

# When Transfer Learning Meets Cross-City Urban Flow Prediction: Spatio-Temporal Adaptation Matters

阅读汇报：林宇菁 @ 北航 BIGSCITY 实验室

# 目录

CONTENTS



研究背景



研究方法



实验分析



总结



1

# 研究背景

---

*BackGround*

**城市交通流量预测：**智能交通系统（*ITS*）当中典型的时空预测任务，旨在根据历史观测数据（如出租车和自行车出行记录）预测整个城市区域的**未来人群流量**。

**城市交通迁移学习：**从交通数据丰富的大城市数据集当中学习交通预测模型，并使其在一些交通数据稀疏的城市也能进行准确地进行预测。

- 如何捕获不同城市域无关的**空间**相关性
- 如何捕获不同城市域无关的**时间**相关性


区域:  $R_c = \{r_{1,1}, \dots, r_{i,j}, \dots, r_{h,w}\}$

交通流: 
$$x_t^{(r_{i,j}, in)} = \sum_{\tau \in \mathbb{T}} |g_{t-1} \notin r_{i,j} \& g_t \in r_{i,j}|$$
$$x_t^{(r_{i,j}, out)} = \sum_{\tau \in \mathbb{T}} |g_{t-1} \in r_{i,j} \& g_t \notin r_{i,j}|$$

城市交通流快照:  $X_t \in \mathcal{R}^{h \times w \times 2}$ .

城市交通流序列:  $\mathcal{X} = \{X_{t-k+1}, \dots, X_{t-1}, X_t\} \in \mathcal{R}^{k \times h \times w \times 2}$

迁移学习: 
$$\min_f \sum_{t+1} error(Y_{t+1}^T, \hat{Y}_{t+1}^T), \hat{Y}_{t+1}^T = f(\mathcal{X}^S, \mathcal{X}^T)$$



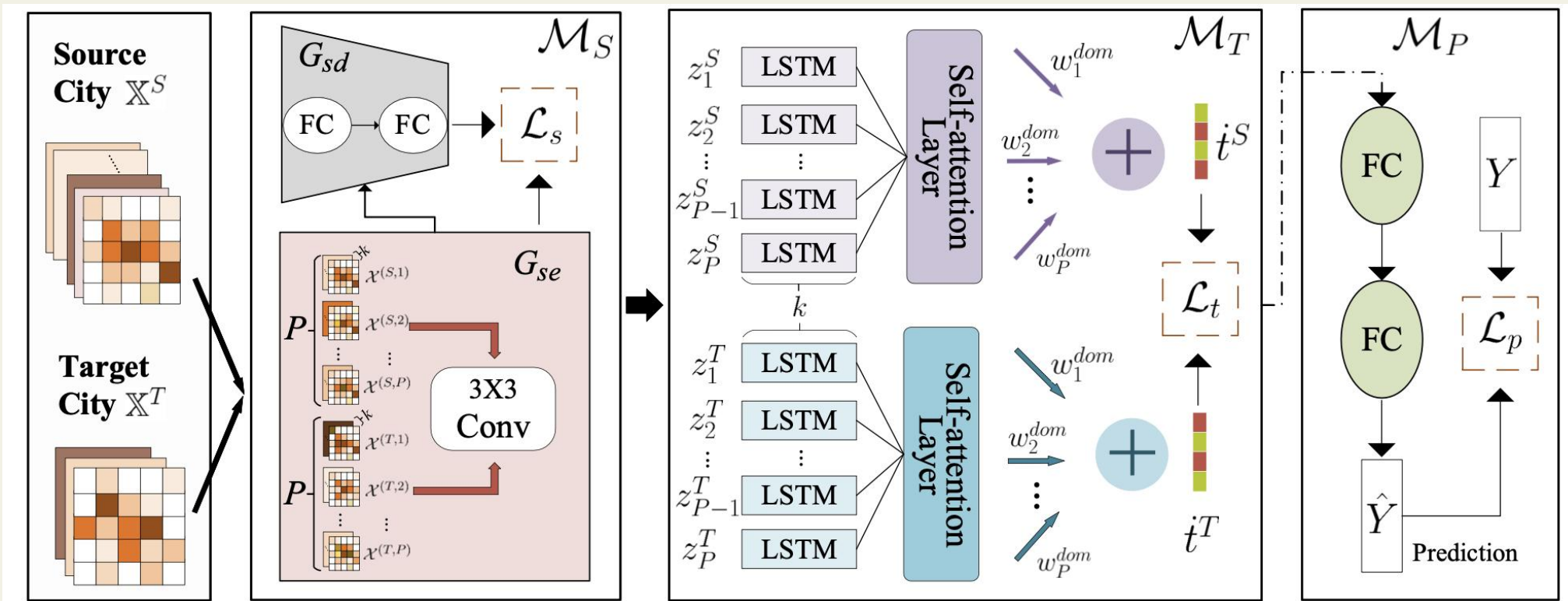
2

研究方法

---

*Method*

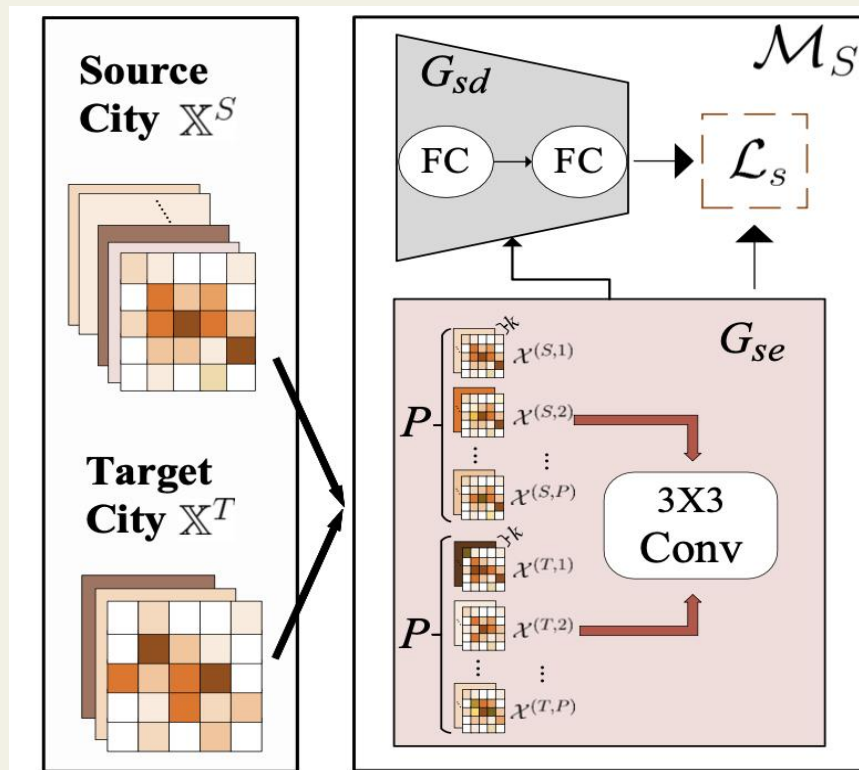
# STAN 模型概览：空间域适应模块 + 时间域适应模块 + 预测模块



空间域适应:

1. 使用 *CNN* 提取特征  $Z^S = \{z_1^S, \dots, z_{P-1}^S, z_P^S\}$

2. 域适应判别器: 类似 *GAN*, 判断表征来自源城市还是目标城市





时间域适应：最小化特征的  $MMD$  距离；在单个时间步和全局两个尺度上进行

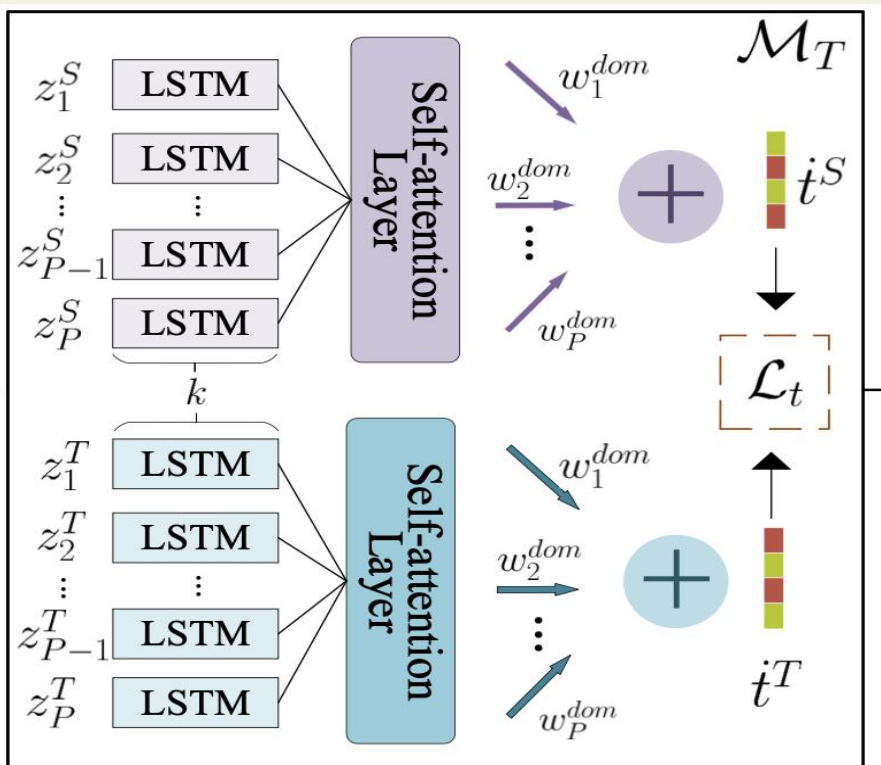
- 单时间步： $LSTM$  提取特征，代入  $MMD$  公式

$$d_p(S, T) = \frac{1}{M} \sum_{i=1}^M \|E(t_{p,i}^S) - E(t_{p,i}^T)\|_{\mathcal{H}}^2$$

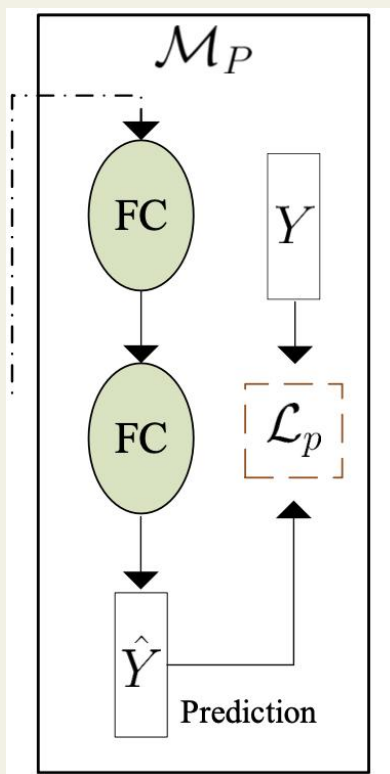
- 全局尺度：对各个时间步进行加权得到全局表征，随后代入  $MMD$  公式


$$d_p^g = |d_g - d_p|, \quad w_p^{dom} = \frac{e^{d_p^g}}{\sum_{p \in P} d_p^g}$$

$$\dot{t}^S = \sum_{p \in P} w_p^{dom} t_p^S, \quad \dot{t}^T = \sum_{p \in P} w_p^{dom} t_p^T$$



预测：使用两个全连接层进行预测





3

实验分析

---

*Experiment*

**数据集信息：** 在纽约、北京出租车租赁数据，纽约自行车数据等数据集上验证。  
都是网格流量数据，时间片长一小时或半小时

<b>Datasets</b>	<b>NYCTaxi</b>	<b>NYCBike</b>	<b>CHIBike</b>	<b>BJTaxi</b>	<b>Chengdu</b>
# of Trips	About 120 million	About 10 million	About 3 million	/	About 7 million
Longitude	(-74.02, -73.93)	(-74.02, -73.93)	(-87.73, -87.55)	(116.36, 116.50)	(103.80, 104.30)
Latitude	(40.65, 40.79)	(40.65, 40.79)	(41.76, 42.01)	(39.85, 39.99)	(30.50, 30.80)
Time range	2015.1-2015.12	2015.1-2015.12	2015.1-2015.12	2015.11-2016.4	2016.11
# of Regions	16 × 16	16 × 16	16 × 16	32 × 32	32 × 32
Time interval	1 hour	1 hour	1 hour	30 minutes	1 hour

Table 1: Statistics of the Evaluated Datasets

迁移实验：

*NYCTaxi* → *NYCBike*

*NYCBike* → *CHIBike*

*BJTaxi* → *Chengdu*

<b>Datasets</b>	<b>NYCTaxi</b>	<b>NYCBike</b>	<b>CHIBike</b>	<b>BJTaxi</b>	<b>Chengdu</b>
# of Trips	About 120 million	About 10 million	About 3 million	/	About 7 million
Longitude	(-74.02, -73.93)	(-74.02, -73.93)	(-87.73, -87.55)	(116.36, 116.50)	(103.80, 104.30)
Latitude	(40.65, 40.79)	(40.65, 40.79)	(41.76, 42.01)	(39.85, 39.99)	(30.50, 30.80)
Time range	2015.1-2015.12	2015.1-2015.12	2015.1-2015.12	2015.11-2016.4	2016.11
# of Regions	16 × 16	16 × 16	16 × 16	32 × 32	32 × 32
Time interval	1 hour	1 hour	1 hour	30 minutes	1 hour

Table 1: Statistics of the Evaluated Datasets

迁移实验:

*NYCTaxi* → *NYCBike*

*NYCBike* → *CHIBike*

*BJTaxi* → *Chengdu*

性能实验：同三类不同的模型比较，*STAN* 在 *RMSE*、*MAE* 指标上的表现均较好

Methods		NYCTaxi → NYCBike		CHIBike → NYCBike		NYCBike → CHIBike		BJTaxi → Chengdu	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Deep-Learning	ConvLSTM	—	—	0.1058	0.0807	0.1040	0.0802	0.0756	0.0389
	STResNet	—	—	0.0649	0.0417	0.0896	0.0577	0.0541	0.0283
Fine-Tuned	ConvLSTM-FT	0.0613	0.0454	0.0505	0.0345	0.0219	0.0105	0.0361	0.0214
	STResNet-FT	0.0497	0.0292	0.0480	0.0281	0.0319	0.0176	0.0444	0.0230
	DCRNN-FT	0.0411	0.0292	0.0392	0.0279	0.0392	0.0276	—	—
Transfer-Learning	MetaST	0.0526	0.0349	0.0453	0.0226	0.1003	0.0656	0.0390	0.0275
	ST-DAAN	0.0381	0.0116	0.0381	0.0116	0.0234	0.0070	0.0909	0.0557
	STAN-No-SAAM	0.0344	0.0119	0.0371	0.0117	0.0215	0.0068	0.0357	0.0202
	STAN-No-TAAM	0.0345	0.0119	0.0362	0.0116	0.0215	0.0069	0.0368	0.0209
	STAN	<b>0.0327</b>	<b>0.0111</b>	<b>0.0361</b>	<b>0.0114</b>	<b>0.0196</b>	<b>0.0049</b>	<b>0.0345</b>	<b>0.0196</b>

Table 2: The Performance Comparison of All Methods

消融实验：去除时间、空间相关性的域适应模块，模型效果均有所下降（但仍好于基线）

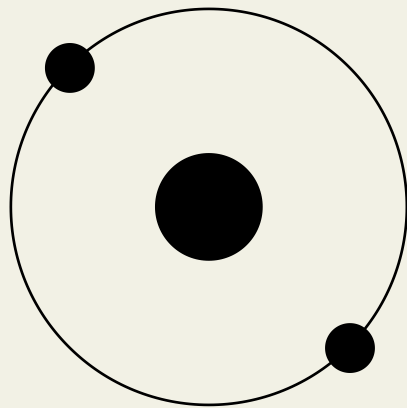
# 总结

CONCLUSION

城市流量预测的迁移学习与传统的交通预测研究范式基本是一致的，都是从时间空间两个角度上进行特征抽取。

难点在于如何使得抽取出来的特征在源城市和目标城市都具有好的预测效果。

这一点通常是使用特征域适应做到的，其做法就是让源城市和目标城市抽取出来的特征具有相似的分布。具体方法有对抗损失、分布距离等。



# 谢谢观看

---

Thanks for watching