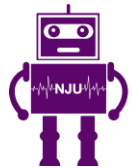




南京大學
NANJING UNIVERSITY



RINC

Robotic Intelligence & Neural Computing Group

基于空洞卷积和空间注意力的 长短期时序预测研究

答辩人：李菲菲 MF21330047

导师：申富饶 教授

日期：2024年5月16日

誠樸雄偉 勵學敦行

1 研究背景

2 研究内容

- 基于浅层空洞卷积的长期时间序列预测算法
- 基于稀疏动态图的短期时空预测算法

3 实际应用

- 交通流量可视化与预测系统

4 研究生期间工作成果

5 工作总结

目录

第一部分

研究背景

Research Background

1.1 时间序列预测

时间序列：利用传感器按照时间顺序进行采样得到的数据点序列

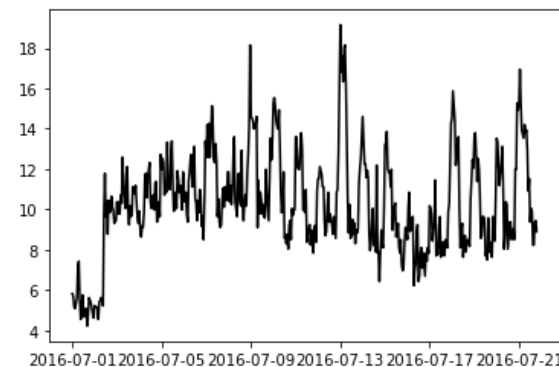
特点：

1. 时间相关性：

不同时间点的观测值之间具有相关性，**周期性、趋势性等**

2. 空间相关性：（如交通流量时序数据）

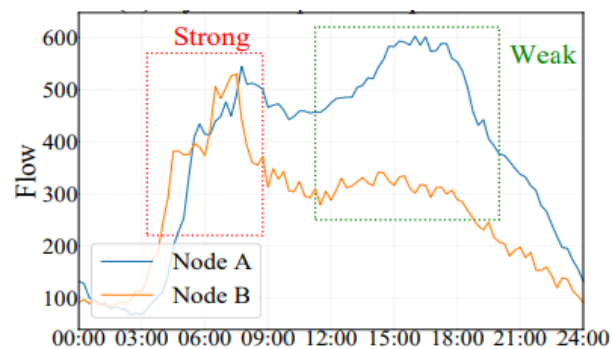
地理空间中每个采集点通过其位置相互关联，形成**空间关系网络**。空间关系网络分为**预定义图（基于地理距离构建）**和**动态图**



时间序列数据



静态地理空间图



关系强弱随时间动态变化

1.1 时间序列预测

时间序列预测：利用过去一段时间的数据来预测未来一段时间内的信息

一天

短期预测

长期预测

- 预测未来几分钟、几小时内的趋势
- 更注重预测的短时精确性和实时性
- 空间关系+时间关系

- 预测未来几天、几周内的趋势
- 更注重趋势和模式的识别，空间关系影响小
- 时间关系

研究问题

1. 针对不具有空间关系的数据，如何有效地进行长期时间序列预测
2. 对于具有显式空间关系的数据，如何进行短期时空序列预测，以充分利用空间信息来提高预测的准确性和效率

1.1 研究意义

广泛的应用场景



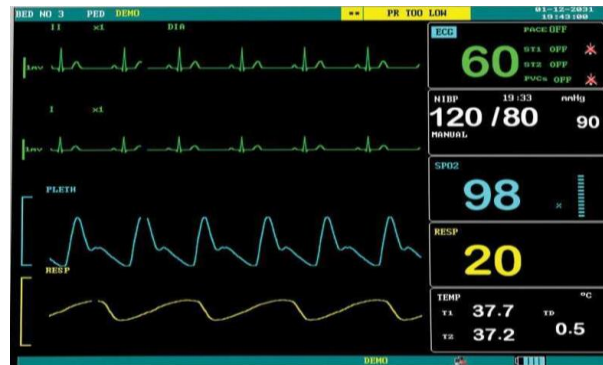
能源消耗预测



股票价格预测



交通流量预测



心率预测

为短期和长远的决策提供科学、及时且有效的数据支持

1.3

研究挑战

挑战:

- 建模长期时间依赖时，现有方法难以兼顾序列局部波动和全局趋势形态的学习
- 现有方法具有高时间复杂度和高空间复杂度，无法广泛应用于实时响应应用场景
- 基于距离等构建的预定义图是静态的，难以满足空间关系动态变化
- 构建动态图时，现有方法生成的动态图过密，存在冗余连接

一天

短期预测

长期预测

工作一:

基于浅层空洞卷积的长期时间序列预测算法
MSDCN

工作二:

基于稀疏动态图的短期时空预测算法
ThermoNet

第二部分

研究内容

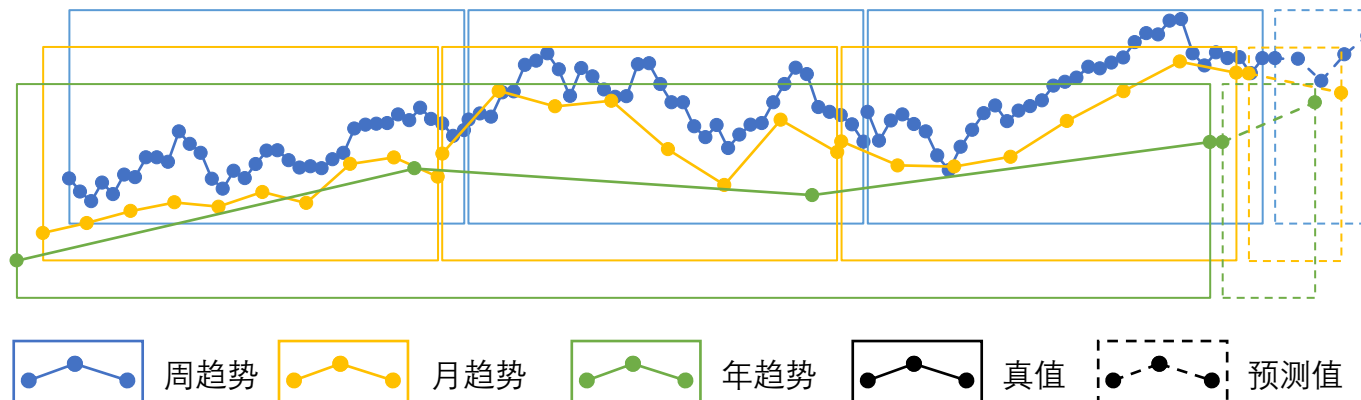
Research Content

基于浅层空洞卷积的长期时间序列预测算法 MSDCN

基于稀疏动态图的短期时空预测算法 ThermoNet

2.1 长期时序预测算法 MSDCN

研究动机



时间序列可以被视为不同尺度成分（周、月、年等）的叠加

↓
 现有方法不足

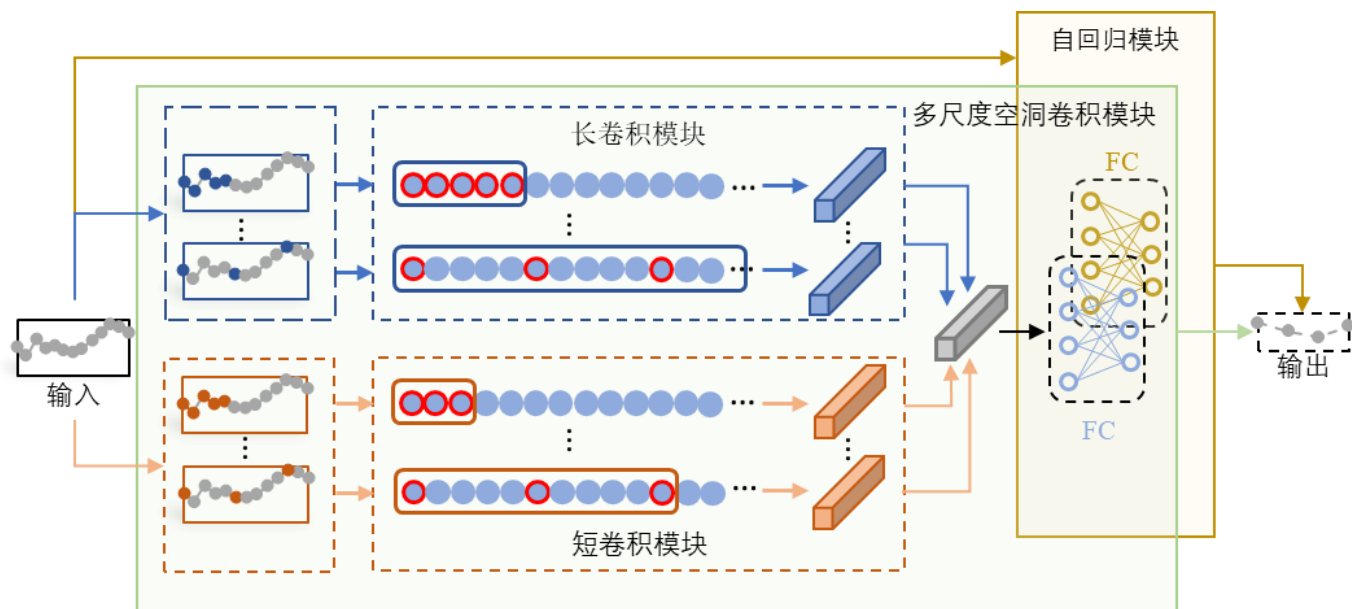
- 单一尺寸的卷积核
→ 无法提取多尺度信息
- 多尺度普通卷积核
→ 模型参数量大；训练和推理时间慢

↓
 思路

- ✓ 多尺度的空洞卷积核，来捕获局部波动和全局趋势形态
- ✓ 利用空洞卷积的扩张特性，在不增加参数的情况下，捕获更长的序列信息

2.1 长期时序预测算法 MSDCN

■ 整体结构



- 多尺度空洞卷积模块 Convolution Module
 - 多尺度卷积核：不同卷积核大小（长/短卷积模块）、不同空洞数（长/短卷积模块内部）
 - 加权融合不同卷积核得到的特征表示

$$H = [h_1, h_2, \dots, h_{n+m}]$$

$$M = \sum_i^{n+m} H_i \odot W_i'$$

- 自回归模块 Autoregressive Module

- 传统自回归数学公式

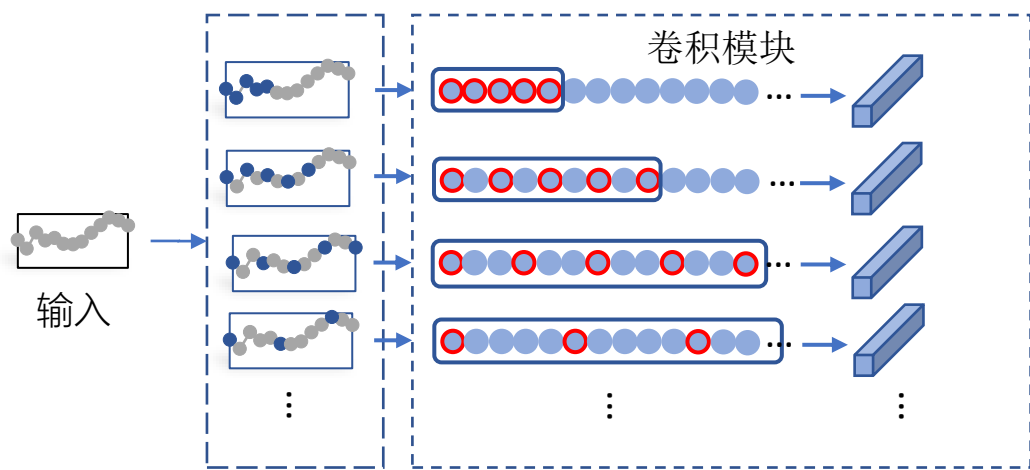
$$Y_t = \sum_{i=1}^p \phi_i Y_{i-1} + \epsilon_t$$

- 自回归模块

$$\hat{Y}_h = W_2 X + b_2$$

2.1 长期时序预测算法 MSDCN

■ 关键细节



- 每一个一维卷积是一个分通道的空洞卷积

$$h_i = \text{ConvBlock}_i(X)$$

$$u_i = \text{DilatedConv1d}(\text{Padding}(X))$$

$$h_i = \text{ReLU}(\text{BatchNorm1d}(u_i))$$

- 不同的一维卷积并列，空洞数以近似2的倍数增长：

$$\text{dilation factors} = \{1, 2^0 + 1, 2^1 + 1, \dots, 2^K + 1\}$$

2.1 长期时序预测算法 MSDCN

■ 对比实验

表2-1-1 公共数据集MSDCN实验结果

模型 指标		MSDCN (Ours)		TimesNet		Dlinear		ETSformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.174	0.256	<u>0.187</u>	<u>0.267</u>	0.193	0.292	0.189	0.280
	192	0.239	0.298	<u>0.249</u>	<u>0.309</u>	0.284	0.362	0.253	0.319
	336	0.296	0.332	0.321	<u>0.351</u>	0.369	0.427	<u>0.314</u>	0.357
	720	0.395	0.392	<u>0.408</u>	<u>0.403</u>	0.554	0.522	0.414	0.413
ETTh2	96	0.286	0.333	0.340	<u>0.374</u>	<u>0.333</u>	0.387	0.340	0.391
	192	0.364	0.385	<u>0.402</u>	<u>0.414</u>	0.477	0.476	0.430	0.439
	336	0.407	0.420	<u>0.452</u>	<u>0.452</u>	0.594	0.541	0.485	0.479
	720	0.418	0.438	<u>0.462</u>	<u>0.468</u>	0.831	0.657	0.500	0.497
Electricity	96	<u>0.175</u>	0.265	0.168	<u>0.272</u>	0.197	0.282	0.187	0.304
	192	0.183	0.271	<u>0.184</u>	0.289	0.196	<u>0.285</u>	0.199	0.315
	336	<u>0.199</u>	0.287	0.198	<u>0.300</u>	0.209	0.301	0.212	0.329
	720	0.238	0.320	0.220	<u>0.320</u>	0.245	0.333	<u>0.233</u>	0.345
Weather	96	0.169	0.217	<u>0.172</u>	<u>0.220</u>	0.196	0.255	0.197	0.281
	192	0.215	0.259	<u>0.219</u>	<u>0.261</u>	0.237	0.296	0.237	0.312
	336	0.269	0.299	<u>0.280</u>	<u>0.306</u>	0.283	0.335	0.298	0.353
	720	<u>0.350</u>	0.352	0.365	<u>0.359</u>	0.345	0.381	0.352	0.388

- 输入序列长度：96；输出序列长度：96-720
(以ETTm2为例，输入历史1h数据，输出未来1h-7h数据)
- 评估指标：均方误差 (MSE)、平均绝对误差 (MAE)
- 在多个公共数据集上，我们的方法性能指标平均提升了3%
 - 我们的方法采用的浅层空洞卷积结构相较于TimesNet中使用的深层二维普通卷积结构，在预测性能上具有优势
 - 不依赖于复杂的Transformer架构，我们的方法也能够有效地处理长序列

2.1 长期时序预测算法 MSDCN

■ 对比实验

表2-1-2 MSDCN与主流模型的时间效率和空间效率对比结果

模型	参数量	内存	乘累加操作数	推理时间
MSDCN (Ours)	239.67K	19.16MB	8.75M	3.68s
Dlinear	139.68K	13.48MB	0.14M	1.50s
Autoformer	14.91M	126.33MB	4.18G	62.39s
PatchTST	7.59M	800.96MB	12.69G	80.42s
TimesNet	301.75M	1628.31MB	1.23T	1.49min

- 我们的方法在时间效率和空间效率上仅次于只包含少量池化层和线性层的Dlinear模型
- 与预测结果最优的基于普通卷积结构的TimesNet模型相比，我们的模型参数量约为其 $\frac{1}{1000}$ ，推理时间约为其 $\frac{1}{24}$ ，在推理速度上具有显著优势

2.1 长期时序预测算法 MSDCN

■ 对比实验

表2-1-2 MSDCN与主流模型的时间效率和空间效率对比结果

模型	参数量	内存	乘累加操作数	推理时间
MSDCN (Ours)	239.67K	19.16MB	8.75M	3.68s
Dlinear	139.68K	13.48MB	0.14M	1.50s
Autoformer	14.91M	126.33MB	4.18G	62.39s
PatchTST	7.59M	800.96MB	12.69G	80.42s
TimesNet	301.75M	1628.31MB	1.23T	1.49min

- 我们的方法在时间效率和空间效率上仅次于只包含少量池化层和线性层的Dlinear模型
- 与预测结果最优的基于普通卷积结构的TimesNet模型相比，我们的模型参数量约为其 $\frac{1}{1000}$ ，推理时间约为其 $\frac{1}{24}$ ，在推理速度上具有显著优势

2.1 长期时序预测算法 MSDCN

■ 消融实验

表2-1-3 不同尺度卷积模块对预测结构的影响

卷积模块		评估指标	ETTm2				Electricity			
长	短		96	192	336	720	96	192	336	720
×	×	MSE	0.182	0.246	0.306	0.408	0.198	0.198	0.212	0.254
		MAE	0.265	0.304	0.341	0.398	0.275	0.278	0.293	0.326
√	×	MSE	0.176	0.239	0.300	0.398	0.176	0.184	0.200	0.241
		MAE	0.257	0.299	0.335	0.392	0.265	0.272	0.287	0.321
×	√	MSE	0.177	0.242	0.302	0.407	0.194	0.193	0.208	0.250
		MAE	0.256	0.300	0.336	0.399	0.271	0.274	0.289	0.322
√	√	MSE	0.174	0.239	0.296	0.395	0.175	0.183	0.199	0.239
		MAE	0.256	0.298	0.332	0.392	0.265	0.271	0.287	0.320

- 分别引入长卷积模块和短卷积模块都可以带来预测性能的提升
- 同时引入后, 模型预测性能平均提升4.03%, 达到最佳水平

4.03%↑

表2-1-4 自回归模块对预测结果的影响

模型		自回归模块		卷积模块		MSDCN (自回归模块+卷积模块)	
评估指标		MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.198	0.275	0.187	0.279	0.175	0.265
	192	0.198	0.278	0.198	0.286	0.183	0.271
	336	0.212	0.293	0.214	0.302	0.199	0.287
	720	0.254	0.326	0.253	0.332	0.238	0.320

- 自回归模块的加入可以增强模型表达能力, 预测精度提高4.71%

4.71%↑

2.1 长期时序预测算法 MSDCN

■ 分析实验

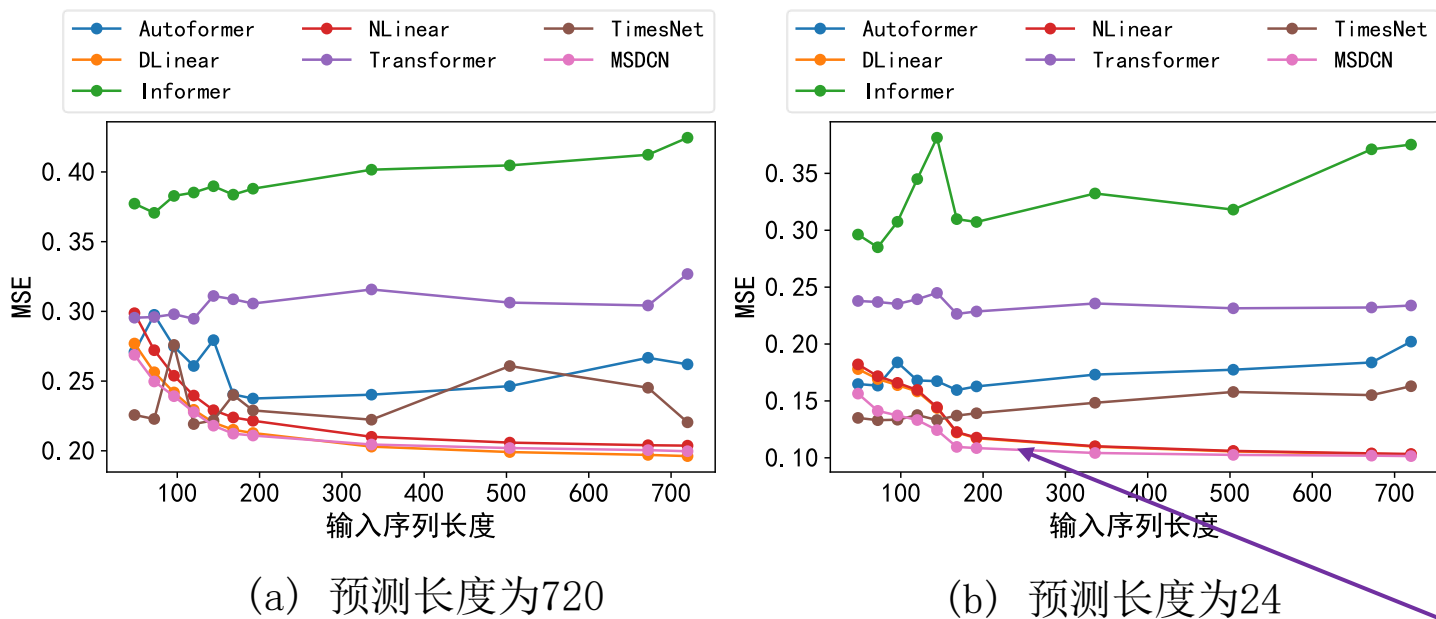


图2-1-1 不同输入长度下的预测性能比较

- 实验设置：固定输出序列长度，增加输入序列长度
- 基于Transformer的模型Autoformer、Informer和基于普通卷积的模型TimesNet模型预测均方误差逐渐上升，预测性能逐渐下降
- 我们的方法能够在较长的输入序列下保持好的预测性能，这说明我们的方法能更好地有效提取输入序列的依赖信息

2.2 短期时空预测 ThermoNet

■ 研究动机

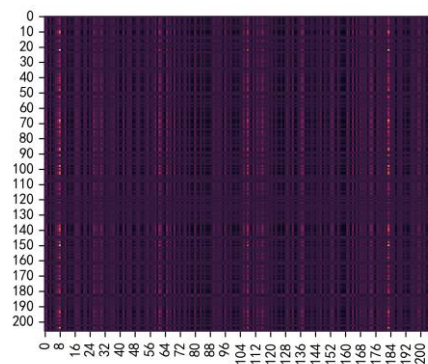
□ 预定义图难以满足空间关系动态变化

- 预定义空间图采用基于距离的权重分配方法，以量化地理空间中的相互作用

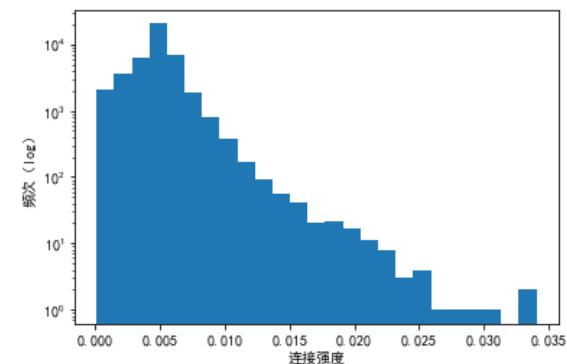
$$A_{ij} = \begin{cases} \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{2\sigma^2}\right), & \text{如果 } \text{dist}(v_i, v_j) \leq k, \\ 0 & \text{否则} \end{cases}$$

- 地理位置相距较远，但功能相似的区域可能展现出强烈的空间依赖性

□ 动态图难以区分关系强弱



(a) 热力图



(b) 对应的分布直方图

图2-2-1 现有方法生成动态图的分布

- 现有方法生成的动态图所有节点间都存在连接关系
- 过多的弱连接稀释了强相关性连接的影响力。

对动态图进行稀疏化，减少弱连接，增强强连接
 联合预定义图和动态图二者空间信息

2.2 短期时空预测 ThermoNet

■ 整体结构

② 编码层

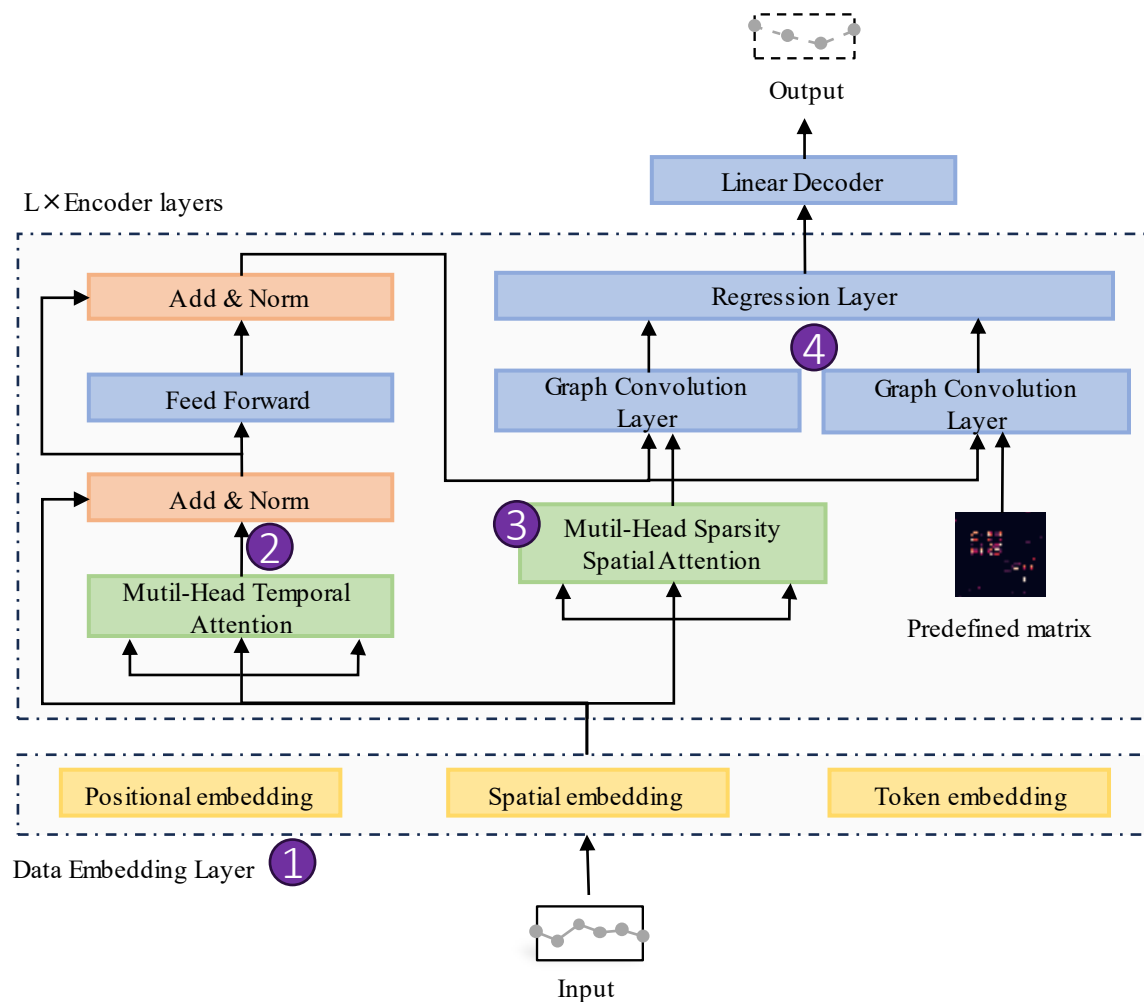
利用原始Transformer的编码层提取时间依赖

① 数据嵌入层

位置嵌入：序列相对位置信息

空间嵌入：空间信息

标识嵌入：局部信息



④ 图卷积层

对预定义图和动态图提取空间依赖关系

③ 生成动态空间图

多头稀疏空间注意力机制

2.2 短期时空预测 ThermoNet

■ 关键细节

空间注意力矩阵 $M_n \in \mathbb{R}^{N \times N}$

$$M_n = \sigma' \left(\frac{Q_n K_n^T}{\sqrt{D_k}} \right)$$

• 稀疏化操作2——硬阈值函数

✓ 消除低于设定阈值的注意力权重

✓ 阈值根据均匀分布值设置, 即 $\frac{1}{\text{节点数}}$

$$A_s = g(M_n, \tau) = \begin{cases} 0, & \text{如果 } M_n(i, j) < \tau, \\ M_n(i, j), & \text{否则.} \end{cases}$$

• 稀疏化操作1——Softmax函数变体

✓ 修改原始Softmax函数, 引入 β , 改变注意力分布

✓ 让重要位置的注意力分数更大, 不重要位置的注意力分数更小

$$\sigma'(z) = \frac{e^{\beta z}}{\sum_{i=1}^n e^{\beta z}}$$

✓ β 取值根据训练过程中生成图是否达到稀疏度要求来进行动态调整

2.2 短期时空预测 ThermoNet

■ 对比实验

表2-2-1 公共数据集ThermoNet实验结果1

数据集	指标	STGCN	DCRNN	GMAN	STTN	GTS	STNorm	ThermoNet
PEMS04	MAE	21.76	22.74	19.14	19.48	20.96	<u>18.96</u>	18.86
	RMSE	34.77	36.58	31.60	31.91	32.95	<u>30.98</u>	28.72
	MAPE	13.87	14.75	13.19	13.63	14.66	12.69	<u>13.10</u>
PEMS08	MAE	17.84	18.19	<u>15.31</u>	15.48	16.49	15.41	14.90
	RMSE	27.12	28.18	<u>24.92</u>	24.97	26.08	24.77	23.03
	MAPE	11.21	11.24	10.13	10.34	10.54	<u>9.76</u>	9.45

表2-2-2 公共数据集ThermoNet实验结果2

数据集	预测范围	指标	STGCN	DCRNN	GMAN	STTN	GTS	STNorm	ThermoNet
METR-LA	15min	MAE	2.88	2.77	2.80	2.79	<u>2.75</u>	2.81	2.70
		RMSE	5.74	5.38	5.55	5.48	<u>5.27</u>	5.57	5.14
		MAPE	7.60	7.30	7.41	7.19	<u>7.12</u>	7.40	6.93
	30min	MAE	3.47	3.15	<u>3.12</u>	3.16	3.14	3.18	3.11
		RMSE	7.24	6.45	6.49	6.50	<u>6.33</u>	6.59	6.17
		MAPE	9.60	8.80	8.73	8.53	<u>8.62</u>	8.47	<u>8.53</u>
PEMS-BAY	15min	MAE	1.36	1.38	1.34	1.36	1.37	<u>1.33</u>	1.32
		RMSE	2.96	2.95	2.82	2.87	2.92	<u>2.82</u>	2.80
		MAPE	2.90	2.90	2.81	2.89	2.85	2.76	<u>2.81</u>
	30min	MAE	1.81	1.74	1.62	1.67	1.72	1.65	<u>1.65</u>
		RMSE	4.27	3.97	3.72	3.79	3.86	3.77	<u>3.75</u>
		MAPE	4.20	3.90	3.63	3.78	3.88	<u>3.66</u>	3.73

- **数据集**：四个公共交通数据集
PEMS04、PEMS08、METR-LA、PEMS-BAY
- **指标**：平均绝对误差(MAE)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)
- 在预定义图极为稀疏的PEMS04和PEMS08数据集上，我们的方法取得了预测性能**4.14%的提升**
- 在预定义图较为密集的METR-LA和PEMS-BAY数据集上，我们的方法也**与其他方法相当**

2.2 短期时空预测 ThermoNet

■ 消融实验

表2-2-3 时间模块和空间模块对预测结果的影响

预测范围	15min			30min			60min		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
w/o M_T	2.86	5.31	7.53	3.26	6.27	8.99	3.76	7.44	10.64
w/o M_S	2.99	5.94	7.96	3.62	7.33	10.31	4.54	9.16	13.80
w/o A_g	2.75	5.24	7.20	3.12	6.20	8.63	3.55	7.24	10.28
w/o A_s	2.81	5.35	7.35	3.27	6.55	9.14	3.86	7.86	11.35
ThermoNet	2.70	5.14	6.93	3.11	6.17	8.53	3.56	7.24	10.23

- 仅保留时间依赖模块 (w/o M_S) 的预测效果不如仅保留空间依赖模块 (w/o M_T) 的情况, 空间依赖对短期预测问题非常重要
- 移除任何一个图 (w/o A_g 、w/o A_s) 均会导致预测准确性的降低, 且动态图的移除 (w/o A_s) 对性能的影响更大
- 联合预定义图 A_g 和动态图 A_s 取得最佳性能

2.2 短期时空预测 ThermoNet

■ 消融实验

表2-2-4 不同Embedding对预测结果的影响

Embedding			PEMS04		
Positional	Spatial	Token	MAE	RMSE	MAPE
✓			20.41	30.61	14.44
	✓	✓	19.23	29.26	14.11
✓	✓		19.12	29.05	12.94
✓	✓	✓	18.86	28.72	13.10

} 1.2%↑

表2-2-5 不同稀疏操作对预测结果的影响

操作		PEMS04		
硬阈值函数	Softmax变体	MAE	RMSE	MAPE
		18.95	29.07	13.12
✓		19.05	29.06	13.86
	✓	19.11	29.29	13.15
✓	✓	18.86	28.72	13.10

- 考虑数据的序列顺序性以及节点的空间表示是非常重要的
- 引入标识嵌入来提取关键的局部序列依赖信息，性能提升1.2%
- 三种嵌入技术联合使用，效果最佳
- 同时施加两种稀疏操作，效果最佳
- 采用 Softmax 函数变体会使得连接强度逐渐偏向“两极化”，让不重要的相关连接值更小，此时利用硬阈值函数对这种不重要的值进行裁剪。能够弱化弱连接，强化强连接

第三部分

实际应用

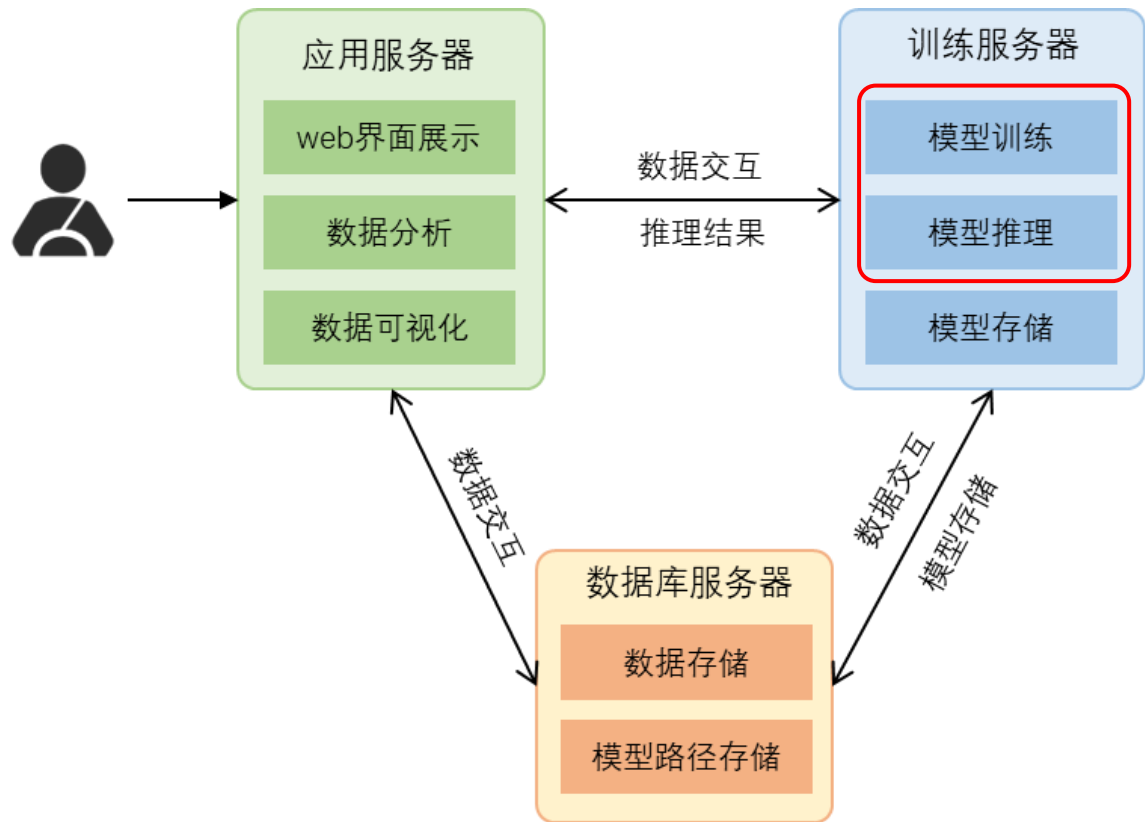
Applications

交通流量可视化与预测系统

3.1 系统概述

系统需求：

- 实时交通流量数据分析
- 未来交通流量数据预测
 - 长期交通流量预测
 - 短期交通流量预测
- 未来出行细粒度建议
 - 各路口通行建议



上述两个算法模型

3.2 系统展示

历史交通流量查询



历史流量查询

输入查询时间、区域

城市
江苏省南京市

区域
仙林大道

起始日期
2024/03/15

结束日期
2024/03/15

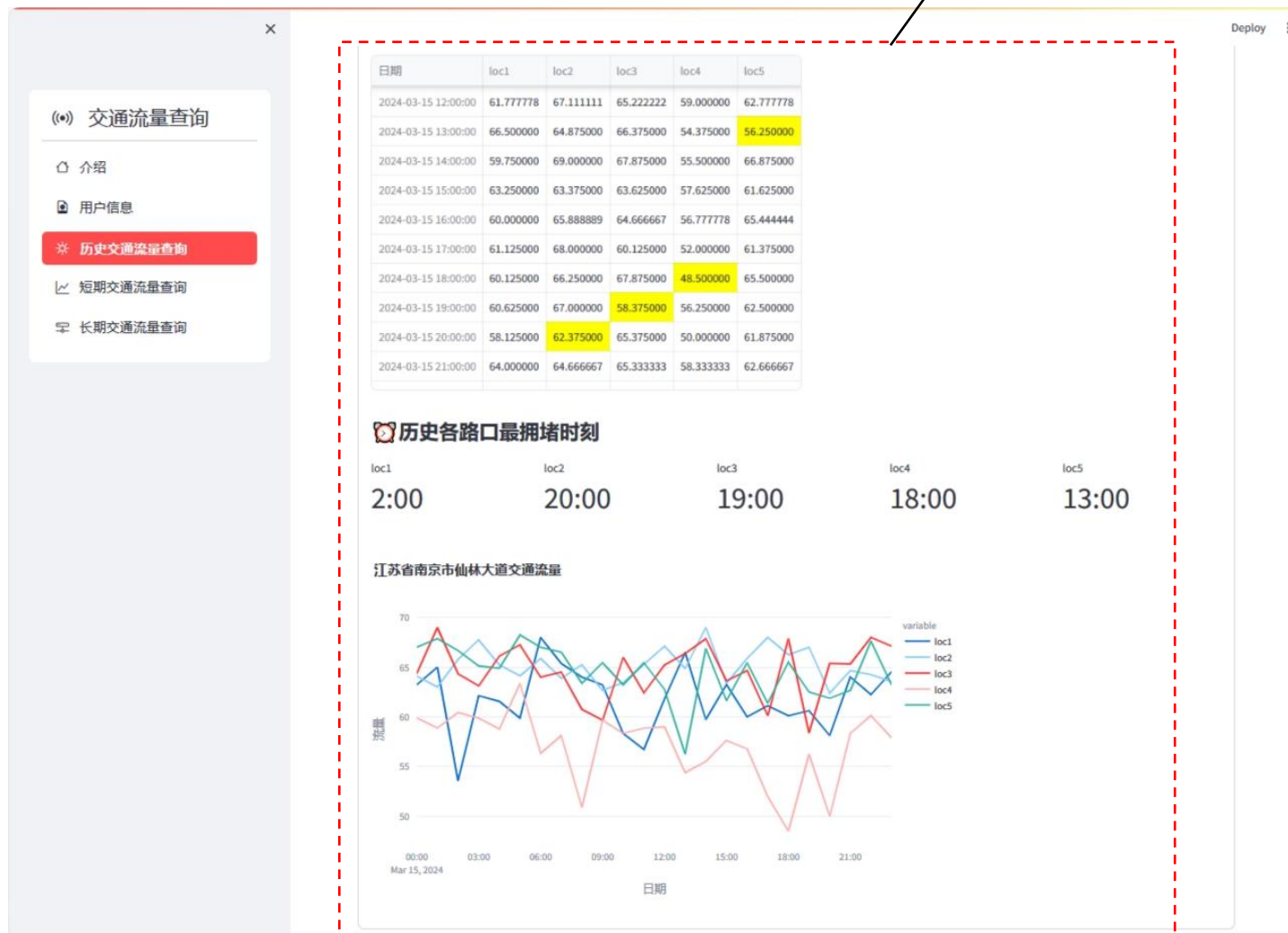
起始时间
00:00

结束时间
23:00

提交

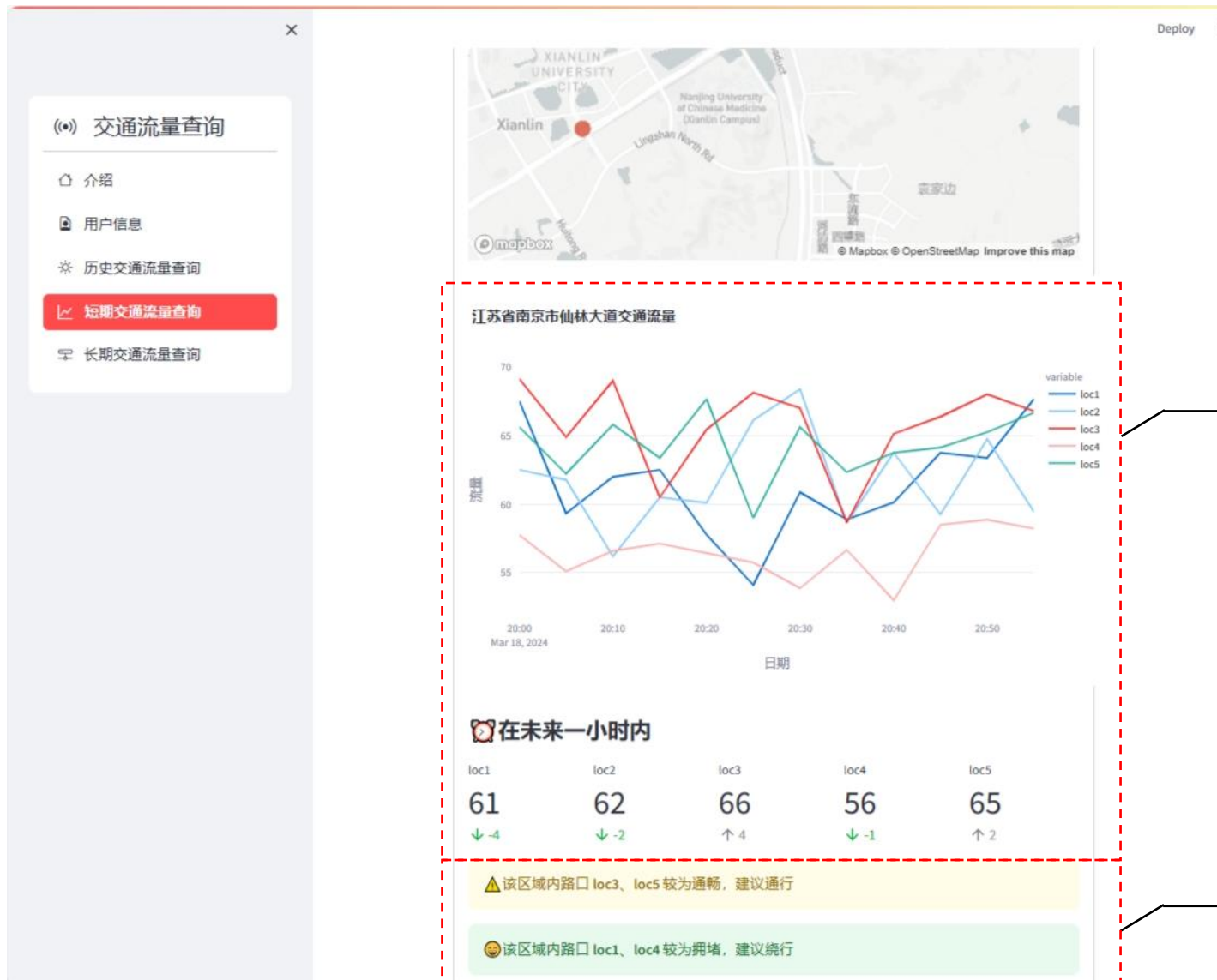
3.2 系统展示

历史交通流量查询



3.2 系统展示

未来交通流量查询



预测流量分析与可视化

路口通行建议

第四部分

研究生期间工作成果

Work Product

➤ 专利

1. 申富饶,李菲菲,郭苏涵,韩峰,赵健 一种基于多尺度空洞卷积的时间序列预测方法。
专利申请号: 202410110896.2

➤ 论文

1. Feifei Li, Suhan Guo, Feng Han, Jian Zhao, Furao Shen. Multi-Scale Dilated Convolution Network for Long-Term Time Series Forecasting. arXiv preprint arXiv:2405.05499, 2024.

➤ 项目

1. 科技部重大项目“基于神经可塑性的脉冲网络高效学习机制与类脑智能系统”（项目编号2021ZD0201300，课题年限2021年9月——2024年6月），负责时间序列预测相关问题的研究
2. 国家自然科学基金面上项目“面向增量式无监督学习的新型神经网络研究”（项目编号62276127，课题年限2023年1月——2024年6月），负责神经网络模型相关研究

第五部分

工作总结

Summary

基于浅层空洞卷积的长期时间序列预测算法

- 将多尺度空洞卷积模块与传统自回归模型相结合
- 利用空洞卷积模块的空洞性质，以在不增加计算成本的前提下，灵活调整卷积核的感受野大小
- 在推理速度和预测准确率方面均优于当前主流的预测模型

基于稀疏动态图的短期时空预测算法

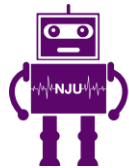
- 通过空间注意力机制构建了随时间而变化的空间动态图
- 运用硬阈值函数与Softmax函数的变体对注意力矩阵进行稀疏化处理，强化模型对空间图中关键连接的关注，同时减少对非关键连接的依赖
- 联合预定义图和动态图进行图卷积，充分提取空间依赖关系

交通流量可视化与预测系统

- 基于浅层空洞卷积的长期时间序列预测算法为系统提供“长期预测”功能
- 基于稀疏动态图的短期时空预测算法为系统提供“短期预测”功能
- 系统支持对历史流量和预测流量的数据分析与可视化



南京大學
NANJING UNIVERSITY



RINC

Robotic Intelligence & Neural Computing Group

感谢各位老师!

答辩人: 李菲菲 MF21330047

导师: 申富饶 教授

日期: 2024年5月16日

誠樸雄偉 勵學敦行