A Brief Introduction To Contrastive learning

Three key procedures of contrastive learning:

- 1. Design instances.
- 2. Define (dis)similar instances.
- 3. Define encoders.

Problems: treating two different but semantic similar samples as a negative pair may hurt the performance.

A Brief Introduction To Contrastive learning

- *Can we evaluate the spatial-temporal similarity of traffic pattern between locations?*

  (For the purpose of defining positive/negative sample pairs)

Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction

AAAI 2019, CITATION 120+

Introduction

- The challenge of traffic prediction lies in how to model the complex spatial dependencies and temporal dynamics
- Existing work make strong assumptions about spatial dependence and temporal dynamics. i.e. spatial dependence is stationary in time, and temporal dynamics is strictly periodical. However, in practice the spatial dependence could be dynamic, and the temporal dynamics could have some perturbation from one period to another period.
- The authors make two observations:

(1) The spatial dependencies between locations could be dynamics;

(2) The temporal dependency follows daily and weekly pattern but it is not strictly periodic for its dynamic temporal shifting.

- To address these two issues, the authors propose a Spatial-Temporal Dynamic Network(STDN), using a flow gating mechanism to learn the dynamic similarity between locations, and a periodically shifted attention mechanism is designed to handle long-term periodic temporal shifting.

Notations and Problem Formulation

- Spatial partition: Split the whole city to an a x b grid map with n regions in total(n= a x b), and use {1, 2, …, n} to denote them.
- Temporal partition: Split the whole time period into m equal-length continuous time intervals.
- Start/end traffic volume: $y_{i,t}^{s}$ and $y_{i,t}^{e}$ stand for the start/end traffic volume for region I during the t-th time interval.
- Traffic flow: the traffic flow starting from region i in time interval t and ending in region j in time interval τ is denoted as $f_{i,t}^{j,\tau}$
- Problem: Given the data until time interval t, the traffic volume prediction problem aims to predict the start and end traffic volume at time interval t+1.

Spatial-Temporal Dynamic Network

- How to model the spatial dependency?

Local CNN without considering spatial dynamic similarity

$$\mathbf{Y}_{i,t}^{(k)} = \mathrm{ReLU}(\mathbf{W}^{(k)} * \mathbf{Y}_{i,t}^{(k-1)} + \mathbf{b}^{(k)}), \qquad (1)$$

- How to represent the spatial similarity?

If there are more flows existing between two regions, the relation between them is stronger.

After considering spatial dynamic similarity:

$$\mathbf{F}_{i,t}^{(k)} = \mathrm{ReLU}(\mathbf{W}_f^{(k)} * \mathbf{F}_{i,t}^{(k-1)} + \mathbf{b}_f^{(k)}), \qquad (3)$$

$$\mathbf{Y}_{i,t}^{(k)} = \mathrm{ReLU}(\mathbf{W}^{(k)} * \mathbf{Y}_{i,t}^{(k-1)} + \mathbf{b}^{(k)}) \otimes \sigma(\mathbf{F}_t^{i,k-1}), \quad (4)$$

Spatial-Temporal Dynamic Network

- How to model the temporal dependency? LSTM

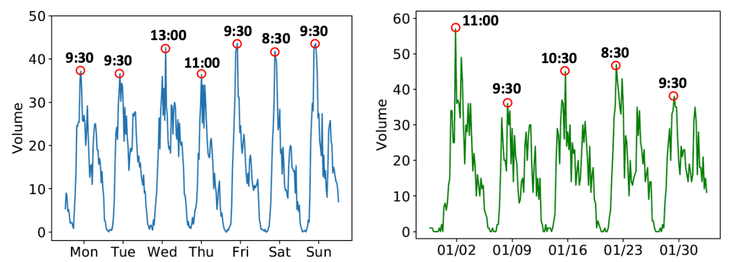
$$\mathbf{h}_{i,t} = \text{LSTM}([\mathbf{y}_{i,t}; \mathbf{e}_{i,t}], \mathbf{h}_{i,t-1}), \qquad (2)$$

- How to handle the long-term information?

  since the increasing length enlarges the risk of gradient vanishing, thus significantly weaken the effects of periodicity.

  To address this issue, relative time intervals of the predicting target (e.g., same time of yesterday, and the day before yesterday) should be explicitly modeled.

  However, purely incorporating relative time intervals is insufficient ignores temporal shifting of periodicity, i.e., traffic data is not strictly periodic.



(a) (b)

Figure 2: The temporal shifting of periodicity. (a) Temporal shifting between different days. (b) Temporal shifting between different weeks. Note that, each time in these figures represents a time interval (e.g., 9:30am means 9:00-9:30am).

Spatial-Temporal Dynamic Network

- How to address the shifting in daily periodicity?

(1) relative time intervals from previous P days are included for handling the periodic dependency.

(2) we further select Q time intervals from each day in Q.

$$\mathbf{h}_{i,t}^{p,q} = \mathrm{LSTM}([\mathbf{y}_{i,t}^{p,q}; \mathbf{e}_{i,t}^{p,q}], \mathbf{h}_{i,t}^{p,q-1}), \quad (5)$$

$$\mathbf{h}_{i,t}^{p} = \sum \alpha_{i,t}^{p,q} \mathbf{h}_{i,t}^{p,q}, \quad (6)$$

$$\alpha_{i,t}^{p,q} = \frac{\exp(\mathrm{score}(\mathbf{h}_{\mathrm{i,t}}^{\mathrm{p,q}}, \mathbf{h}_{\mathrm{i,t}}))}{\sum_{q \in Q} \exp(\mathrm{score}(\mathbf{h}_{\mathrm{i,t}}^{\mathrm{p,q}}, \mathbf{h}_{\mathrm{i,t}}))}. \quad (7)$$

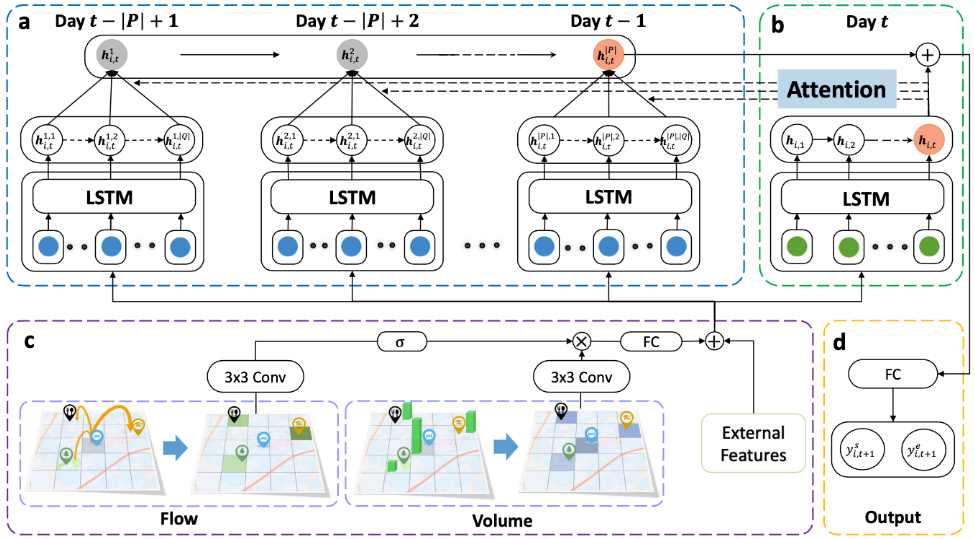
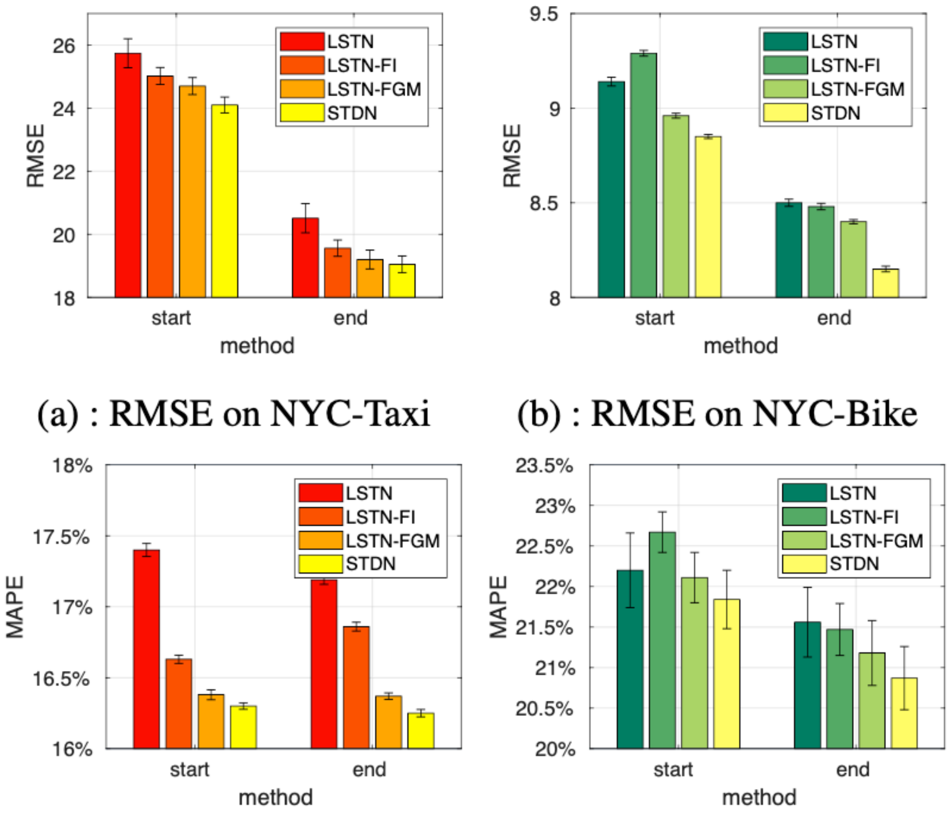


Figure 1: The architecture of STDN. (a) Periodically shifted attention mechanism captures the long-term periodic dependenc and temporal shifting. For each day, we also use LSTM to capture the sequential information. (b) The short-term tempora dependency is captured by one LSTM. (c) The flow gating mechanism tracks the dynamic spatial similarity representation b controlling the spatial information propagation; FC means fully connected layers and Conv means several convolutional layers (d) A unified multi-task prediction component predicts two types of traffic volumes simultaneously.



(a) : RMSE on NYC-Taxi (b) : RMSE on NYC-Bike

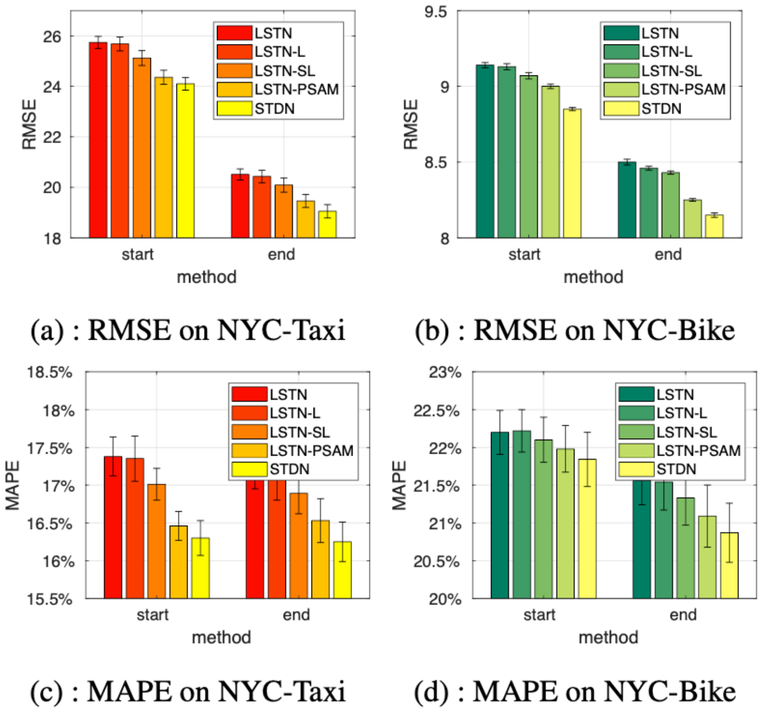
(c) : MAPE on NYC-Taxi (d) : MAPE on NYC-Bike

Figure 3: Evaluation of flow gating mechanism (FGM) and its variants.

Table 1: Comparison with Different Baselines

Dataset	Method	Start		End	
		RMSE	MAPE	RMSE	MAPE
NYC-Taxi	HA	43.82	23.18%	33.83	21.14%
	ARIMA	36.53	22.21%	27.25	20.91%
	LR	28.51	19.94%	24.38	20.07%
	MLP	26.67±0.56	18.43±0.62%	22.08±0.50	18.31±0.83%
	XGBoost	26.07	19.35%	21.72	18.70%
	LinUOTD	28.48	19.91%	24.39	20.03%
	ConvLSTM	28.13±0.25	20.50±0.10%	23.67±0.20	20.70±0.20%
	DeepSD	26.35±0.53	18.12±0.38%	21.95±0.35	18.15±0.62%
	ST-ResNet	26.23±0.33	21.13±0.63%	21.63±0.25	21.09±0.51%
	DMVST-Net	25.74±0.26	17.38±0.46%	20.51±0.46	17.14±0.32%
	STDN	**24.10±0.25***	**16.30±0.23%***	**19.05±0.31***	**16.25±0.26%***
NYC-Bike	HA	12.49	27.82%	11.93	27.06%
	ARIMA	11.53	26.35%	11.25	25.79%
	LR	10.92	25.29%	10.33	24.58%
	MLP	9.83±0.19	23.12±0.47%	9.12±0.24	22.40±0.40%
	XGBoost	9.57	23.52%	8.94	22.54%
	LinUOTD	11.04	25.22%	10.44	24.44%
	ConvLSTM	10.40±0.17	25.10±0.45%	9.22±0.19	23.20±0.47%
	DeepSD	9.69	23.62%	9.08	22.36%
	ST-ResNet	9.80±0.12	25.06±0.36%	8.85±0.13	22.98±0.53%
	DMVST-Net	9.14±0.13	22.20±0.33%	8.50±0.19	21.56±0.49%
	STDN	**8.85±0.11***	**21.84±0.36%**	**8.15±0.15***	**20.87±0.39%***

*** (**) means the result is significant according to Students T-test at level 0.01 (0.05) compared to DMVST-Net



(a) : RMSE on NYC-Taxi
(b) : RMSE on NYC-Bike
(c) : MAPE on NYC-Taxi
(d) : MAPE on NYC-Bike

Figure 4: Evaluation of periodically shifted attention mechanism (PSAM) and its variants.

Conclusion and Discussion

- STDN method tracks the dynamics spatial similarity between regions by flow gating mechanism and temporal similarity by periodically shifted attention mechanism.
- The proposed model on other spatial-temporal prediction problems needs to be investigated.
- The feature importance of traffic prediction needs to be further explained, which is important for policy makers.

Further reading

- Brandes, Ulrik, and Jürgen Lerner. "Structural similarity in graphs." *International Symposium on Algorithms and Computation*. Springer, Berlin, Heidelberg, 2004.
- Castrillo, Eduar, Elizabeth León, and Jonatan Gómez. "Dynamic Structural Similarity on Graphs." *arXiv preprint arXiv:1805.01419* (2018).
- Goldsmith, Timothy E., and Daniel M. Davenport. "Assessing structural similarity of graphs." (1990). (Citation 200+)
- Gentner, Dedre, and Arthur B. Markman. "Defining structural similarity." *The Journal of Cognitive Science* 6.1 (2006): 1-20. (Citation 80+)
- Zhou, Yang, Hong Cheng, and Jeffrey Xu Yu. "Graph clustering based on structural/attribute similarities." *Proceedings of the VLDB Endowment* 2.1 (2009): 718-729. (Citation 800+)
- Randić, Milan, and Charles L. Wilkins. "Graph theoretical approach to recognition of structural similarity in molecules." *Journal of Chemical Information and Computer Sciences* 19.1 (1979): 31-37.
- Cheng, Hong, Yang Zhou, and Jeffrey Xu Yu. "Clustering large attributed graphs: A balance between structural and attribute similarities." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5.2 (2011): 1-33.
- Yan, Xifeng, Philip S. Yu, and Jiawei Han. "Substructure similarity search in graph databases." *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. 2005. (Citation 390+)

Thanks !