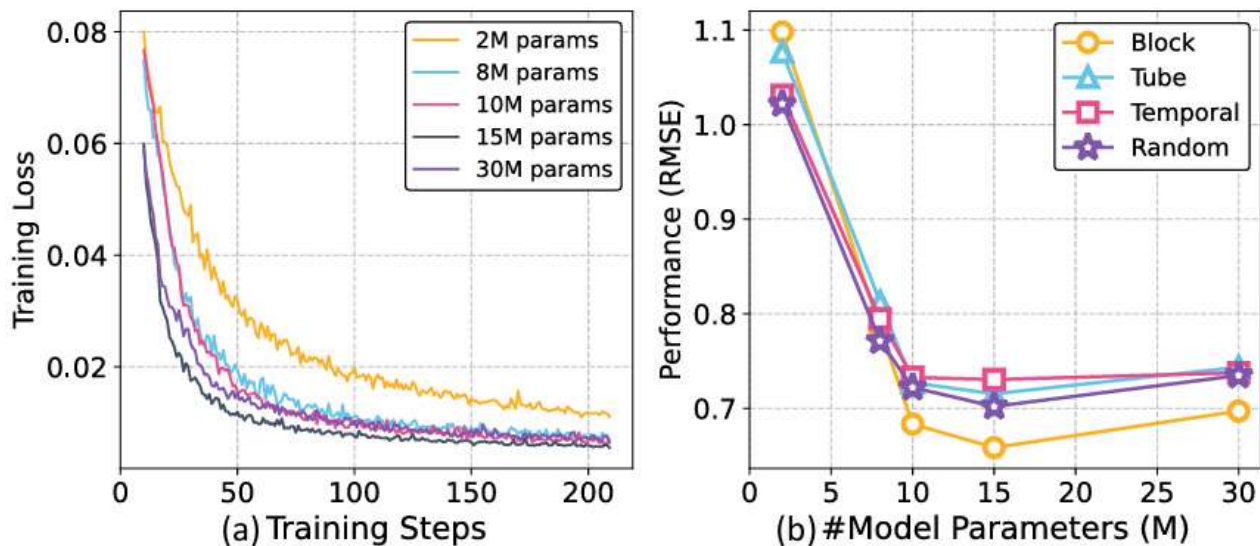
III-F 提示学习器分析、可扩展性

空间和时间记忆池在初始和最终优化状态下的可视化嵌入
嵌入表现出明显的分化。



可拓展性

模型在不同参数大小下的训练损失(左)和在测试集上的RMSE(右)



IV 总结





IV 总结

■ 主要工作

- 设计了一个用于城市时空预测的通用模型UniST
- UniST具备在多场景下进行预测的强大能力，特别是在少样本和零样本场景下

■ 未来方向

- 整合更多时空数据格式，比如：图数据，进一步提升模型的性能和适应性



谢谢！

