

Pre-training Enhanced Spatial-temporal Graph Neural
Network for Multivariate Time Series Forecasting
(KDD ' 22)

基于预训练模型增强的多元时间序列预测时空图神经网络
论文分享

张智博 ZY2306127

2023年7月28日

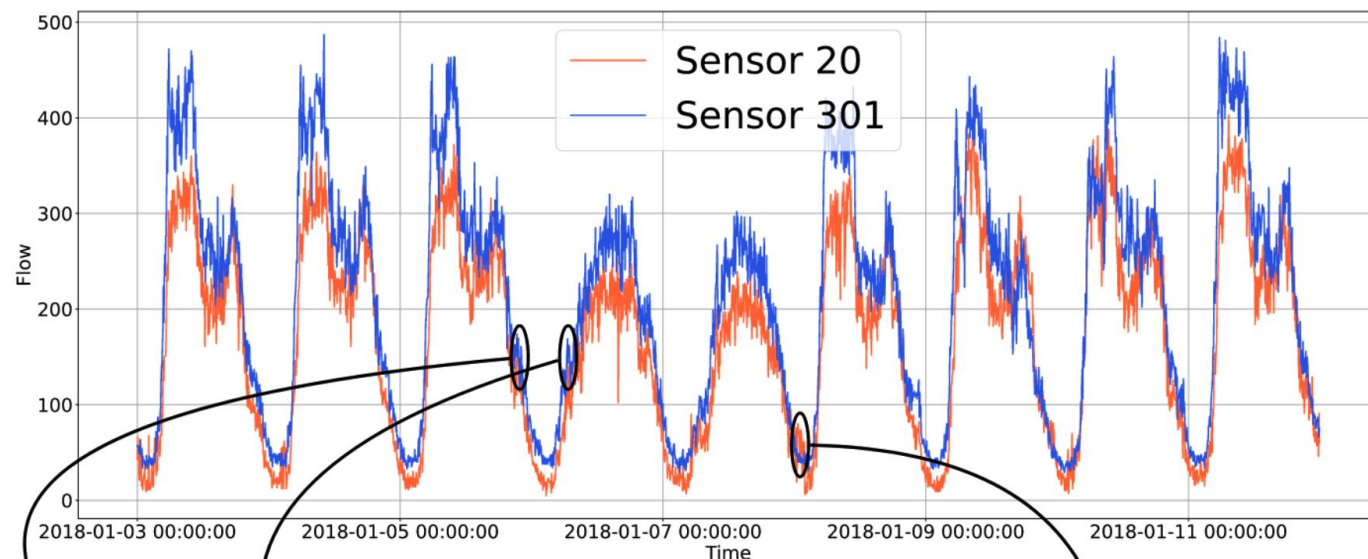
Motivation

时空数据

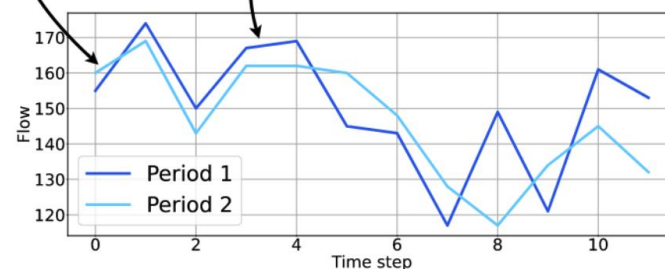
- 数据量大，噪声较大，信息密度低
- 周期特征需要在较长的时间尺度下观测

STGNN模型不足:

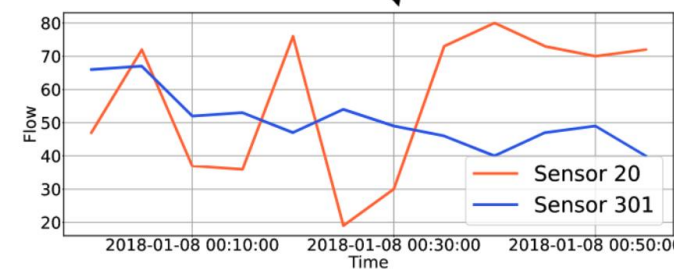
- 难以处理长期历史时空序列
- 基于小窗口的图模式学习不可靠
- 受噪声影响较大



(a) Traffic flow over 9 days in PeMS04 datasets.



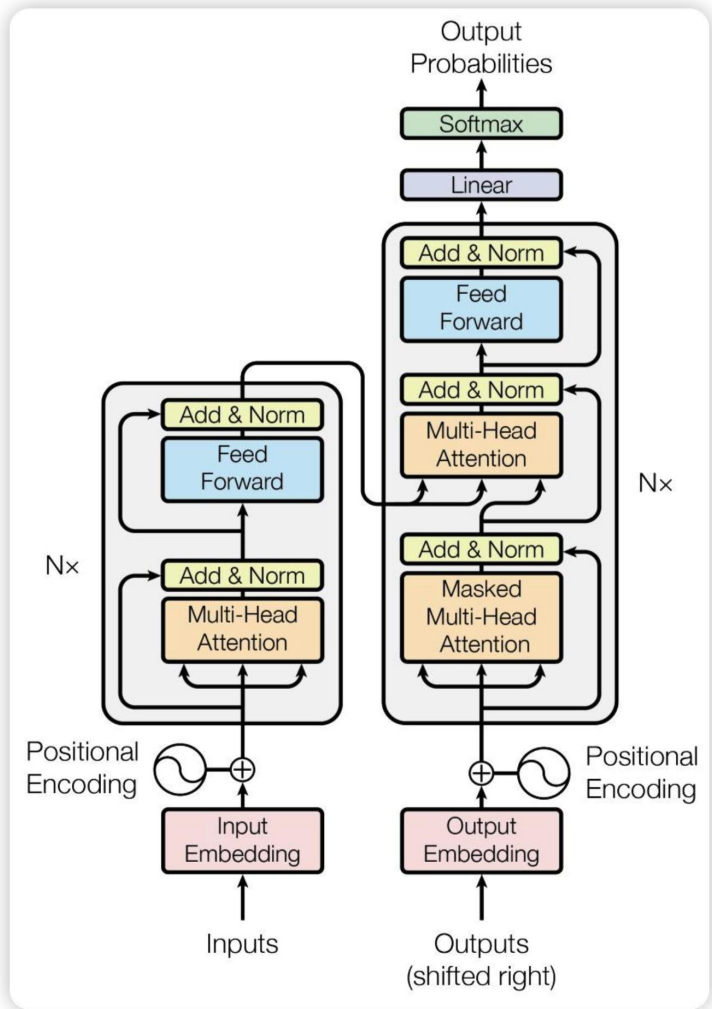
(b) Similar traffic trend in different context.



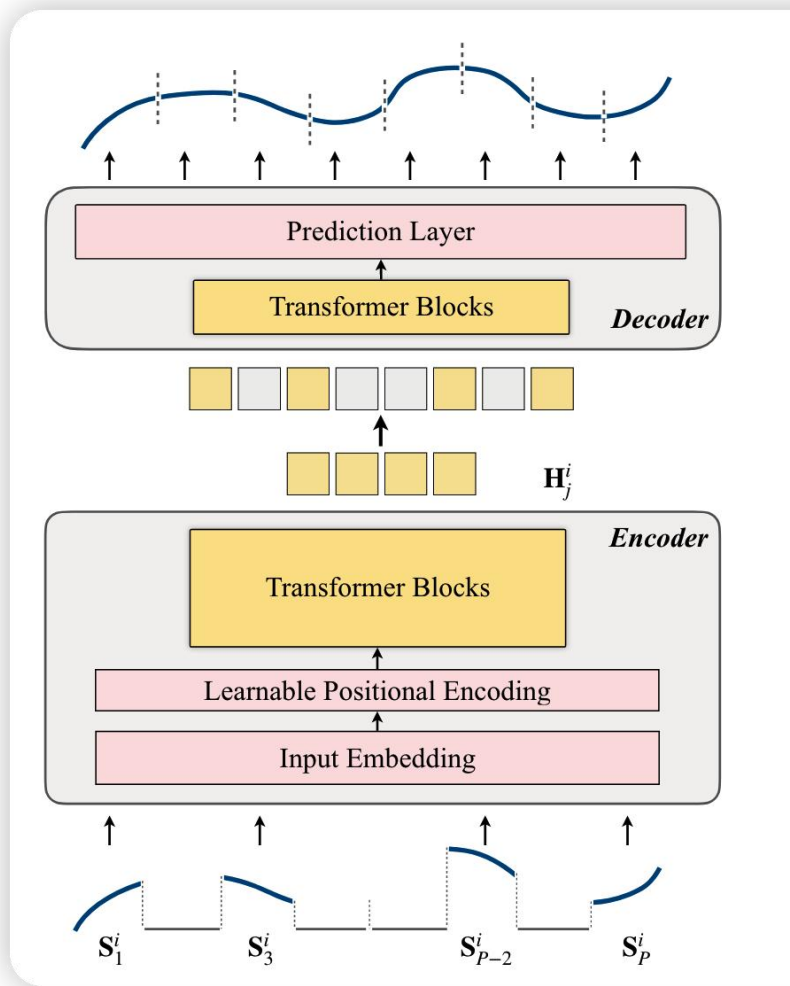
(c) Different traffic trend between similar series.

- TSformer: efficient unsupervised pre-training model for Time Series based on TransFormer blocks
- 使用Transformer捕获超长历史上下文特征以增强STGNN，为短期输入提供丰富的上下文

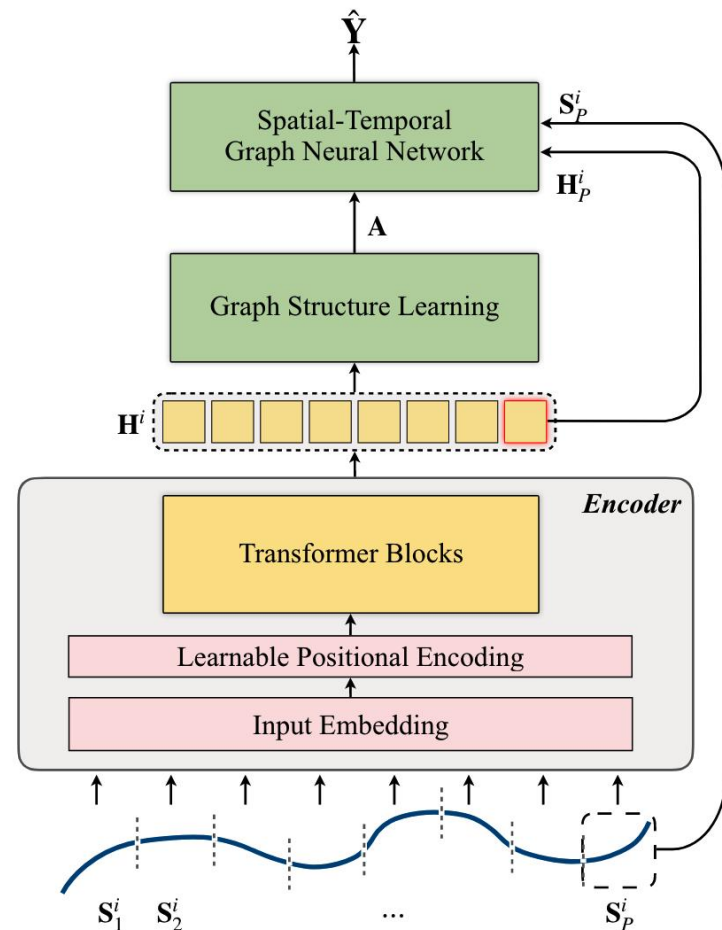
模型结构



Transformer典型结构



STEP-预训练阶段 (pre-training)



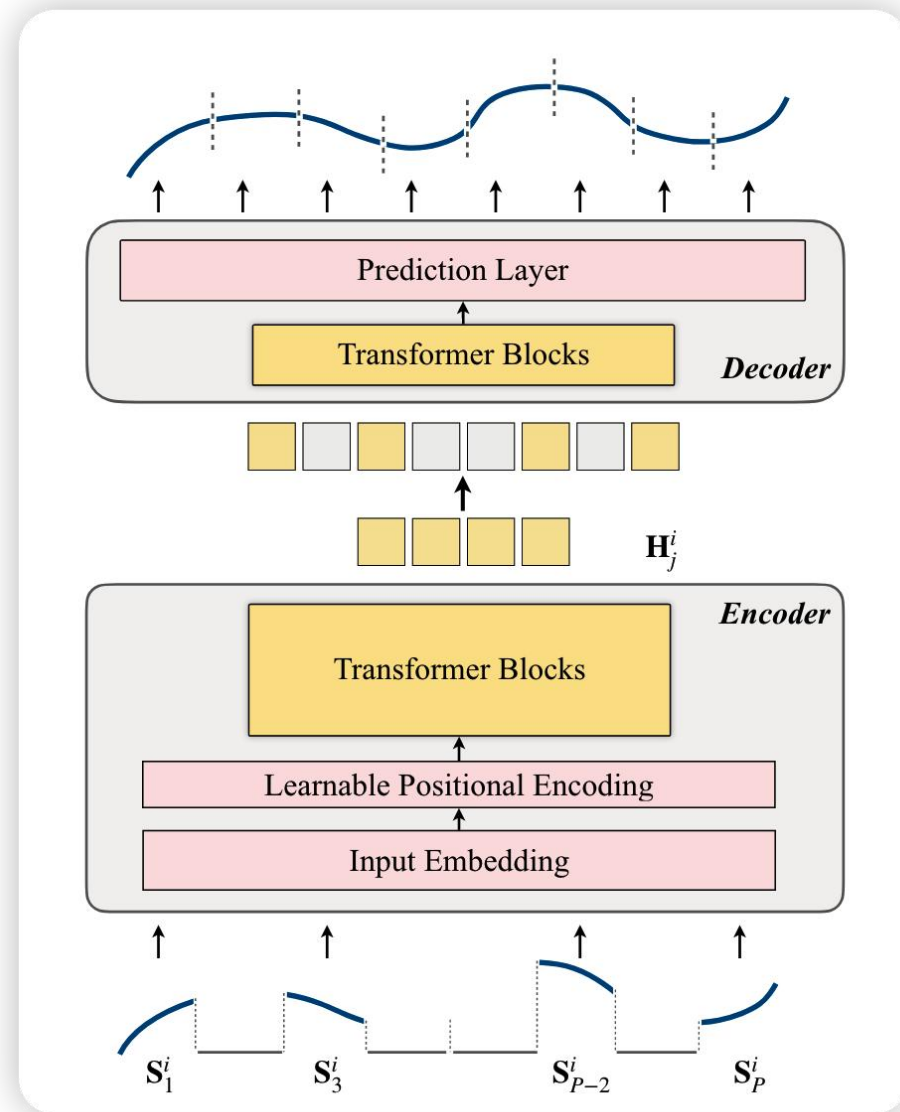
STEP-预测阶段 (forecasting)

预训练(pre-training)阶段

掩码自动编码策略 (MAE) + 可学习位置嵌入

Mask部分:

- 采用固定长度滑动窗口分割时间序列
- 预训练阶段使用**片段块**而非单个数据点作为输入单元
- 平均分为P个块，块长度L与下游STGNN的输入长度相同
- 使用更高的屏蔽率 (75%) 以学习大时间跨度的特征



预训练(pre-training)阶段

Encoder部分:

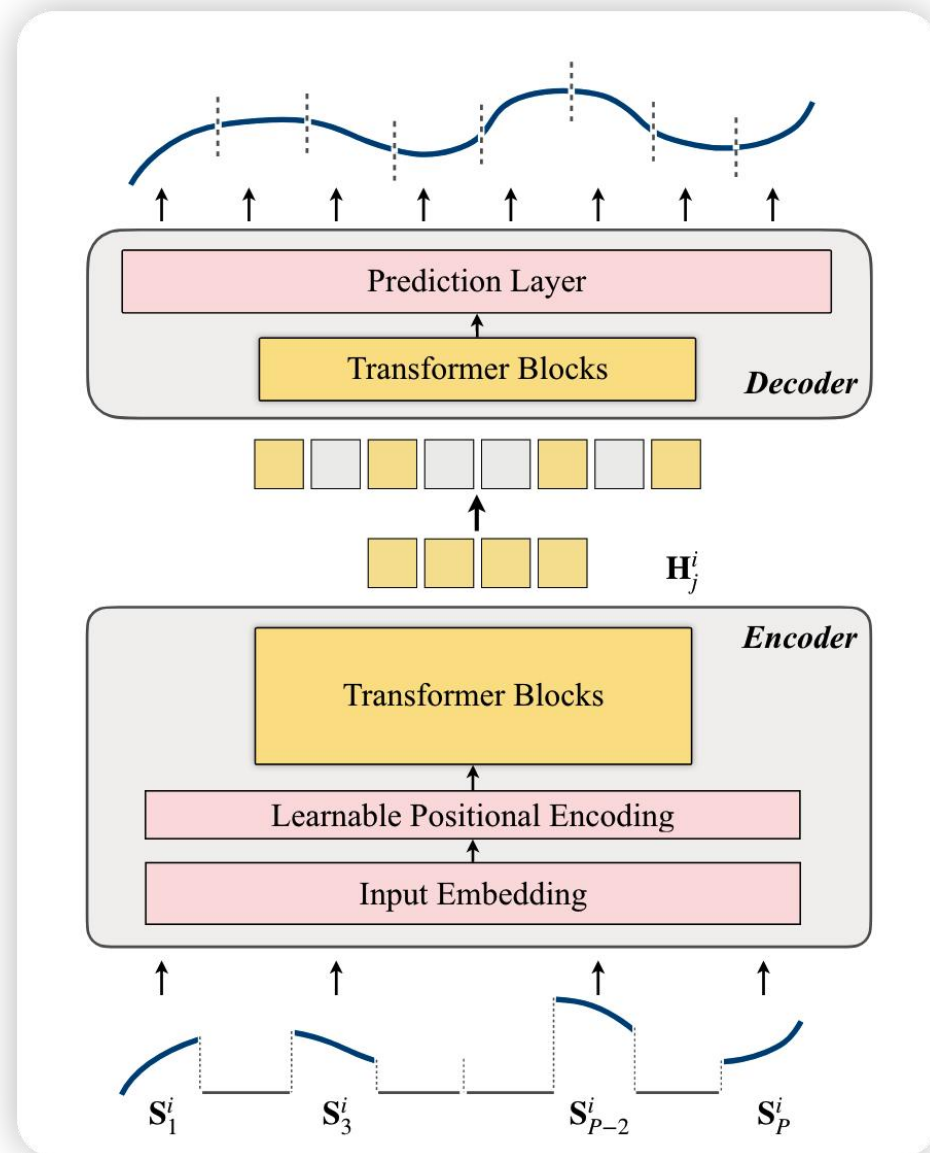
- 屏蔽块使用**可学习的位置嵌入**(positional embedding)而非确定的正弦位置嵌入以捕捉周期性特征
- 使用4层Transformer块, 提取所有**未屏蔽块**的潜在表示
- 预测阶段(STGNN)冻结Encoder参数以提高效率

Decoder部分:

- 仅在预训练阶段使用, 使用**单层**Transformer块
- 不使用positional embedding (Encoder已添加位置信息)
- 输入所有块的潜在表示, 输出重构序列

训练和损失:

- 采用掩码自动编码器(Masked Auto Encoder)训练
- 使用MAE(平均绝对值误差)作为损失函数
- 只计算屏蔽块的损失



预测 (forecasting) 阶段

问题:

- STGNN只能处理较短的序列 (L=12) 直接扩展难以优化
- 图结构学习和STGNN紧密耦合, 且**没有自监督损失信息**

基于TSformer的时空图结构学习器 (Graph structure learning):

- 使用上游Encoder提供的上下文表示为图结构优化提供监督信息
- **?** NN图作为正则化项, 学习离散时空依赖图

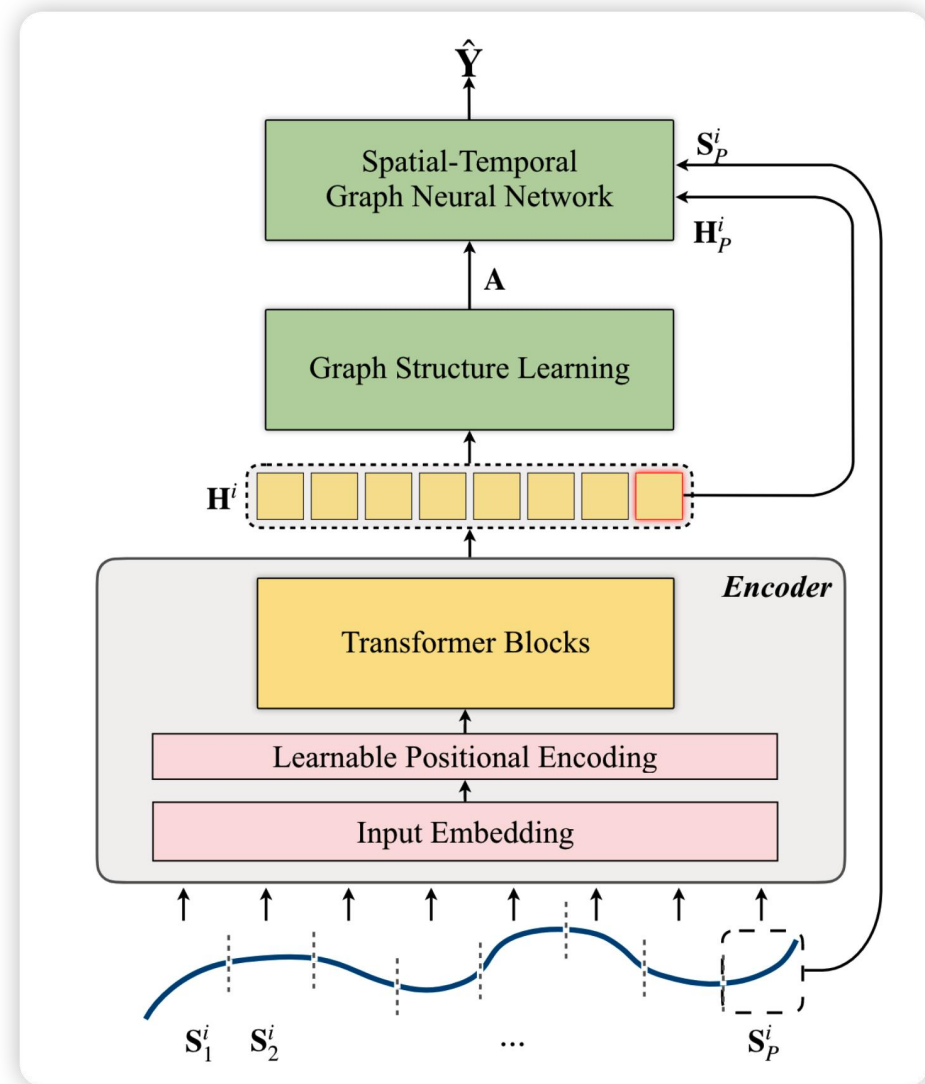
$$\Theta_{ij} = \text{FC}(\text{relu}(\text{FC}(\mathbf{Z}^i \parallel \mathbf{Z}^j)))$$

$$\mathbf{A}_{ij} = \text{softmax}((\Theta_{ij} + \mathbf{g})/\tau),$$

$$\mathbf{Z}^i = \text{relu}(\text{FC}(\mathbf{H}^i)) + \mathbf{G}^i,$$

$$\mathbf{G}^i = \text{FC}(\text{vec}(\text{Conv}(\mathbf{S}_{train}^i))), \quad \mathcal{L}_{graph} = \sum_{ij} -\mathbf{A}_{ij}^a \log \Theta'_{ij} - (1 - \mathbf{A}_{ij}^a) \log(1 - \Theta'_{ij}),$$

- 将表示转换为反映节点关系的稀疏邻接矩阵 \mathbf{A}_{ij}
- 将上下文信息用于指导图结构和STGNN的联合训练



测试与验证

Baseline: STEP + Graph WaveNet

消融实验

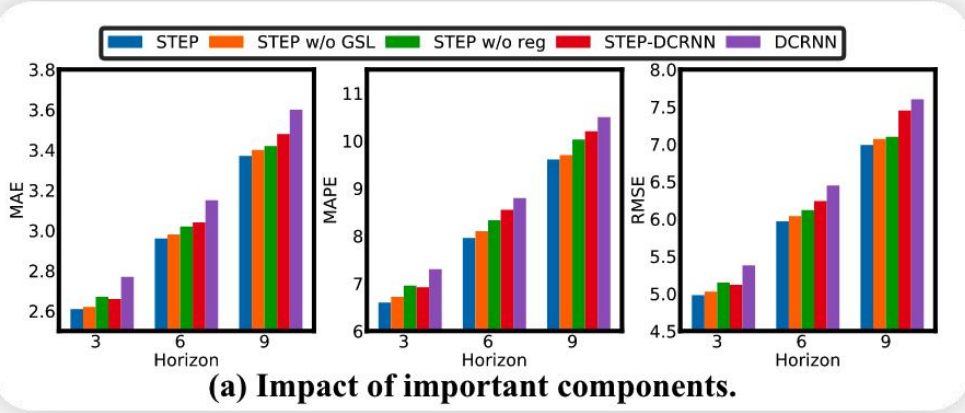


Table 2: Multivariate time series forecasting on the METR-LA, PEMS-BAY, and PEMS04 datasets. Numbers marked with * indicate that the improvement is statistically significant compared with the best baseline (t-test with p-value < 0.05).

Datasets	Methods	Horizon 3			Horizon 6			Horizon 12		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA	HA	4.79	10.00	11.70%	5.47	11.45	13.50%	6.99	13.89	17.54%
	VAR	4.42	7.80	13.00%	5.41	9.13	12.70%	6.52	10.11	15.80%
	SVR	3.39	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%
	FC-LSTM	3.44	6.30	9.60%	3.77	7.23	10.09%	4.37	8.69	14.00%
	DCRNN	2.77	5.38	7.30%	3.15	6.45	8.80%	3.60	7.60	10.50%
	STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%
	Graph WaveNet	2.69	5.15	6.90%	3.07	6.22	8.37%	3.53	7.37	10.01%
	ASTGCN	4.86	9.27	9.21%	5.43	10.61	10.13%	6.51	12.52	11.64%
	STSGCN	3.31	7.62	8.06%	4.13	9.77	10.29%	5.06	11.66	12.91%
	GMAN	2.80	5.55	7.41%	3.12	6.49	8.73%	3.44	7.35	10.07%
	MTGNN	2.69	5.18	6.88%	3.05	6.17	8.19%	3.49	7.23	9.87%
GTS	2.67	5.27	7.21%	3.04	6.25	8.41%	3.46	7.31	9.98%	
STEP	2.61*	4.98*	6.60%*	2.96*	5.97*	7.96%*	3.37*	6.99*	9.61%*	
PEMS-BAY	HA	1.89	4.30	4.16%	2.50	5.82	5.62%	3.31	7.54	7.65%
	VAR	1.74	3.16	3.60%	2.32	4.25	5.00%	2.93	5.44	6.50%
	SVR	1.85	3.59	3.80%	2.48	5.18	5.50%	3.28	7.08	8.00%
	FC-LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%
	DCRNN	1.38	2.95	2.90%	1.74	3.97	3.90%	2.07	4.74	4.90%
	STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%
	Graph WaveNet	1.30	2.74	2.73%	1.63	3.70	3.67%	1.95	4.52	4.63%
	ASTGCN	1.52	3.13	3.22%	2.01	4.27	4.48%	2.61	5.42	6.00%
	STSGCN	1.44	3.01	3.04%	1.83	4.18	4.17%	2.26	5.21	5.40%
	GMAN	1.34	2.91	2.86%	1.63	3.76	3.68%	1.86	4.32	4.37%
	MTGNN	1.32	2.79	2.77%	1.65	3.74	3.69%	1.94	4.49	4.53%
GTS	1.34	2.83	2.82%	1.66	3.78	3.77%	1.95	4.43	4.58%	
STEP	1.26*	2.73*	2.59%*	1.55*	3.58*	3.43%*	1.79*	4.20*	4.18%*	
PEMS04	HA	28.92	42.69	20.31%	33.73	49.37	24.01%	46.97	67.43	35.11%
	VAR	21.94	34.30	16.42%	23.72	36.58	18.02%	26.76	40.28	20.94%
	SVR	22.52	35.30	14.71%	27.63	42.23	18.29%	37.86	56.01	26.72%
	FC-LSTM	21.42	33.37	15.32%	25.83	39.10	20.35%	36.41	50.73	29.92%
	DCRNN	20.34	31.94	13.65%	23.21	36.15	15.70%	29.24	44.81	20.09%
	STGCN	19.35	30.76	12.81%	21.85	34.43	14.13%	26.97	41.11	16.84%
	Graph WaveNet	18.15	29.24	12.27%	19.12	30.62	13.28%	20.69	33.02	14.11%
	ASTGCN	20.15	31.43	14.03%	22.09	34.34	15.47%	26.03	40.02	19.17%
	STSGCN	19.41	30.69	12.82%	21.83	34.33	14.54%	26.27	40.11	14.71%
	GMAN	18.28	29.32	12.35%	18.75	30.77	12.96%	19.95	30.21	12.97%
	MTGNN	18.22	30.13	12.47%	19.27	32.21	13.09%	20.93	34.49	14.02%
GTS	18.97	29.83	13.06%	19.29	30.85	13.92%	21.04	34.81	14.94%	
STEP	17.34*	28.44*	11.57%*	18.12*	29.81*	12.00%*	19.27*	31.33	12.78%*	

可解释性

理解TSFormer: 分析patch之间的余弦相似度

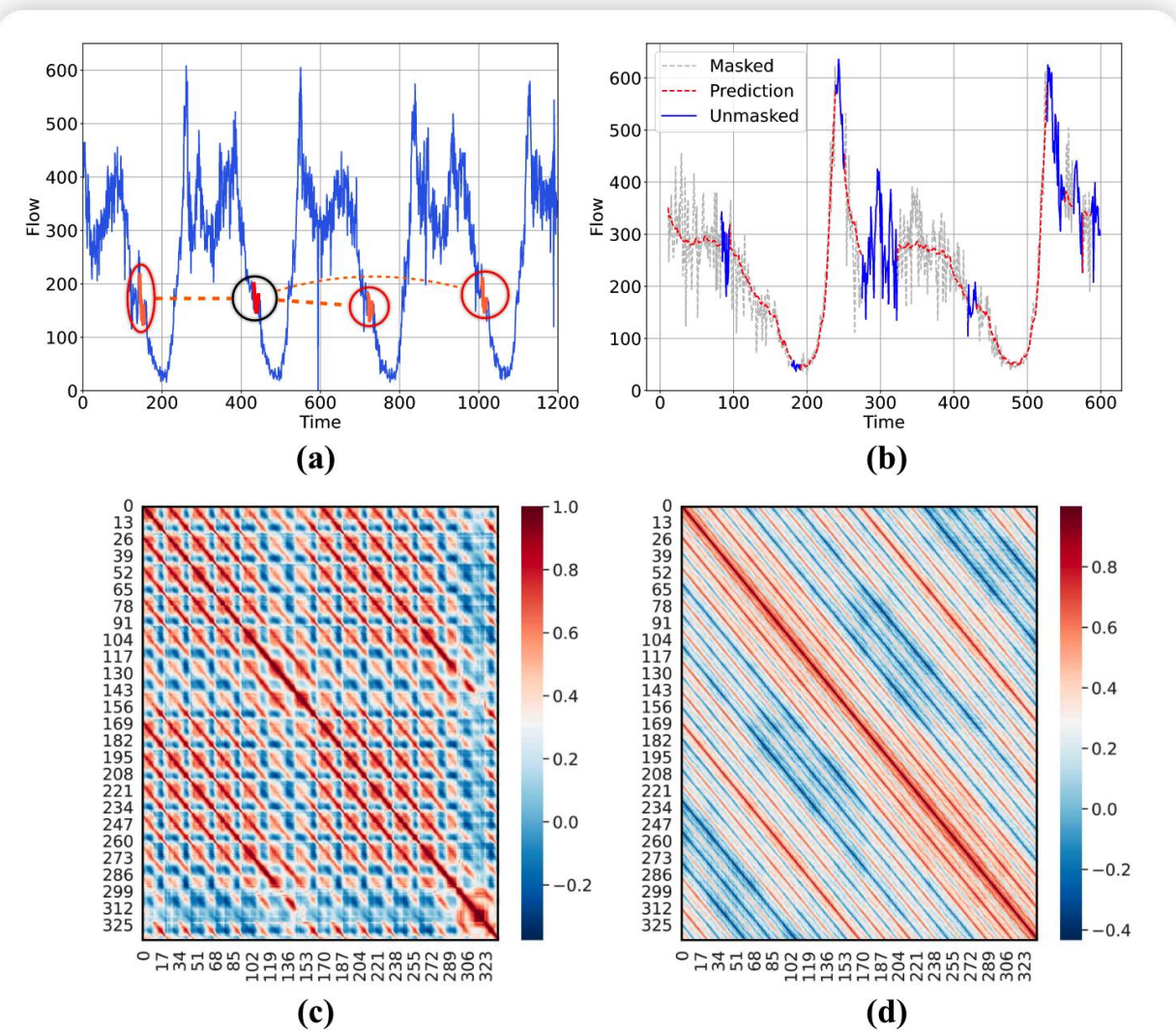
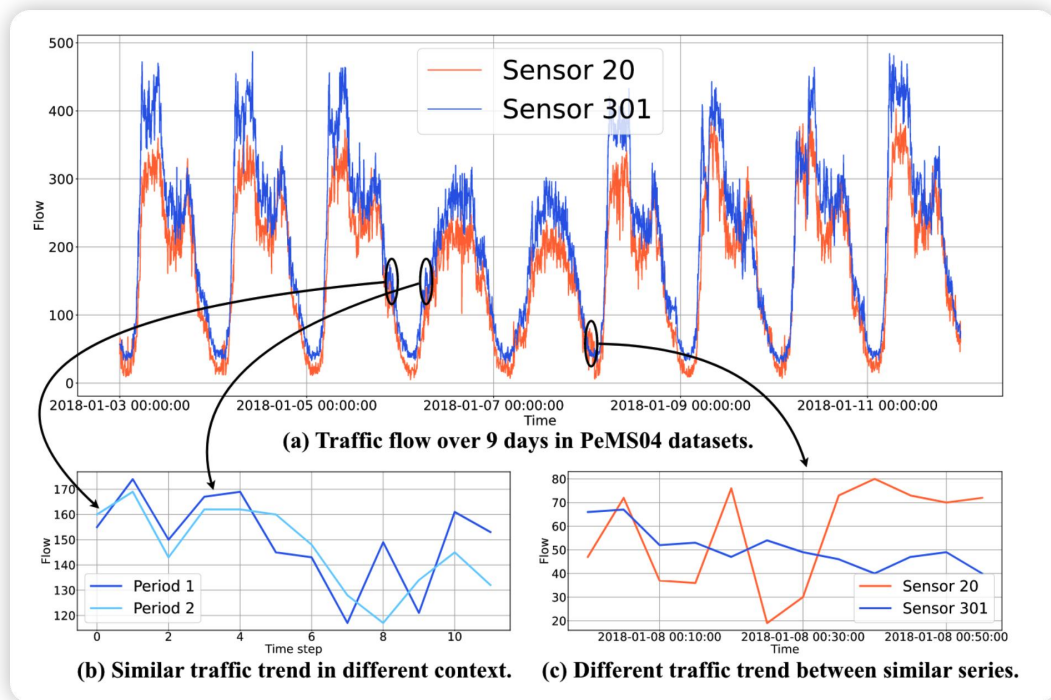


Figure 3: Inspecting the TSFormer. (a) Learned temporal periodicity. (b) Reconstruction. (c) Similarity of latent representations among different patches. (d) Similarity of positional embeddings among different patches.

感谢聆听