ICLR 2022 Oral

Pyraformer

Low-Complexity Pyramidal Attention for Long-Range Time Series Modeling and Forecasting

23.5.25

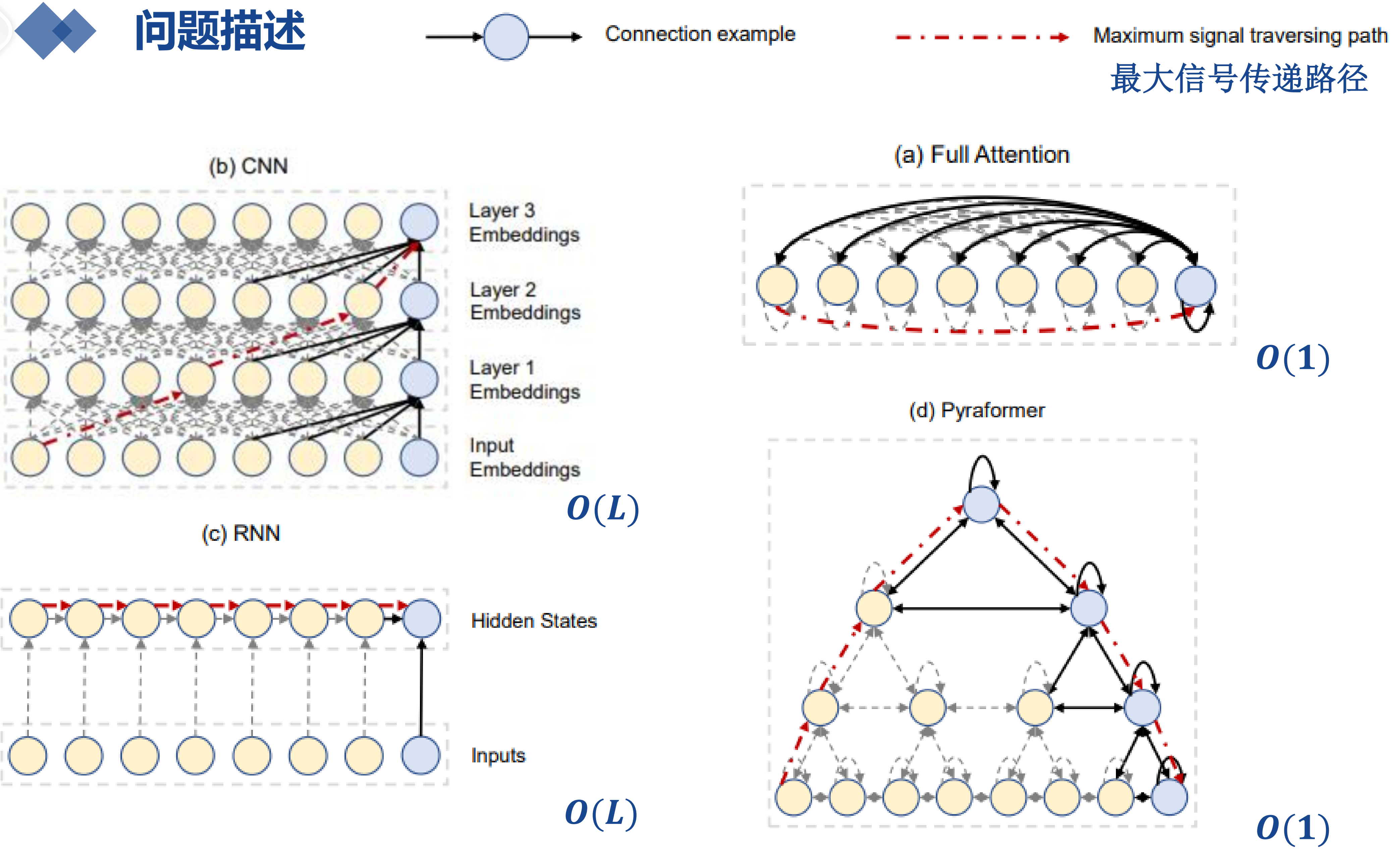
Presented by Yyyq

01 问题描述

- **时序预测任务的难点：**长时依赖问题
- **现有方法的缺陷：**复杂度较高

Method	Complexity per layer
CNN (Munir et al., 2018)	$\mathcal{O}(L)$
RNN (Salinas et al., 2020)	$\mathcal{O}(L)$
Full-Attention (Vaswani et al., 2017)	$\mathcal{O}(L^2)$
ETC (Ainslie et al., 2020)	$\mathcal{O}(GL)$
Longformer (Beltagy et al., 2020)	$\mathcal{O}(L)$
LogTrans (Li et al., 2019)	$\mathcal{O}(L \log L)$
Pyraformer	$\mathcal{O}(L)$

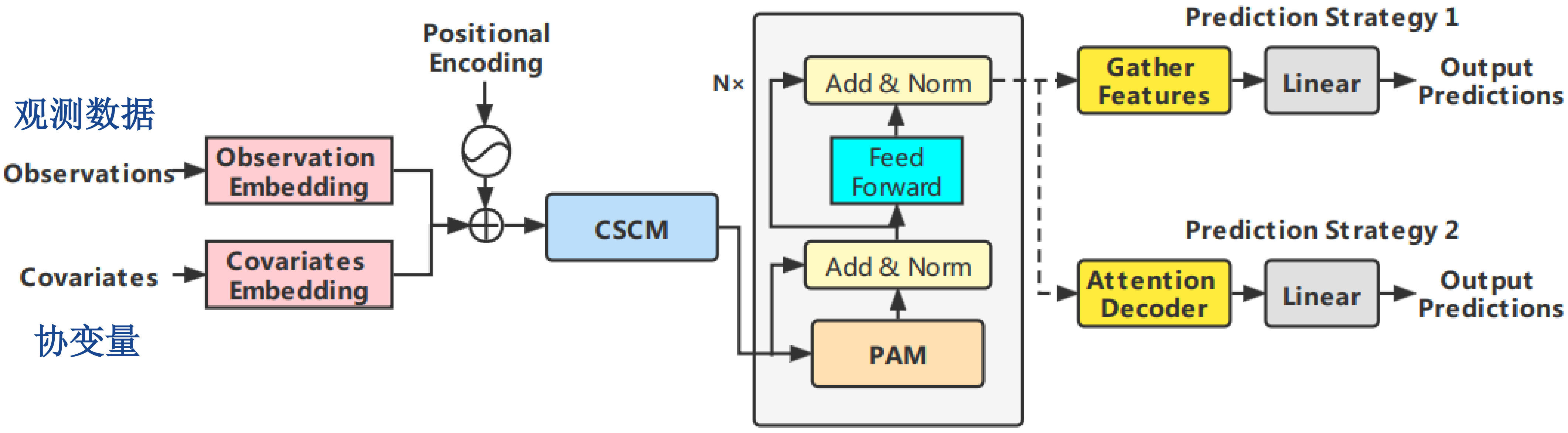
01 问题描述



Connection example
Maximum signal traversing path
最大信号传递路径
(b) CNN
Layer 3
Embeddings
Layer 2
Embeddings
Layer 1
Embeddings
Input
Embeddings
$O(L)$
(c) RNN
Hidden States
Inputs
$O(L)$
(a) Full Attention
$O(1)$
(d) Pyraformer
$O(1)$

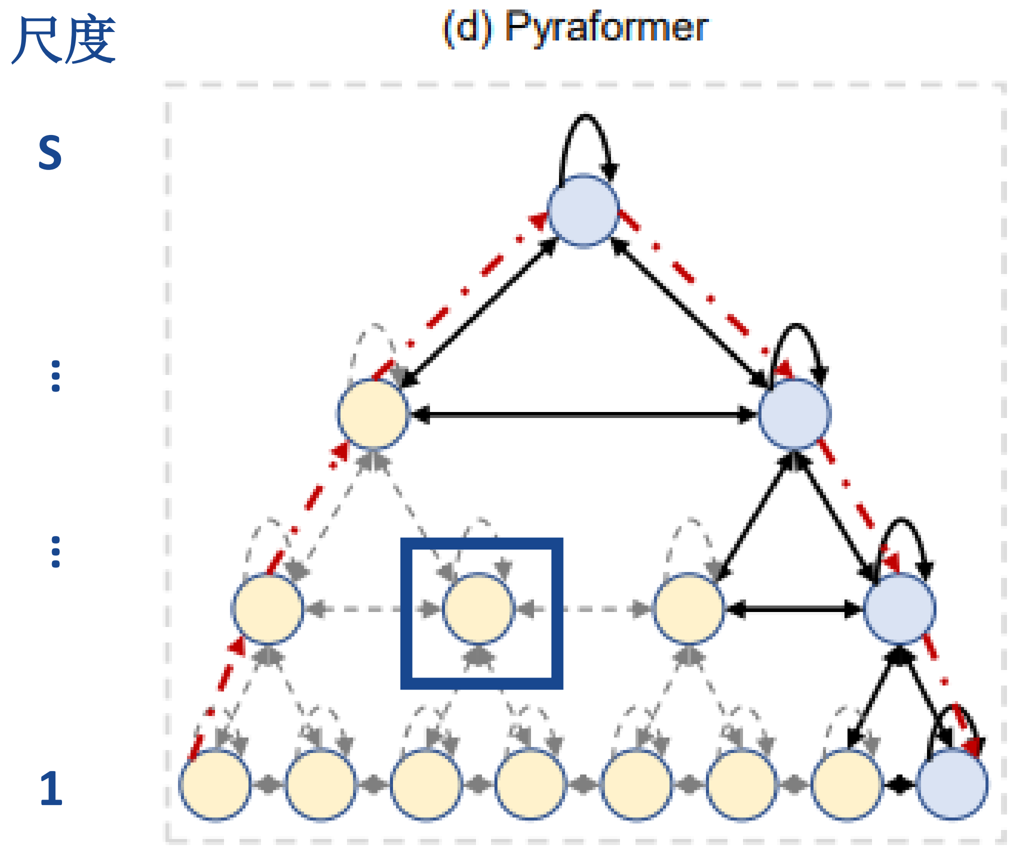
02

整体框架



观测数据
Observations
Observation Embedding
Covariates
Covariates Embedding
协变量
Positional Encoding
CSCM
N×
Add & Norm
Feed Forward
Add & Norm
PAM
Prediction Strategy 1
Gather Features
Linear
Output Predictions
Prediction Strategy 2
Attention Decoder
Linear
Output Predictions

03 算法描述：Pyramidal Attention Module (PAM)



$$\mathbb{N}_\ell^{(s)} \begin{cases} \mathbb{A}_\ell^{(s)} & \text{表示同一级尺度上的相邻节点（包括自己，A=3/5）} \\ \mathbb{C}_\ell^{(s)} & \text{表示C个子节点} \\ \mathbb{P}_\ell^{(s)} & \text{表示父节点} \end{cases}$$

一个节点最多可以连接 A + C + 1 个节点

03 算法描述：Pyramidal Attention Module (PAM)

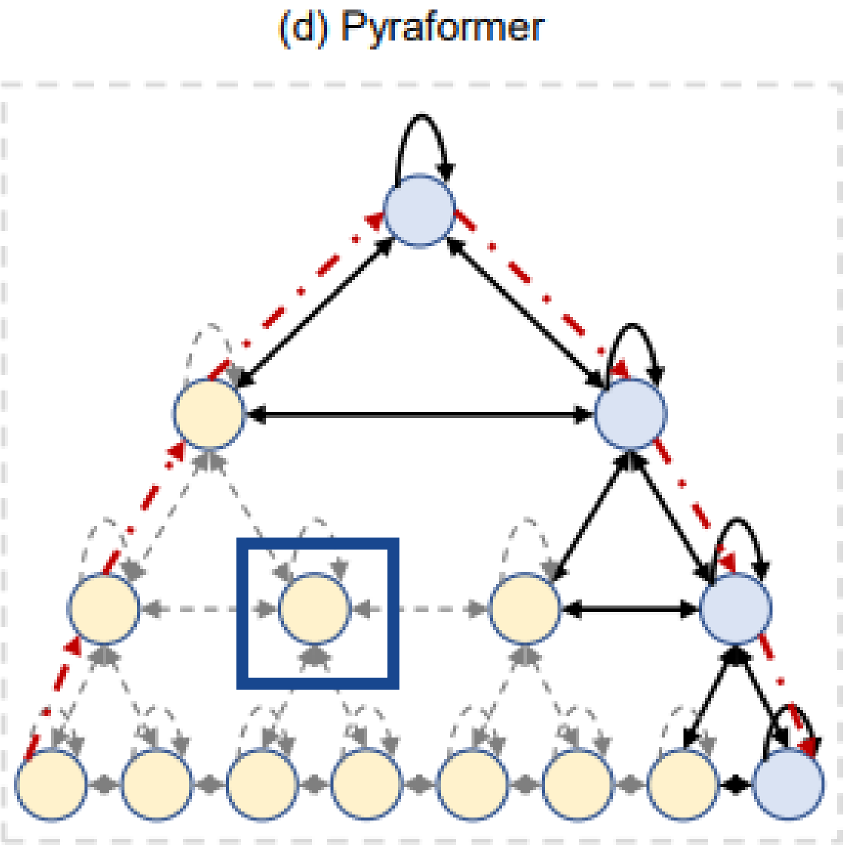
尺度

S

⋮

⋮

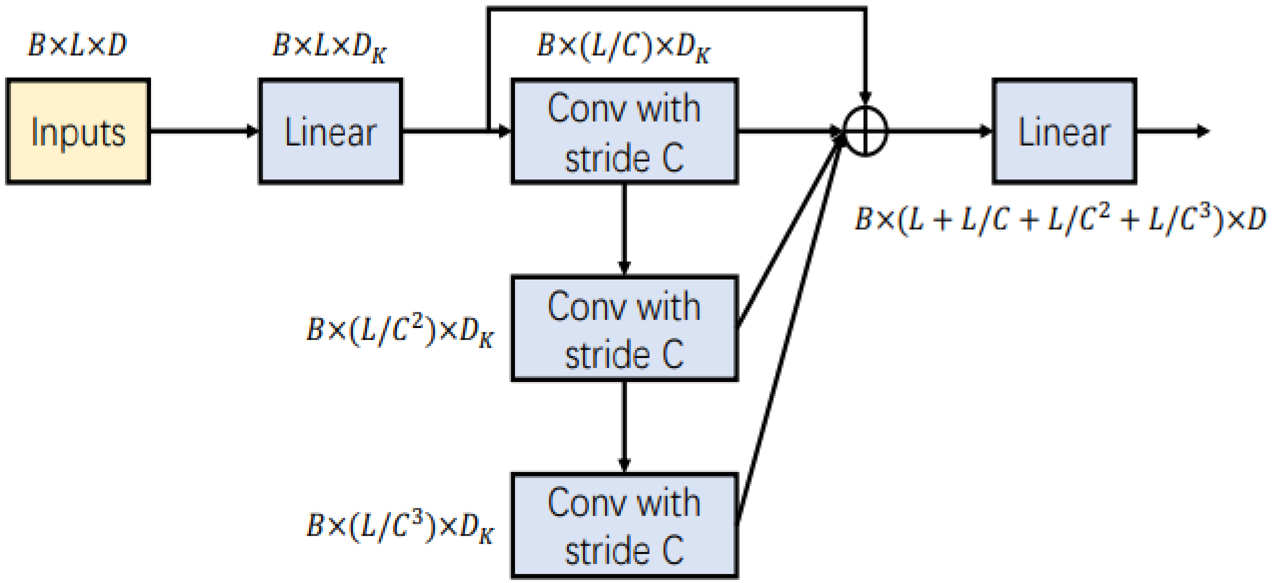
1



$$\boldsymbol{y}_i = \sum_{\ell=1}^{L} \frac{\exp(\boldsymbol{q}_i \boldsymbol{k}_\ell^T / \sqrt{D_K}) \boldsymbol{v}_\ell}{\sum_{\ell=1}^{L} \exp(\boldsymbol{q}_i \boldsymbol{k}_\ell^T / \sqrt{D_K})},$$

$$\boldsymbol{y}_i = \sum_{\ell \in \mathbb{N}_\ell^{(s)}} \frac{\exp(\boldsymbol{q}_i \boldsymbol{k}_\ell^T / \sqrt{d_K}) \boldsymbol{v}_\ell}{\sum_{\ell \in \mathbb{N}_l^{(s)}} \exp(\boldsymbol{q}_i \boldsymbol{k}_\ell^T / \sqrt{d_K})},$$

03 算法描述：Coarser-scale Construction Module (CSCM)



尺度为S，序列长度：$\frac{L}{C^S}$

DASFAA 2023 CCF-B

ISTNet

Inception Spatial Temporal Transformer for Traffic Prediction

23.5.25

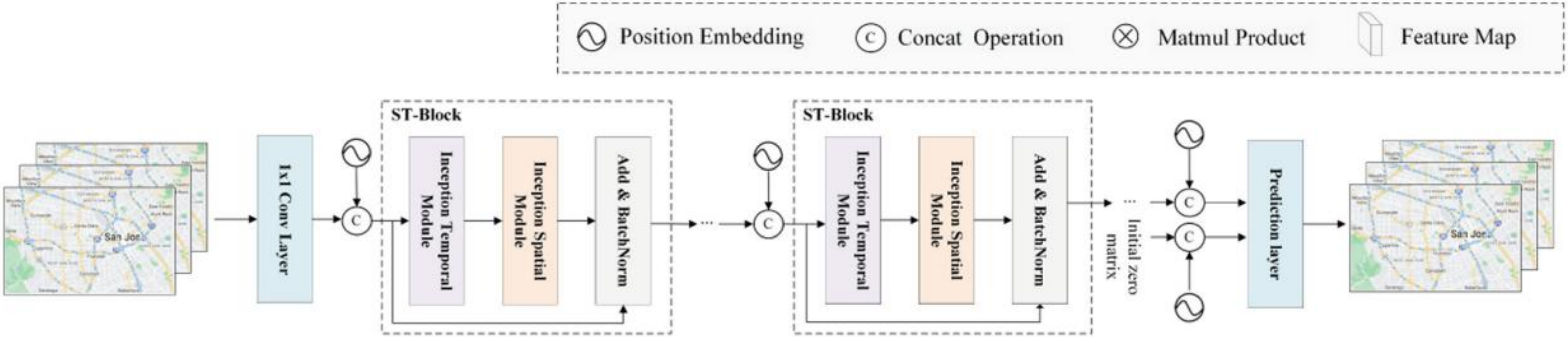
Presented by Yyyq

01

问题描述

- 时间相关性
  - 局部相关性　　CNN：需要堆叠多层，效率低
  - 全局依赖性　　注意力机制：偏好全局信息
- 空间相关性
  - 局部相关性
  - 潜在的全局相关性（语义相关）

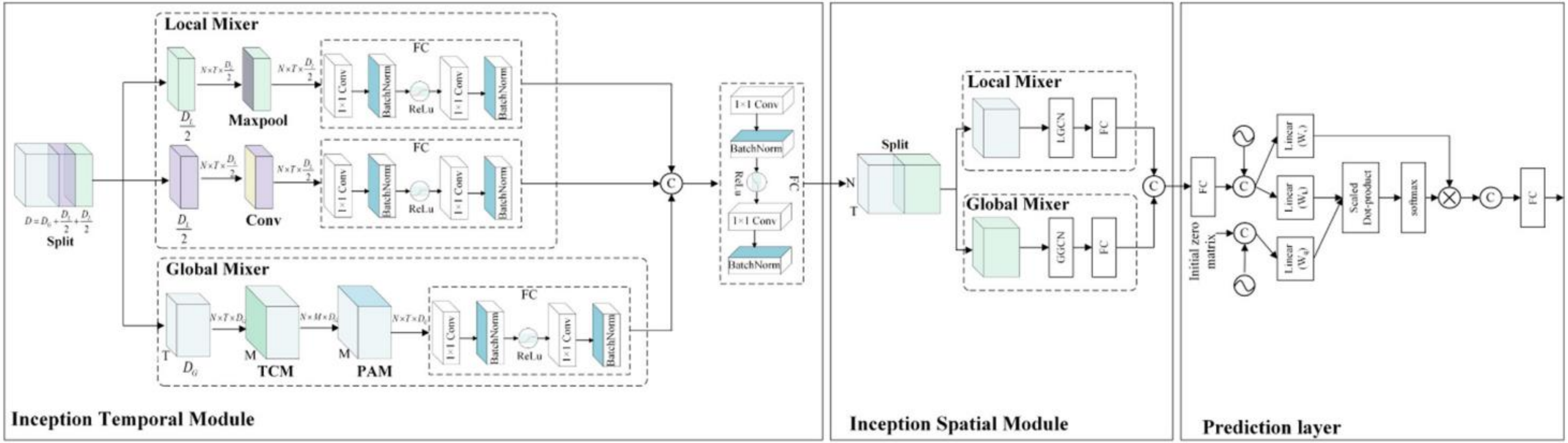
02 整体框架



(a) Overall architrcture of ISTNet

02

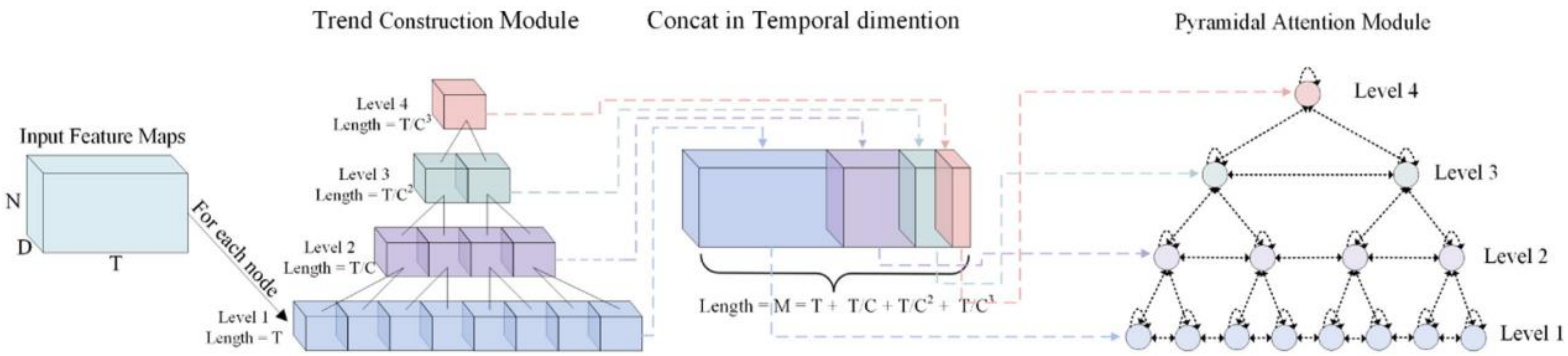
整体框架



Local Mixer
Maxpool
Conv
FC
ReLu
BatchNorm
1×1 Conv
Split
Global Mixer
TCM
PAM
Inception Temporal Module
Local Mixer
LGCN
Global Mixer
GGCN
Split
Inception Spatial Module
Initial zero matrix
Scaled Dot-product
softmax
Prediction layer

03

算法描述：Trend Construction Module (TCM)
Pyramidal Attention Module (PAM)



03 算法描述：Inception Spatial Module

➢ 局部混合器：高斯核函数

$$\mathcal{A}_{i,j} = exp(-\frac{dist(v_i, v_j)}{\mu^2})$$

$$dist(v_i, v_j) \leq \varepsilon \ , \ \mathcal{A}_{i,j} = 0.$$

$$\mathcal{Y}_L^S = FC(\mathcal{A}\mathcal{X}^{local}W_1 + b_1)$$

➢ 全局混合器：自适应邻接矩阵

$$\mathcal{M} = softmax(EE^T/\sqrt{D})$$

$$\begin{aligned} &for \quad i = 1, 2, \cdots, N \\ &\qquad nodeId = argtopk(\mathcal{M}[i, :]) \\ &\qquad \mathcal{M}[i, -nodeId] = 0 \end{aligned}$$

$$\mathcal{Y}_G^S = FC(\mathcal{M}\mathcal{X}^{global}W_2 + b_2)$$

03 算法描述：频率斜坡结构

- **ST-Block堆叠的第一层到第 l 层，平衡局部和全局组件**

底层更喜欢局部信息，而顶层在捕获全局信息方面起着更重要的作用。

通道比率：局部 / 全局 = [2, 1, 1/2]

04

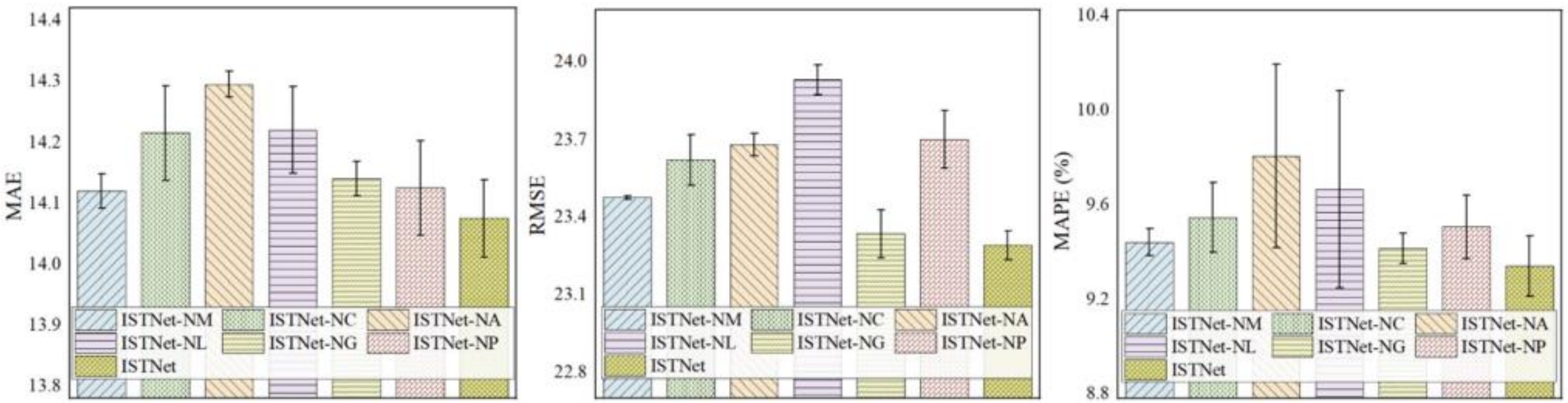
数据集

Datasets	Time Range	Time Steps	Time Interval	Nodes
PEMS03	09/01/2018 - 11/30/2018	26202	5-min	358
PEMS04	01/01/2018 - 02/28/2018	16992	5-min	307
PEMS07	05/01/2017 - 08/31/2017	28224	5-min	883
PEMS08	07/01/2016 - 08/31/2016	17856	5-min	170

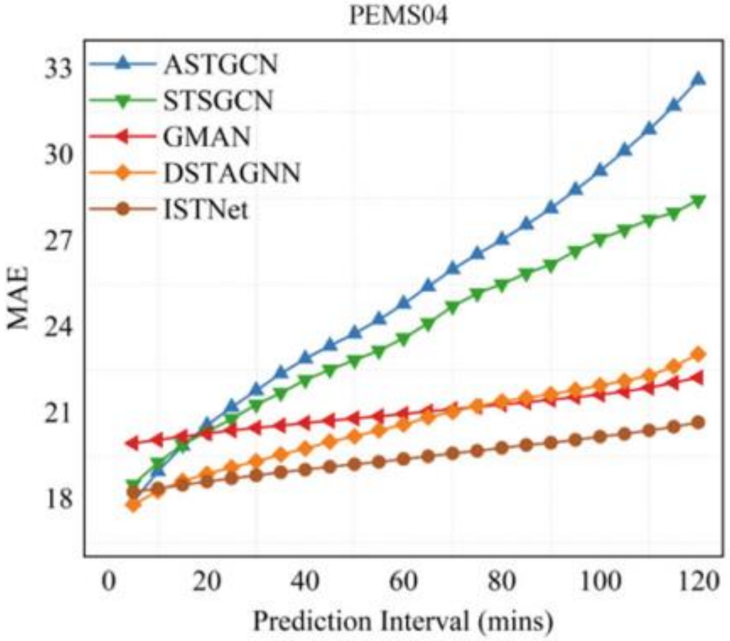
05

实验结果：预测未来一小时

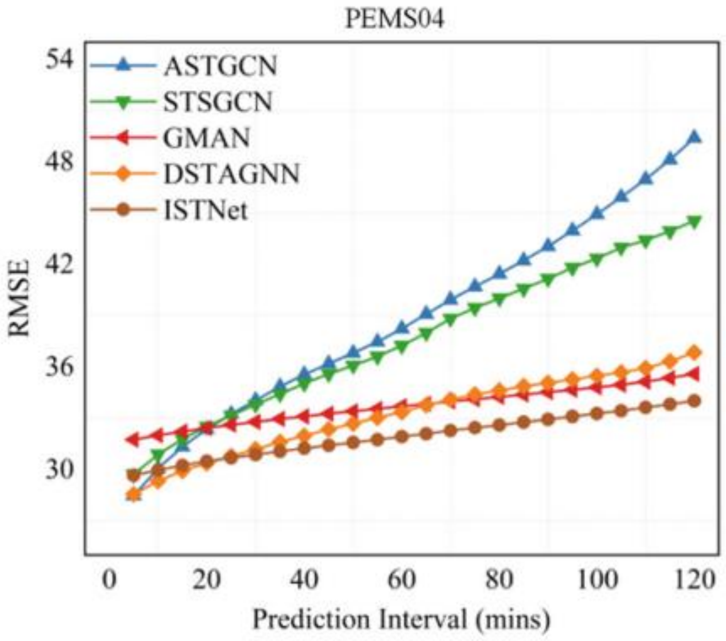
Dataset	Metrics	VAR	SVR	FC-LSTM	DCRNN	STGCN	ASTGCN	Graph WaveNet	STSGCN	GMAN	DSTAGNN	ISTNet
PEMS03	MAE	19.72	19.77	19.56	17.62	19.76	18.67	15.67	17.51	15.52	15.57	**15.03±0.09**
	RMSE	32.38	32.78	33.38	29.86	33.87	30.71	26.42	29.05	26.53	27.21	**24.89±0.25**
	MAPE(%)	20.50	23.04	19.56	16.83	17.33	19.85	15.72	16.92	15.19	**14.68**	15.24±0.19
PEMS04	MAE	24.44	26.18	23.60	24.42	23.90	22.90	19.91	21.52	19.25	19.30	**18.51±0.03**
	RMSE	37.76	38.91	37.11	37.48	36.43	33.59	31.06	34.14	30.85	31.46	**30.36±0.10**
	MAPE(%)	17.27	22.84	16.17	16.86	13.67	16.75	13.62	14.50	13.00	12.70	**12.36±0.16**
PEMS07	MAE	27.96	28.45	34.05	24.45	26.22	28.13	20.83	23.99	20.68	21.42	**19.67±0.13**
	RMSE	41.31	42.67	55.70	37.61	39.18	43.67	33.62	39.32	33.56	34.51	**32.96±0.04**
	MAPE(%)	12.11	14.00	15.31	10.67	10.74	13.31	9.10	10.10	9.31	9.01	**8.57±0.20**
PEMS08	MAE	19.83	20.92	21.18	18.49	18.79	18.72	15.57	17.88	14.87	15.67	**14.08±0.05**
	RMSE	29.24	31.23	31.88	27.30	28.2	28.99	24.32	27.36	24.06	24.77	**23.27±0.12**
	MAPE(%)	13.08	14.24	13.72	11.69	10.55	12.53	10.32	11.71	9.77	9.94	**9.34±0.09**



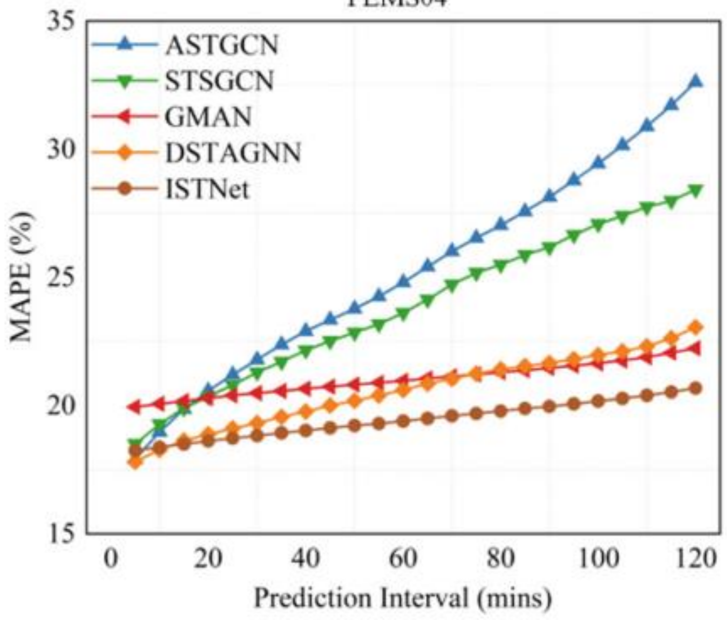
05 实验结果2



(a) MAE per prediction step



(b) RMSE per prediction step



(b) MAPE per prediction step

谢谢观看

MANY THANKS !

23.5.25