Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer

Bin Lu[1], Xiaoying Gan[1], Weinan Zhang[1], Huaxiu Yao[2], Luoyi Fu[1], Xinbing Wang[1]

[1]Shanghai Jiao Tong University, [2]Stanford University

Introduction

➢ Background

- Graphs are ubiquitously used to reveal the interactions among various entities
- Spatio-temporal graph learning is a widely used method for urban computing

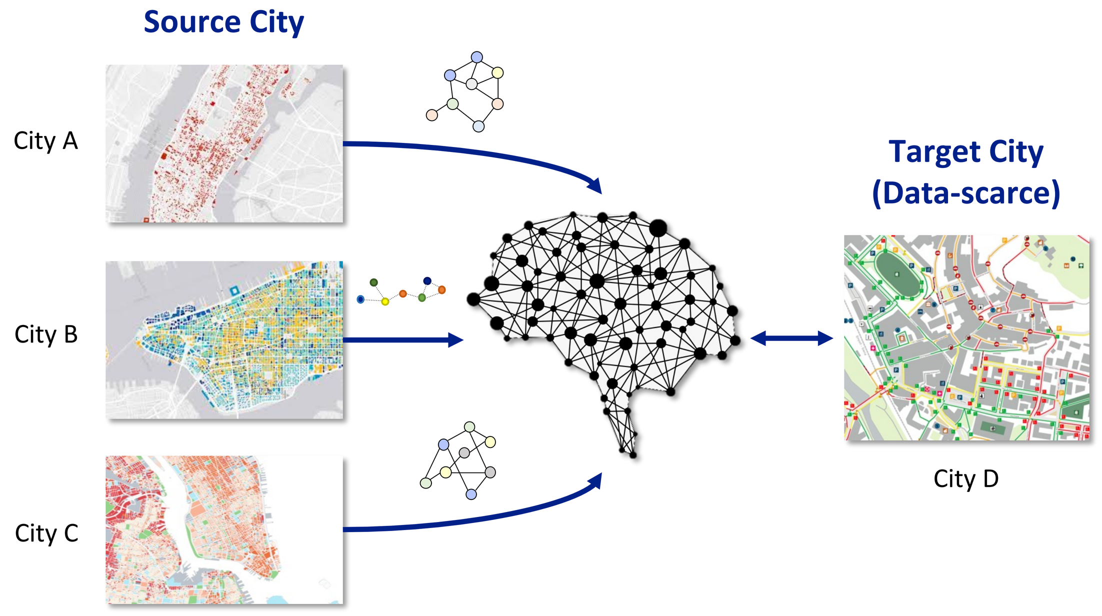
➢ Our Goal

- Transfer the cross-city knowledge in graph-based few-shot learning scenarios.
- Exploring the impact of knowledge transfer across multiple cities.

➢ Challenges

#1 How to adapt feature extraction in target city via the knowledge from multiple source cities?

#2 How to alleviate the impacts of varied graph structure on transferring among different cities?



Methodology

➢ Spatio-Temporal Neural Network (STNN)

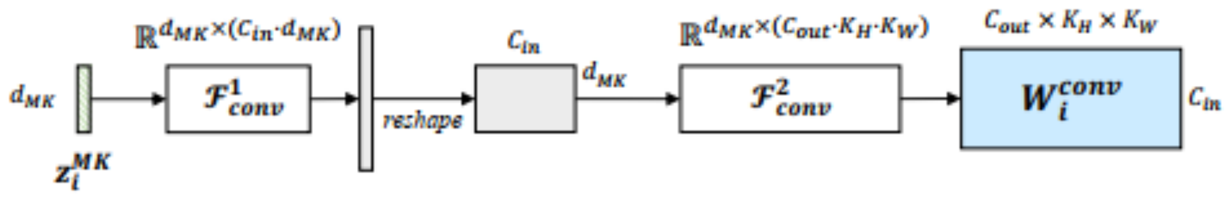
Backbone for feature extractor (time series model, graph neural network, hybrid model, etc.)

➢ Cross-City Knowledge Transfer

1. ST-Meta Learner: obtain the node-level meta knowledge

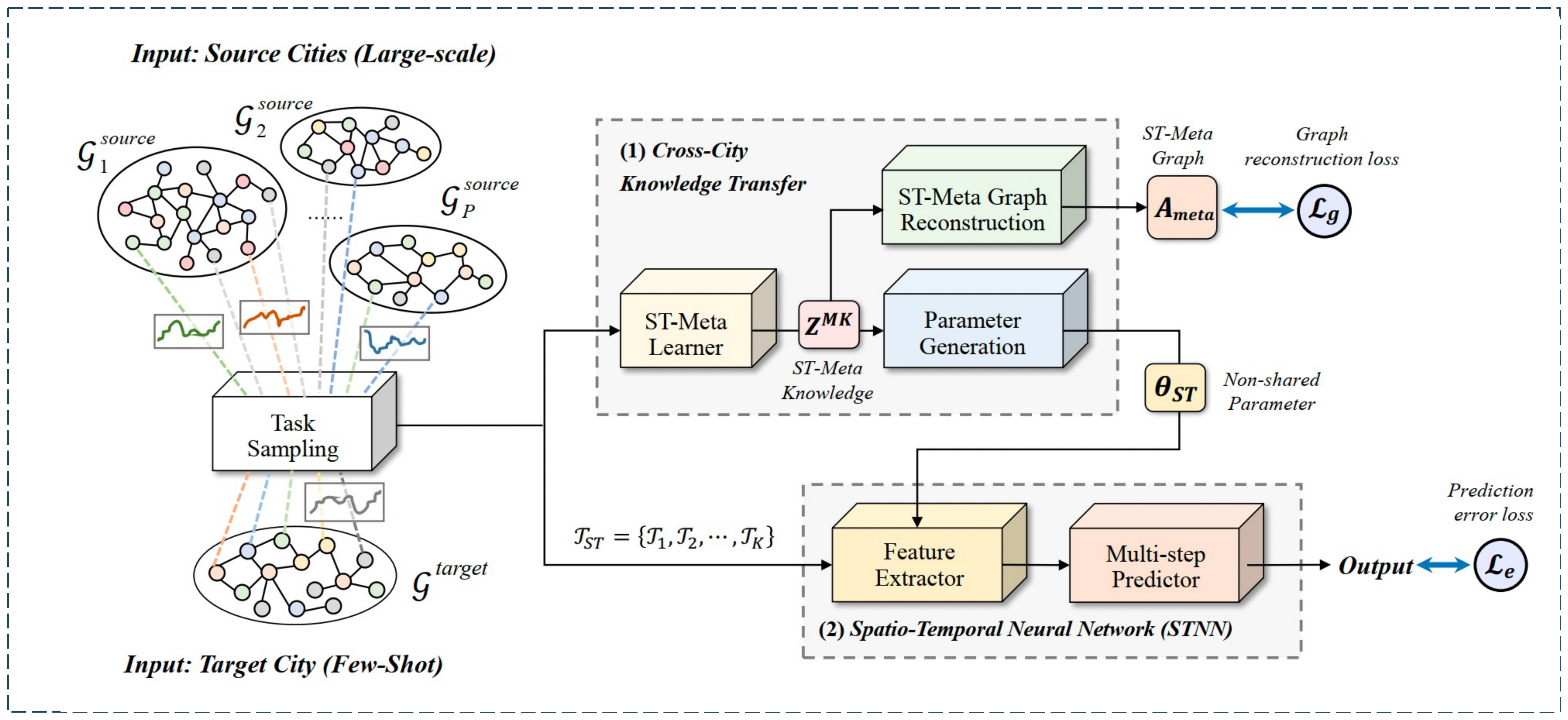
$$Z^{MK} = W^{\gamma}(\gamma \circ Z^{tp} + (1-\gamma) \circ Z^{sp})$$

2. Parameter Generation: customize feature extraction among different cities



3. ST-Meta Graph Reconstruction: structure-aware learning

$$A_{meta} = sigmoid[(Z^{MK})^T \cdot Z^{MK}] \qquad \mathcal{L}_g = \|A_{meta} - A\|^2$$



Overview of proposed ST-GFSL framework

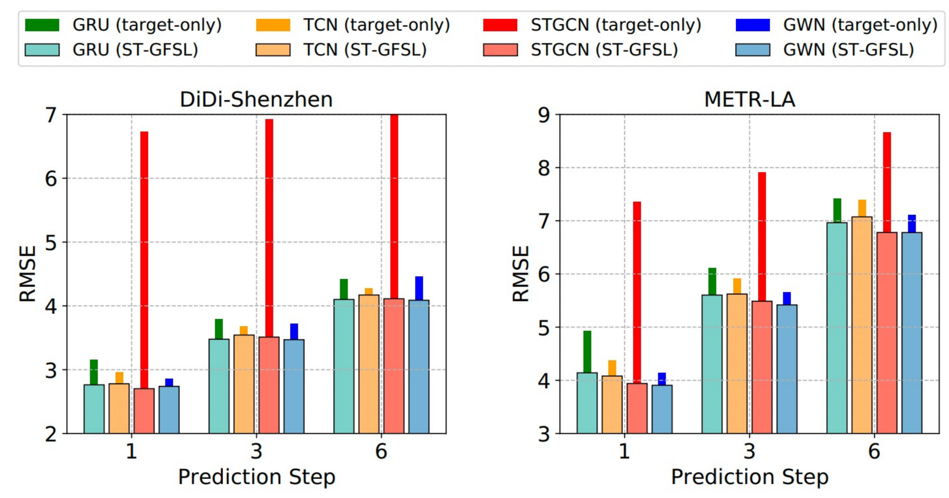
Experiments

① Model Comparison: Non-transfer Methods & Transfer Method

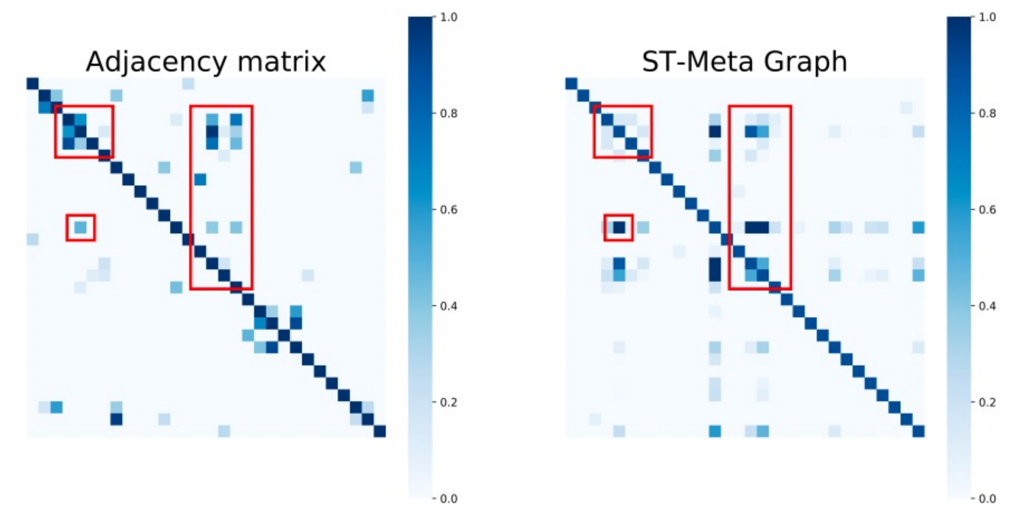
Baselines	PEMS-BAY Dataset						METR-LA Dataset					
	MAE (↓)			RMSE (↓)			MAE (↓)			RMSE (↓)		
	5 min	15 min	30 min	5 min	15 min	30 min	5 min	15 min	30 min	5 min	15 min	30 min
HA	4.373	4.373	4.373	6.745	6.745	6.745	6.021	6.021	6.021	9.483	9.483	9.483
ARIMA	2.019	2.307	2.429	3.929	4.648	5.360	2.900	3.058	3.369	4.179	5.279	7.670
Target-only	1.556	1.920	2.368	3.092	4.043	5.153	2.740	3.229	3.860	4.924	6.118	7.417
Fine-tuned (Vanilla)	1.823	2.166	2.590	3.434	4.280	5.276	2.757	3.277	3.900	4.883	6.123	7.413
Fine-tuned (ST-Meta)	1.371	1.791	2.277	2.699	3.747	4.920	2.647	3.188	3.800	4.368	5.759	7.110
AdaRNN [29]	1.248	1.928	2.749	2.084	3.796	5.725	2.513	2.897	**3.312**	4.298	**5.567**	**6.732**
MAML [14]	1.081	1.600	2.141	1.906	3.291	4.708	2.405	2.960	3.639	4.159	5.710	7.124
ST-GFSL (ours)	**1.073**	**1.560**	**2.073**	**1.865**	**3.180**	**4.584**	**2.355**	**2.896**	3.557	**4.099**	5.588	6.961

Baselines	Didi-Chengdu Dataset						Didi-Shenzhen Dataset					
	MAE (↓)			RMSE (↓)			MAE (↓)			RMSE (↓)		
	10 min	30 min	60 min	10 min	30 min	60 min	10 min	30 min	60 min	10 min	30 min	60 min
HA	3.438	3.438	3.438	4.879	4.879	4.879	2.955	2.955	2.955	4.342	4.342	4.342
ARIMA	2.825	3.305	4.317	3.889	4.253	5.597	2.888	3.056	3.596	4.489	4.764	5.575
Target-only	2.386	2.700	3.085	3.516	4.017	4.569	2.071	2.454	2.834	3.154	3.793	4.422
Fine-tuned (Vanilla)	2.586	2.877	3.246	3.746	4.213	4.751	2.117	2.490	2.867	3.196	3.831	4.442
Fine-tuned (ST-Meta)	2.240	2.693	3.083	3.249	3.956	4.519	2.033	2.454	2.850	2.989	3.719	4.385
AdaRNN [29]	2.260	2.724	3.036	3.231	3.942	**4.324**	2.107	2.473	2.807	3.041	3.674	4.231
MAML [14]	2.215	2.599	2.956	3.215	3.858	4.399	1.917	2.330	2.673	2.825	3.546	4.158
ST-GFSL (ours)	**2.188**	**2.579**	**2.927**	**3.190**	**3.820**	4.339	**1.890**	**2.288**	**2.644**	**2.763**	**3.477**	**4.100**

② Performance Comparison:
Different feature extractors



③ Case Study:
Graph reconstruction loss



④ Ablation Study

Ablation Methods	DiDi-Shenzhen			METR-LA		
	MAE (↓)			MAE (↓)		
	10 min	30 min	60 min	10 min	30 min	60 min
(M1a): Use temporal meta knowledge only	1.910	2.317	2.668	**2.332**	2.905	3.616
(M1b): Use spatial meta knowledge only	1.872	2.300	2.649	2.364	2.915	3.604
(M1c): Use random initialized vectors	1.937	2.332	2.680	2.422	2.949	3.697
(M2): Remove parameter generator	1.917	2.330	2.673	2.405	2.960	3.639
(M3): Remove graph reconstruction loss	2.286	2.652	3.000	3.087	3.585	4.140
ST-GFSL (Ours)	**1.856**	**2.290**	**2.634**	2.387	**2.895**	**3.546**

Conclusion

- We firstly propose a spatio-temporal graph few-shot learning framework called ST-GFSL for cross-city knowledge transfer.
- Extensive experimental results in the running case of traffic speed prediction demonstrate the superiority of ST-GFSL over other baselines.
- In the future, we will further explore source domain selection problem.

Contact Us

- Bin Lu: robinlu1209@sjtu.edu.cn
- Intelligent IoT Research Center: https://iiot.sjtu.edu.cn