http://ivaderlab.unist.ac.kr

Learning to Remember Patterns: Pattern Matching Memory Networks for Traffic Forecasting

Hyunwook Lee, Seungmin Jin, Hyeshin Chu, Hongkyu Lim, Sungahn Ko*

Ulsan National Institute of Science and Technology (UNIST)

ICLR 2022

Contact information:
gusdnr0916@unist.ac.kr

Traffic Forecasting

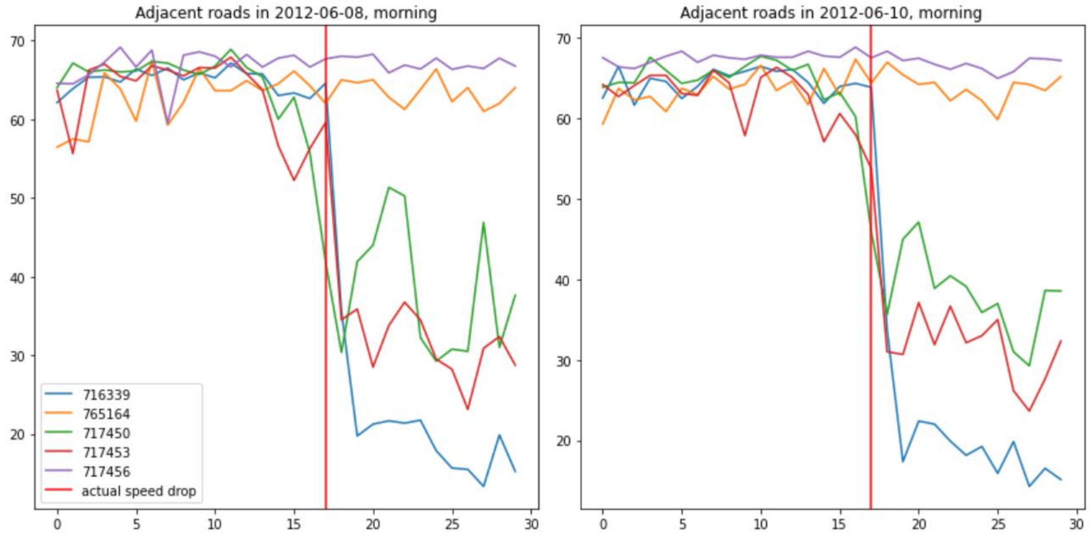
- Traffic forecasting is well-known and well-studied subject in time-series forecasting

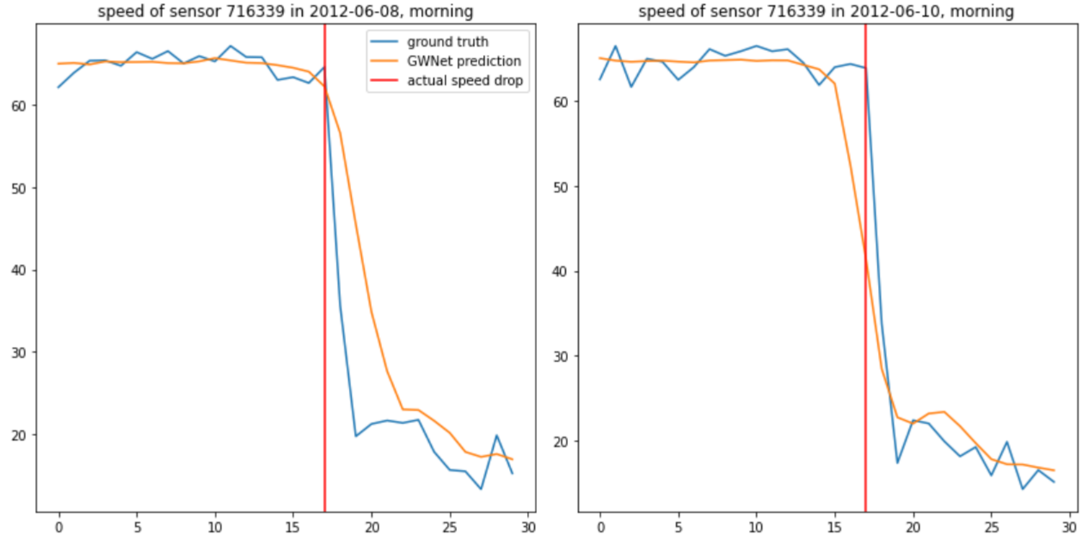
- Traffic forecasting model should deal with both spatial and temporal dependency

- Our main focus is how to model temporal correlation between similar and different circumstances

Challenges in traffic forecasting

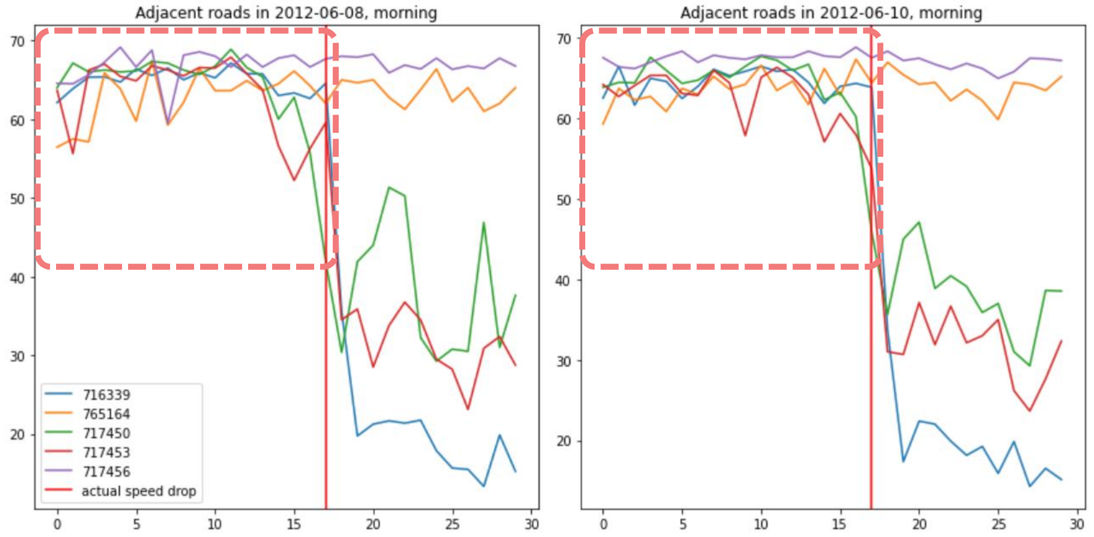
- Sensitive to **noise / fluctuation**

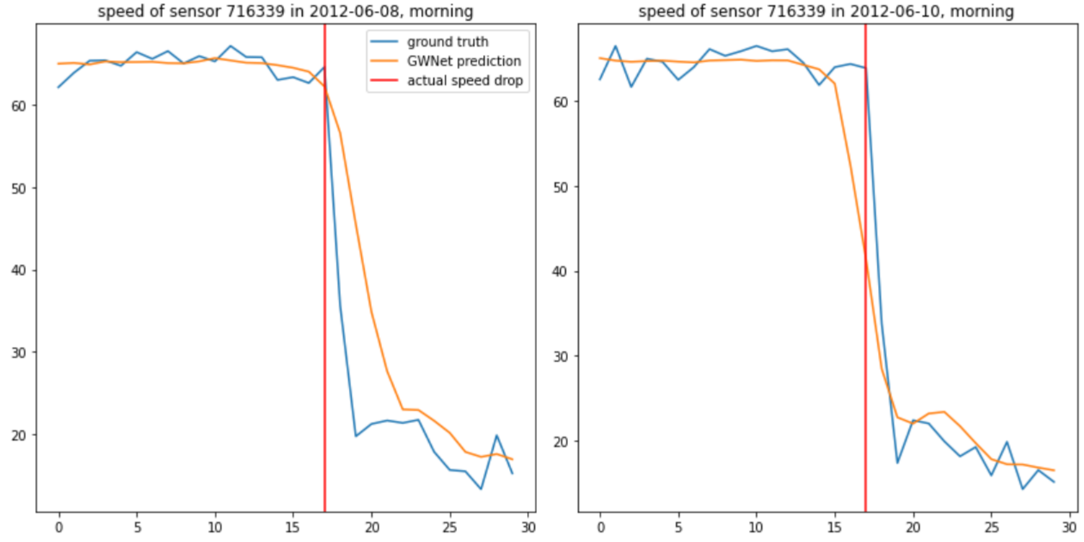




Challenges in traffic forecasting

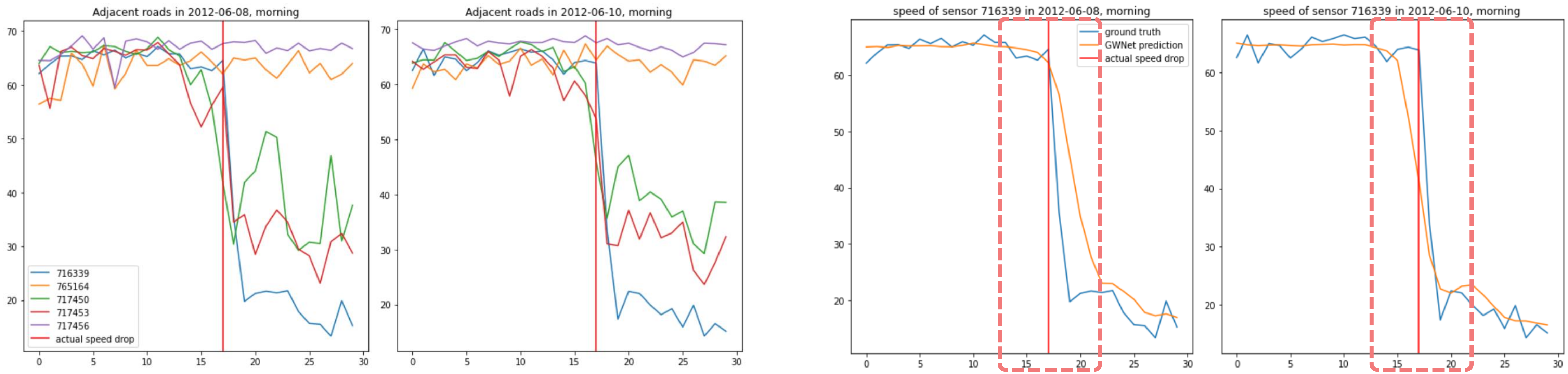
- Sensitive to **noise / fluctuation**





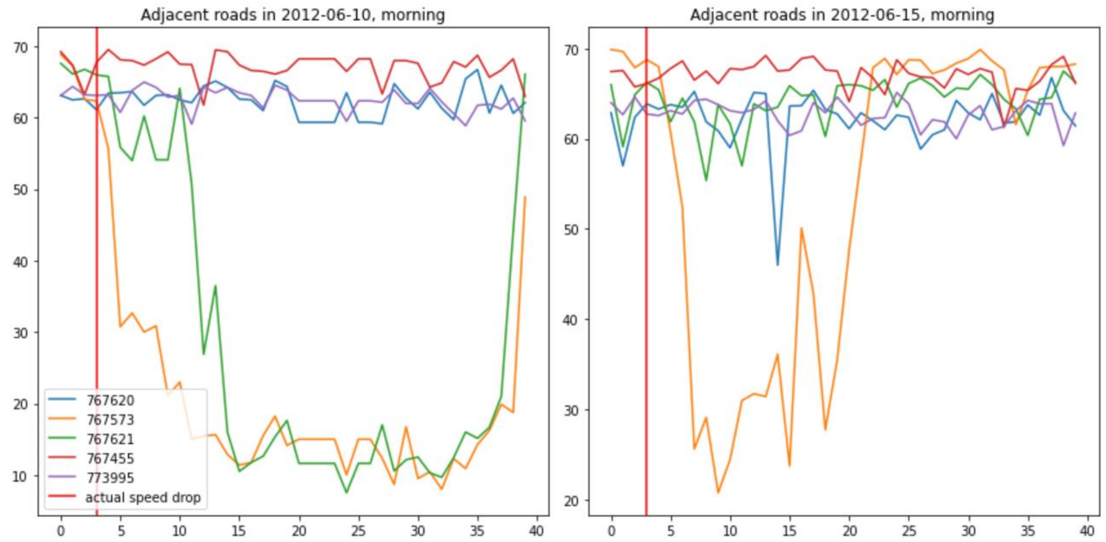
Challenges in traffic forecasting

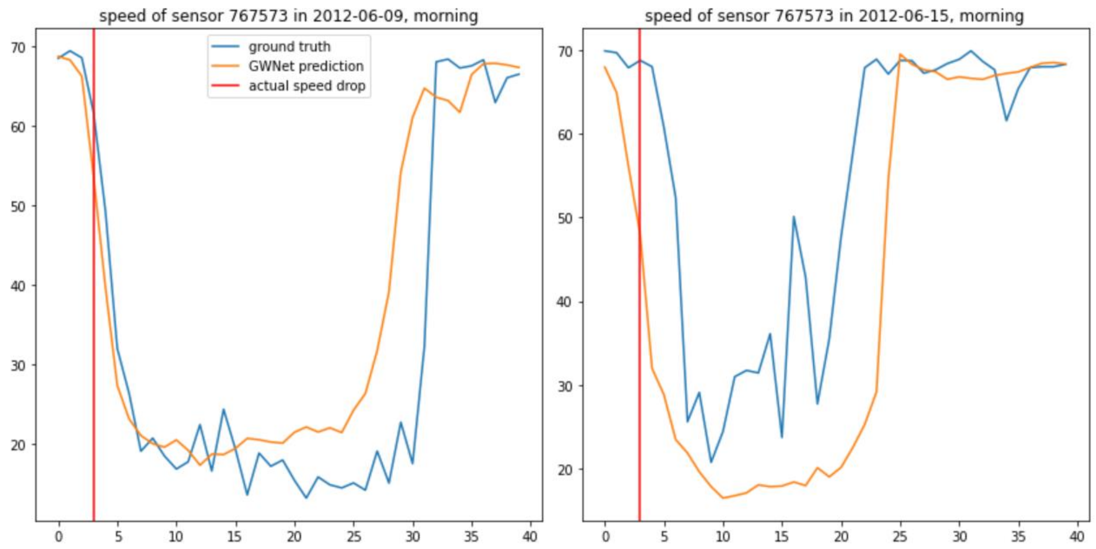
- Sensitive to **noise / fluctuation**



Challenges in traffic forecasting

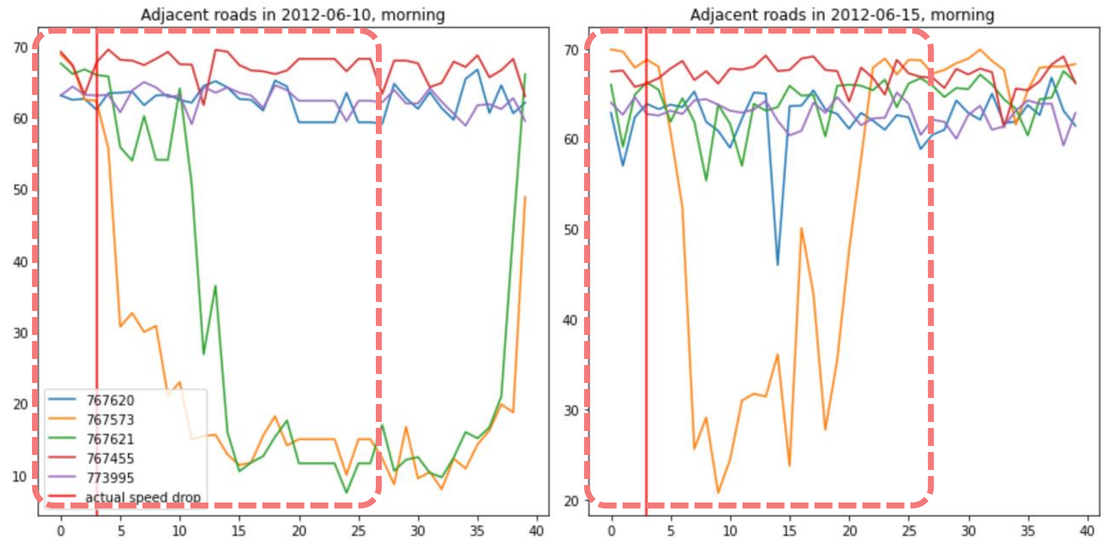
- Sensitive to **noise / fluctuation**
- Rarely recognize situation (or speed pattern) for the prediction



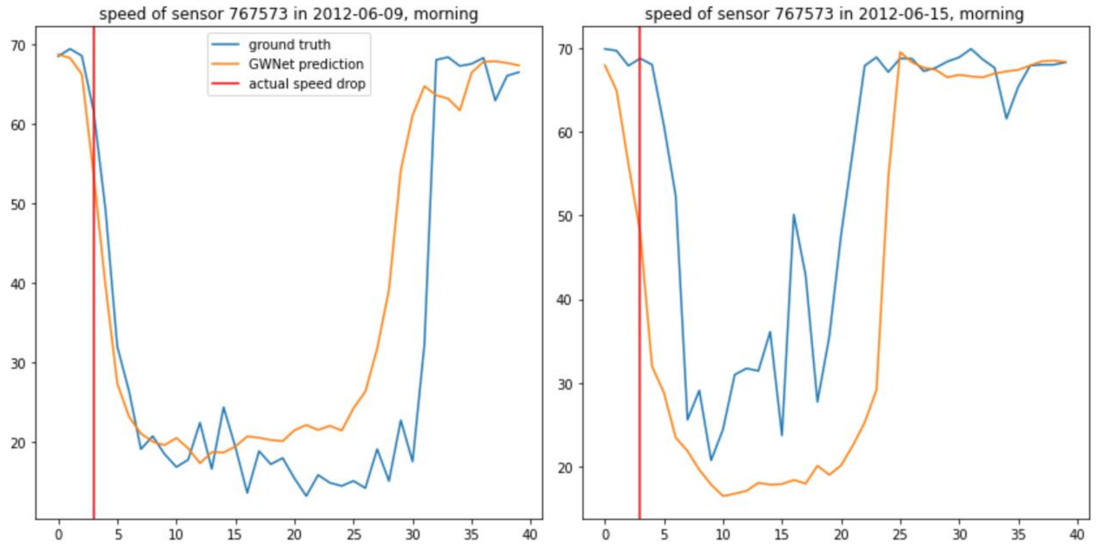


Challenges in traffic forecasting

- Sensitive to **noise / fluctuation**
- Rarely recognize situation (or speed pattern) for the prediction

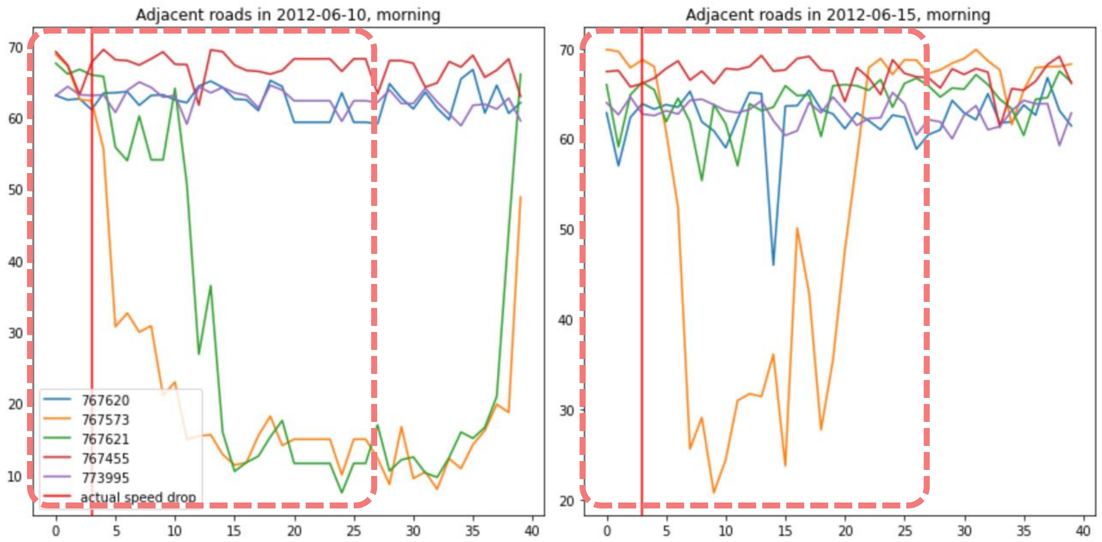


Road #767621 and road #767573 act
differently in two situation

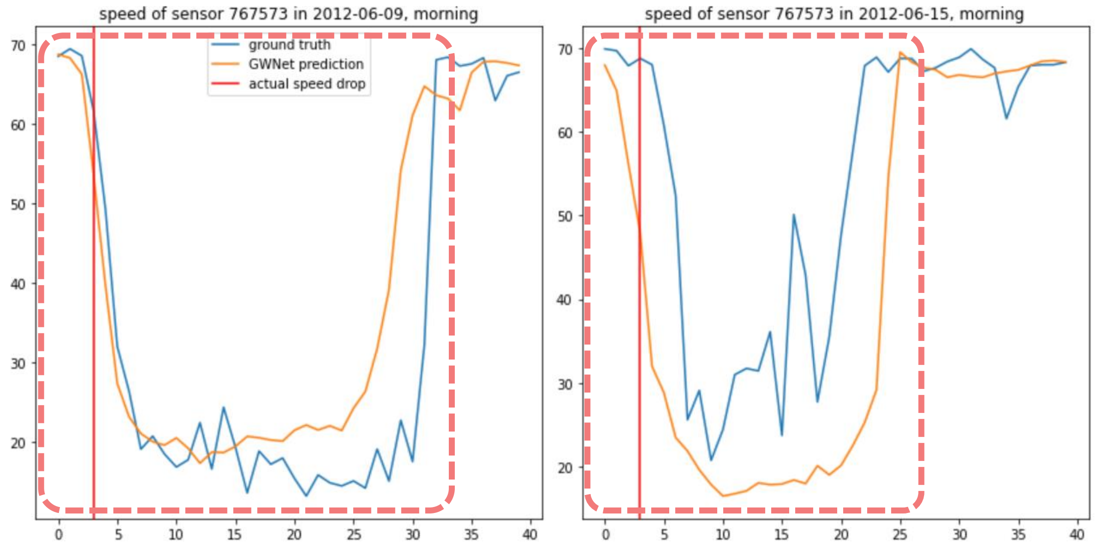


Challenges in traffic forecasting

- Sensitive to **noise / fluctuation**
- Rarely recognize situation (or speed pattern) for the prediction
- How to acquire robust / situation(pattern)-aware forecasting?



Road #767621 and road #767573 act **differently in two situation**



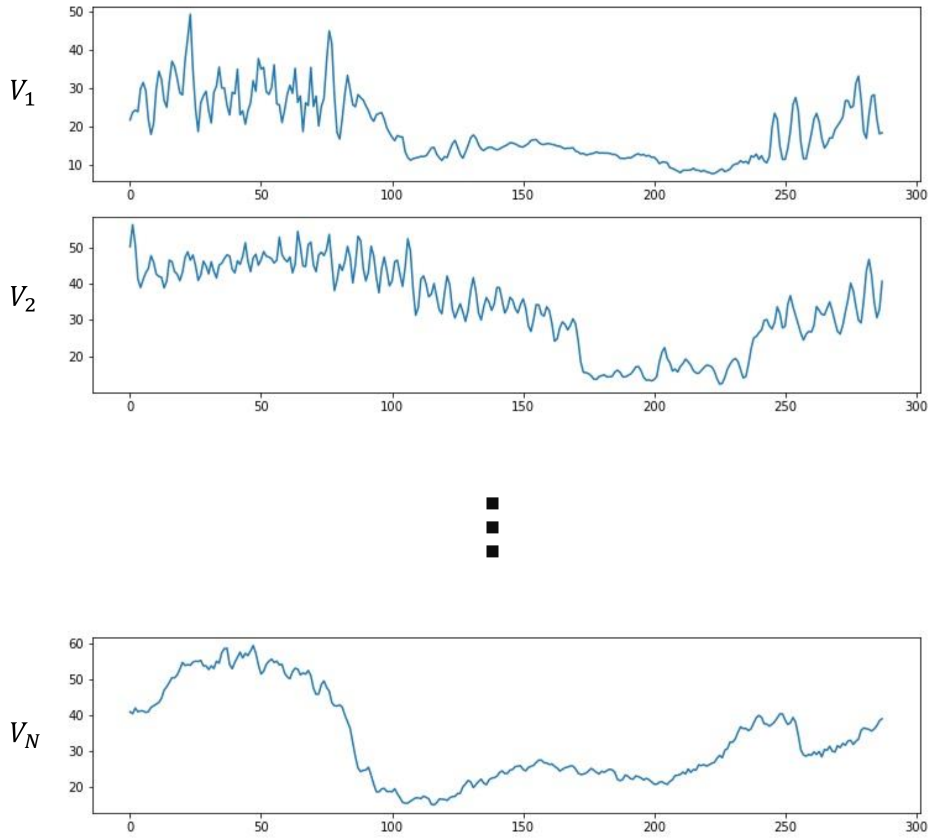
Model cannot recognize situation and have very similar forecasting

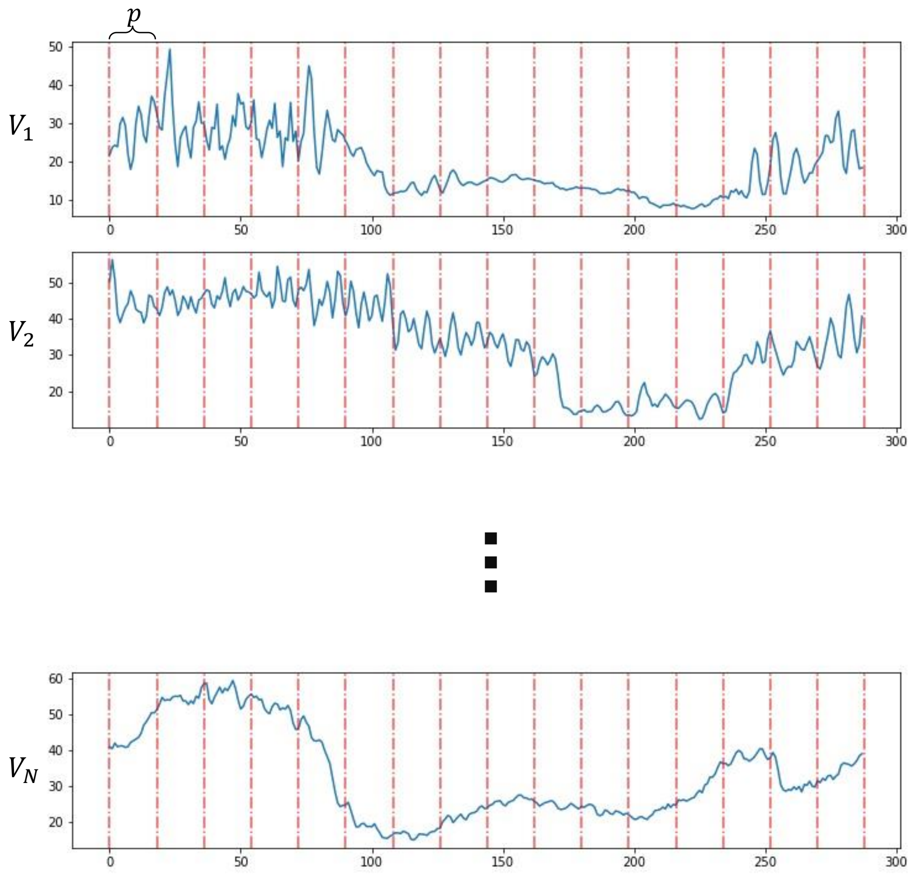
Problem Reformulation

- **Traditionally, traffic forecasting problem is formulated:**
  - $\left[X_{\mathcal{G}}{}^{(t-T'+1)}, \ldots, X_{\mathcal{G}}{}^{(t)}\right] \xrightarrow{f(\cdot)} \left[X_{\mathcal{G}}{}^{(t+1)}, \ldots, X_{\mathcal{G}}{}^{(t+T)}\right]$
  - Not pattern-aware forecasting but **momentary, instance-sensitive**
- **In this work, we reformulate the problem as:**
  - $\left[X_{\mathcal{G}}{}^{(t-T'+1)}, \ldots, X_{\mathcal{G}}{}^{(t)}\right] \xrightarrow{d(\cdot), k-NN} \left[P_1^t, \ldots, P_N^t\right] \xrightarrow{f(\cdot)} \left[X_{\mathcal{G}}{}^{(t+1)}, \ldots, X_{\mathcal{G}}{}^{(t+T)}\right]$
  - Reformulate **point-wise representation learning** into **pattern-matching, pattern-wise representation learning**
- **To solve reformulated problem, we have to build pattern set and novel deep learning model for the memorization**

Pattern Set Extraction

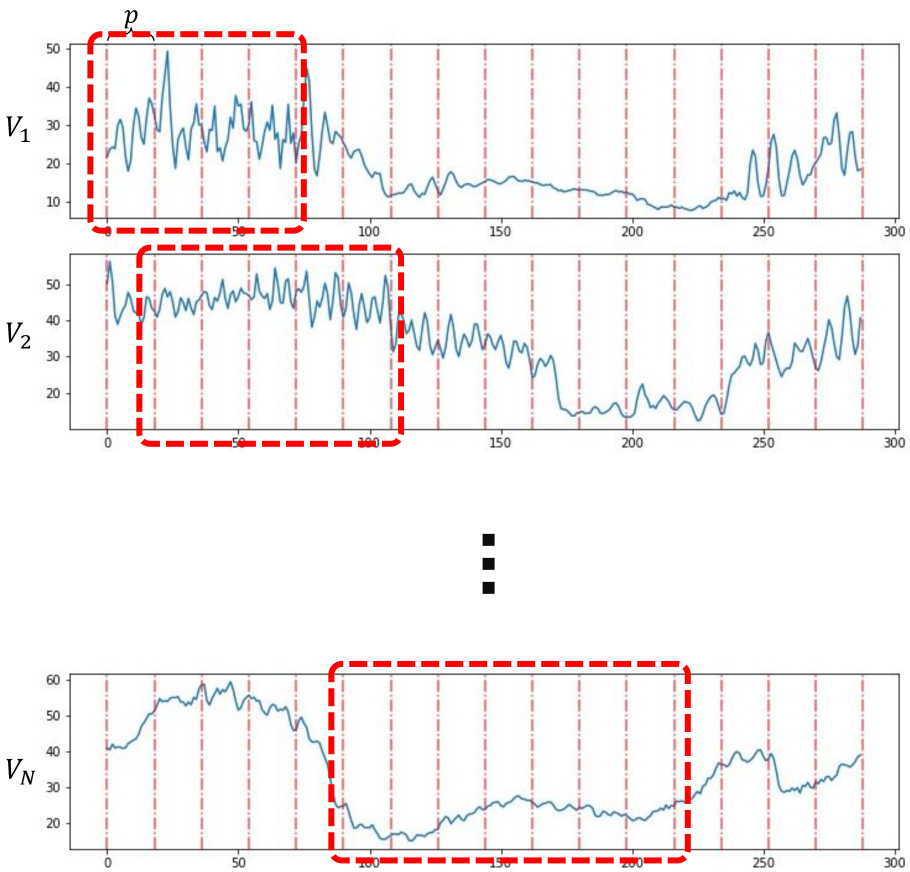
- **Split daily trends into T'-length patterns → prototype patterns**

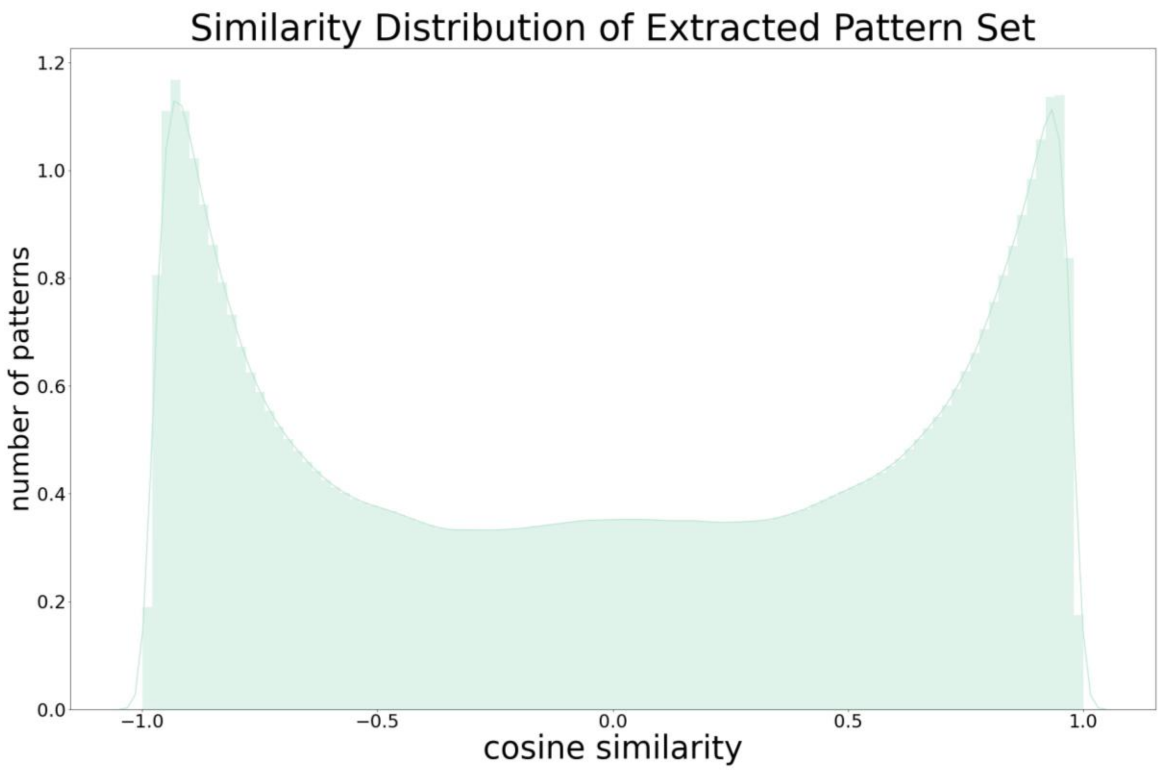




Pattern Set Extraction: Unbiasing

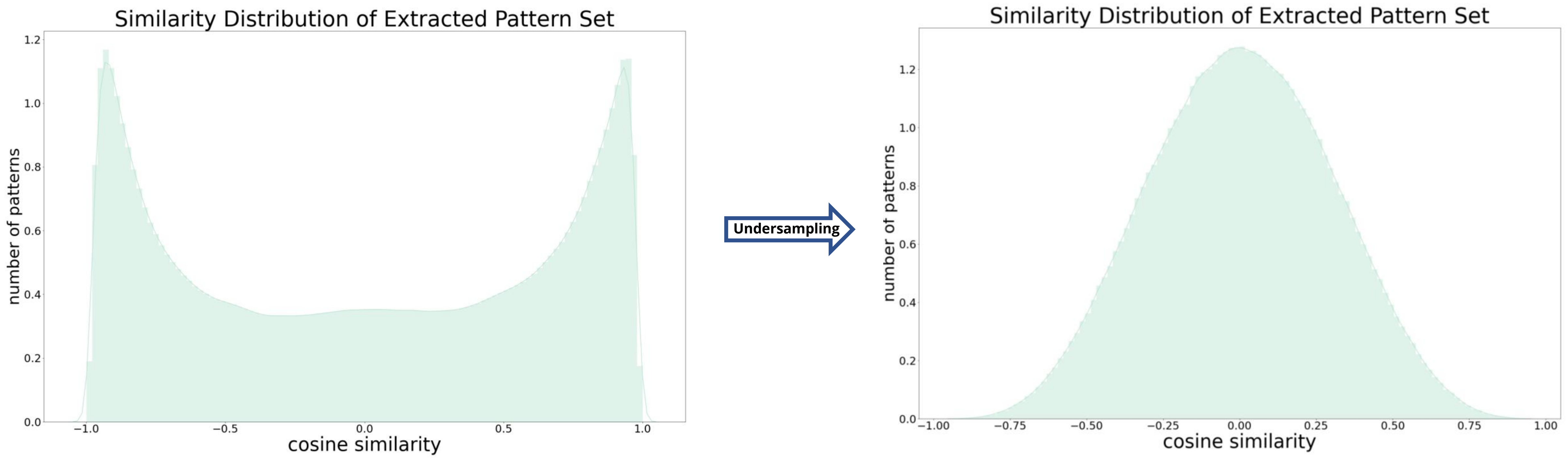
- **Prototype pattern set have unsolved problems**
  - Too many duplicated patterns w/ biased distribution (i.e., class imbalance)
  - Large size ($|P| = N \times \left\lfloor \frac{288}{T'} \right\rfloor > 1000$)





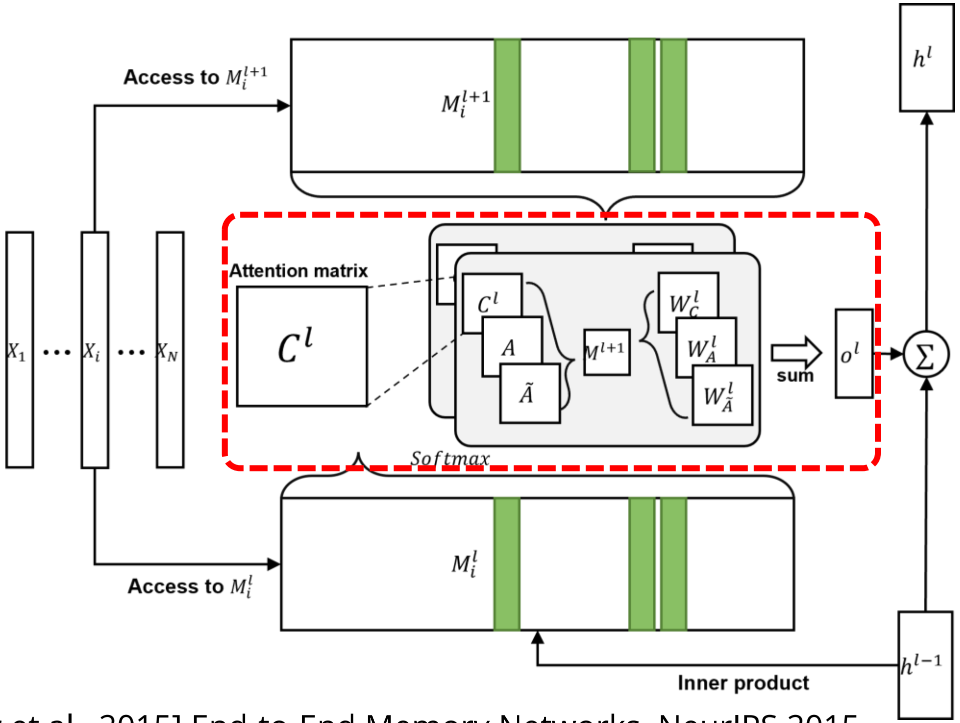
Pattern Set Extraction: Unbiasing

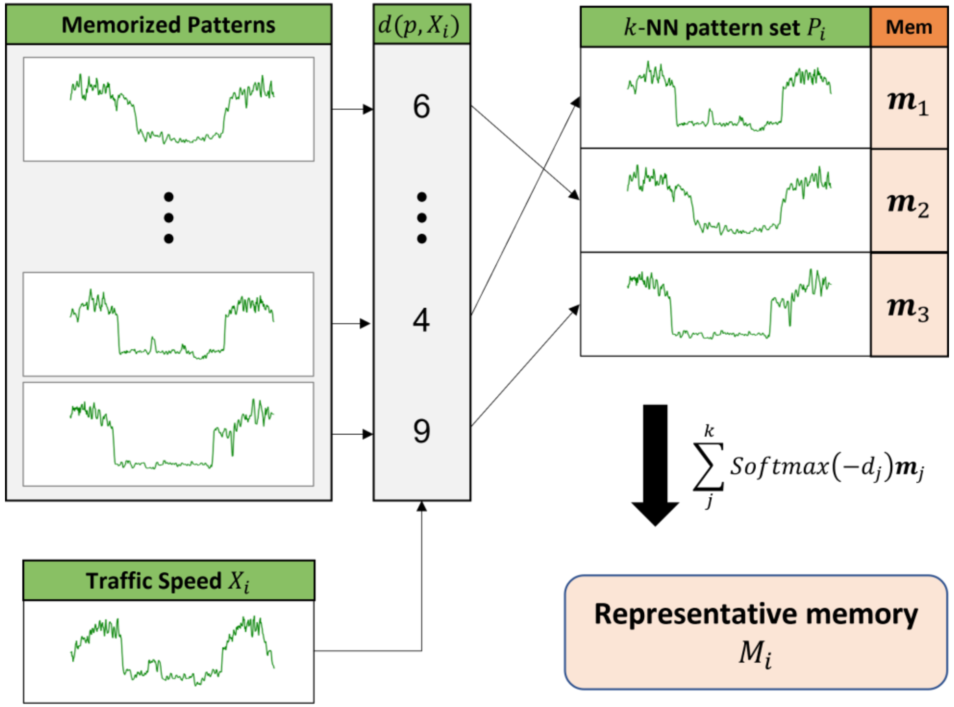
- **By utilizing clustering-based undersampling, we have representative pattern set w/ balanced similarity distribution**



Pattern Matching Memory Networks

- **How to fully utilize representative patterns?**
- **GCMem: a graph-aware memory cell for the traffic prediction**
  - Combine adjacent weight sharing (Sukhbaatar et al., 2015) with **attention and graph convolution**

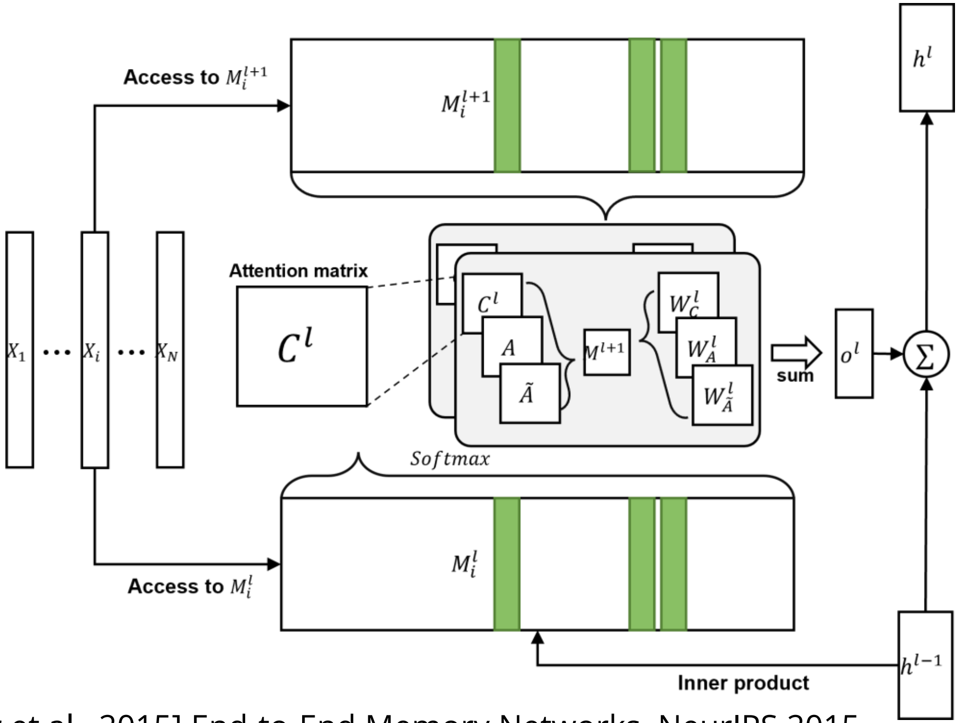


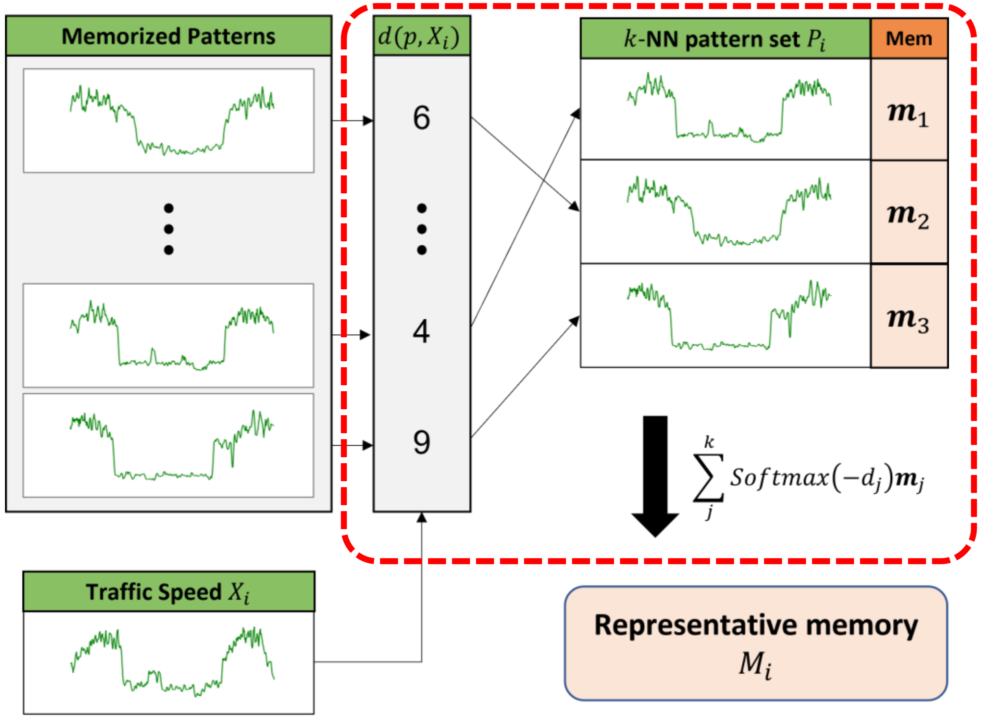


[Sukhbaatar et al., 2015] End-to-End Memory Networks, NeurIPS 2015.

Pattern Matching Memory Networks

- **How to fully utilize representative patterns?**
- **GCMem: a graph-aware memory cell for the traffic prediction**
  - Combine adjacent weight sharing (Sukhbaatar et al., 2015) with **attention and graph convolution**
  - Enhance referring diversity with **k-NN memory access**

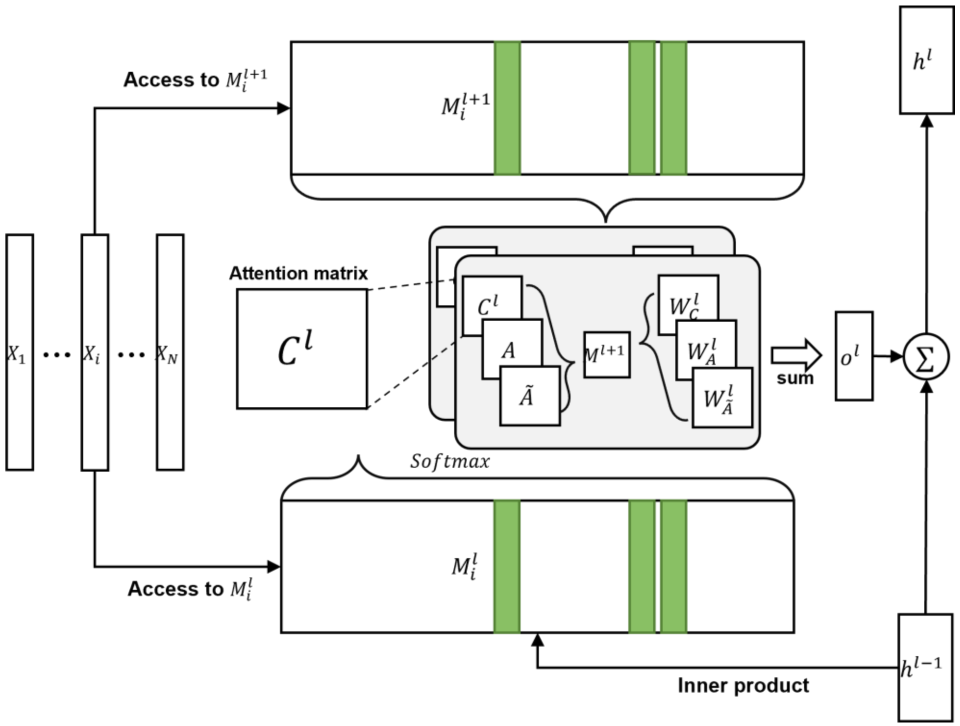


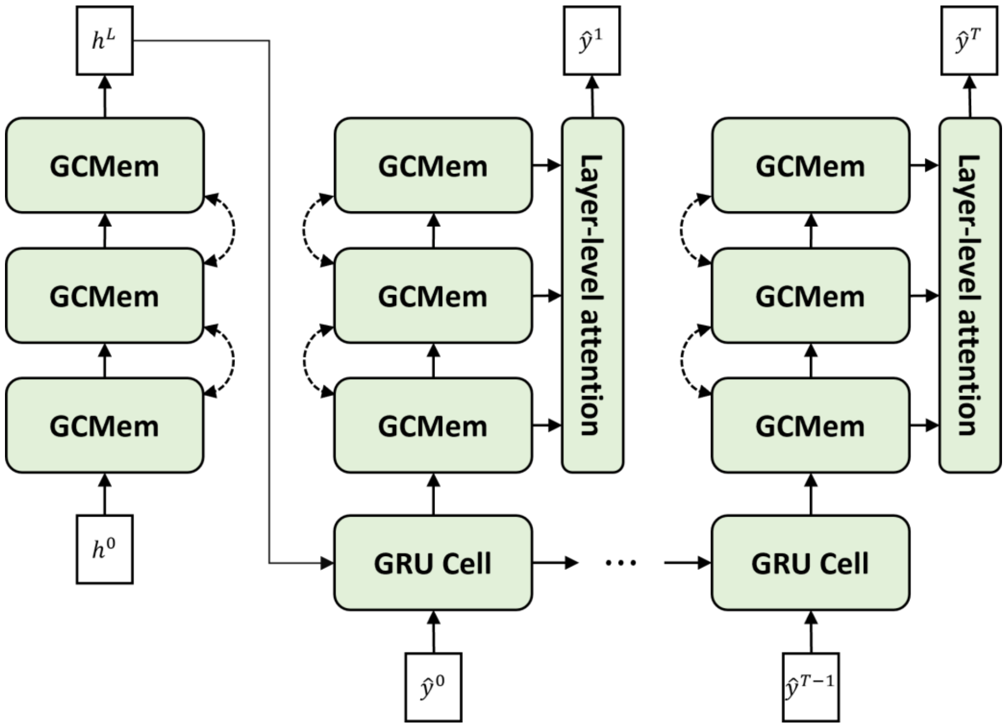


[Sukhbaatar et al., 2015] End-to-End Memory Networks, NeurIPS 2015.

Pattern Matching Memory Networks

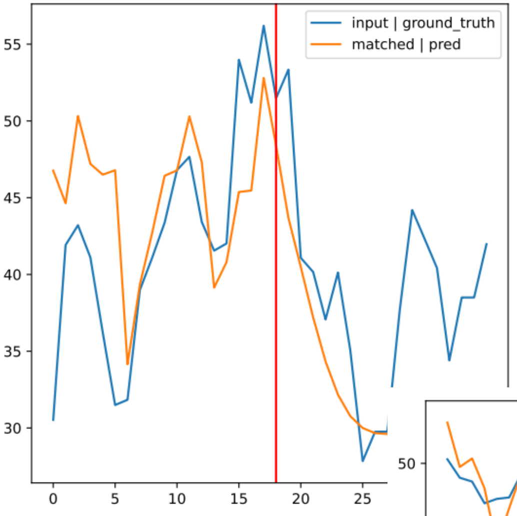
- **PM-MemNet: simple RNN-based forecasting model w/ GCMem**
  - **To handle small noise and matching error, h^0 in encoder is:**
    - $h_i^0 = emb(T) + W_n N_i$, where $N_i = X_i - p_1$
  - **By layer-level attention, model can choose which information will be used for prediction**





Experiments: Performance Evaluation

- **PM-MemNet shown significant improvement across all tasks**
  - Though we utilize **RNN-structure**, PM-MemNet achieves sota performance on **long-term prediction**



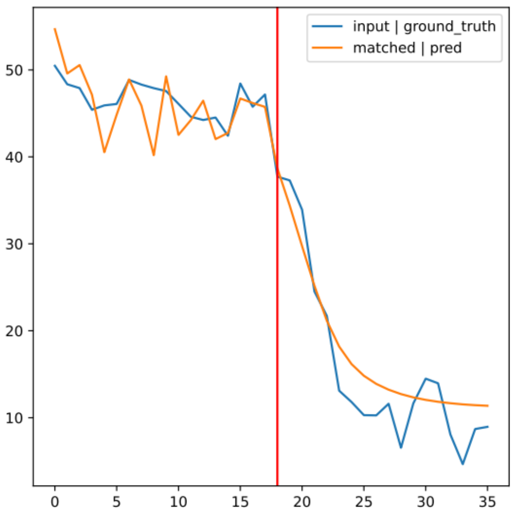


Table 1: Experimental Results for NAVER-Seoul and METR-LA datasets

Dataset	T	Metric	HA	MLP	STGCN	GCRNN	DCRNN	ASTGCN	GWNet	GMAN	PM-MemNet
NAVER-Seoul	15min	MAE	6.54	5.28	4.63	4.87	4.86	5.09	4.91	5.20	**4.57**
		MAPE	18.24	16.86	14.49	15.23	15.35	16.14	14.86	16.98	**14.43**
		RMSE	9.32	7.78	6.92	7.18	7.12	7.44	7.24	8.32	**6.72**
	30min	MAE	7.16	6.13	5.50	5.73	5.67	5.71	5.26	5.35	**5.04**
		MAPE	20.15	20.05	17.37	18.17	18.38	18.78	**16.16**	17.47	16.34
		RMSE	10.18	9.51	8.83	9.03	8.80	8.73	8.13	8.67	**7.86**
	60min	MAE	8.22	7.08	6.77	6.58	6.40	6.22	5.55	5.48	**5.24**
		MAPE	23.37	23.44	20.42	20.95	21.09	20.37	16.97	17.89	**16.94**
		RMSE	11.54	11.13	10.89	10.58	10.06	9.58	8.77	8.94	**8.39**
	90min	MAE	9.24	7.79	8.06	7.14	6.86	6.76	5.87	5.58	**5.40**
		MAPE	26.40	26.08	22.93	22.86	22.74	21.83	17.89	18.18	**17.44**
		RMSE	12.77	12.17	12.86	11.43	10.69	10.32	9.33	9.09	**8.68**
METR-LA	15min	MAE	4.23	2.93	2.61	2.59	**2.56**	3.25	2.72	2.86	2.66
		MAPE	9.76	7.76	**6.59**	6.73	6.67	9.27	7.14	7.67	7.06
		RMSE	7.46	5.81	5.19	**5.12**	5.10	6.28	5.20	5.77	5.28
	30min	MAE	4.80	3.60	3.22	3.08	**3.01**	3.80	3.12	3.14	3.02
		MAPE	11.30	10.00	**8.39**	8.72	8.42	11.28	8.66	8.79	8.49
		RMSE	8.34	7.29	6.63	6.32	6.29	7.59	6.34	6.54	**6.28**
	60min	MAE	5.80	4.69	4.31	3.74	3.60	4.49	3.58	3.48	**3.40**
		MAPE	14.04	13.68	11.13	11.50	10.73	13.69	10.30	10.10	**9.88**
		RMSE	9.86	9.24	8.71	7.71	7.65	8.94	7.53	7.30	**7.24**
	90min	MAE	6.65	5.58	5.41	4.23	4.06	4.97	3.85	3.71	**3.64**
		MAPE	16.37	17.08	13.76	13.49	12.53	15.53	11.39	11.00	**10.74**
		RMSE	10.97	10.52	10.47	8.79	8.58	9.71	8.12	7.71	**7.74**

Experiments: Ablation Study

- **SimpleMem shows importance of spatial modeling in traffic domain**
- **Even with very simple decoder architecture, GCMem outperforms existing works**
- **From the set of experiments, we have proven traffic data can be generalized with small number of representative patterns**

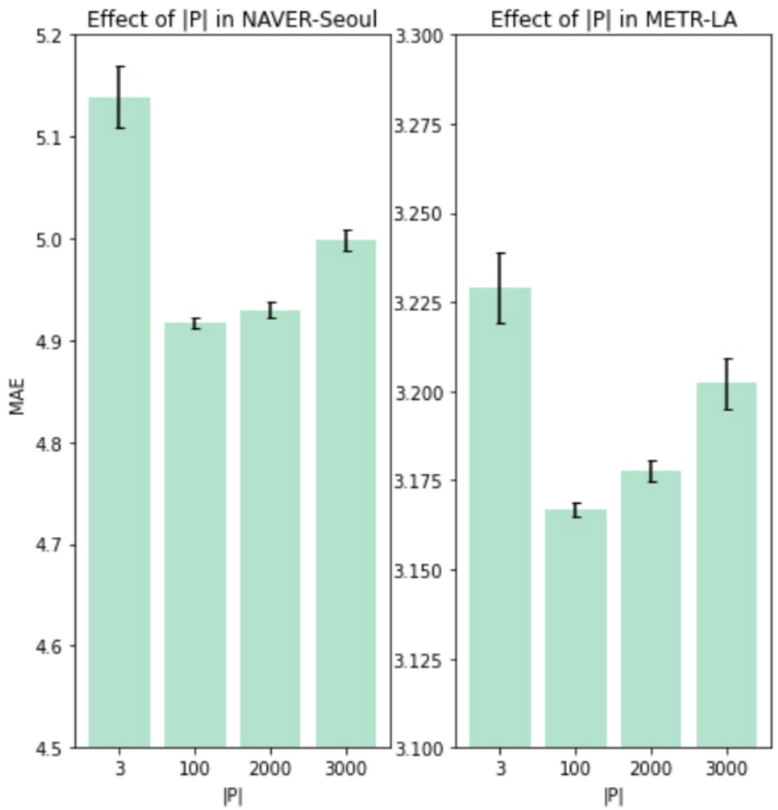
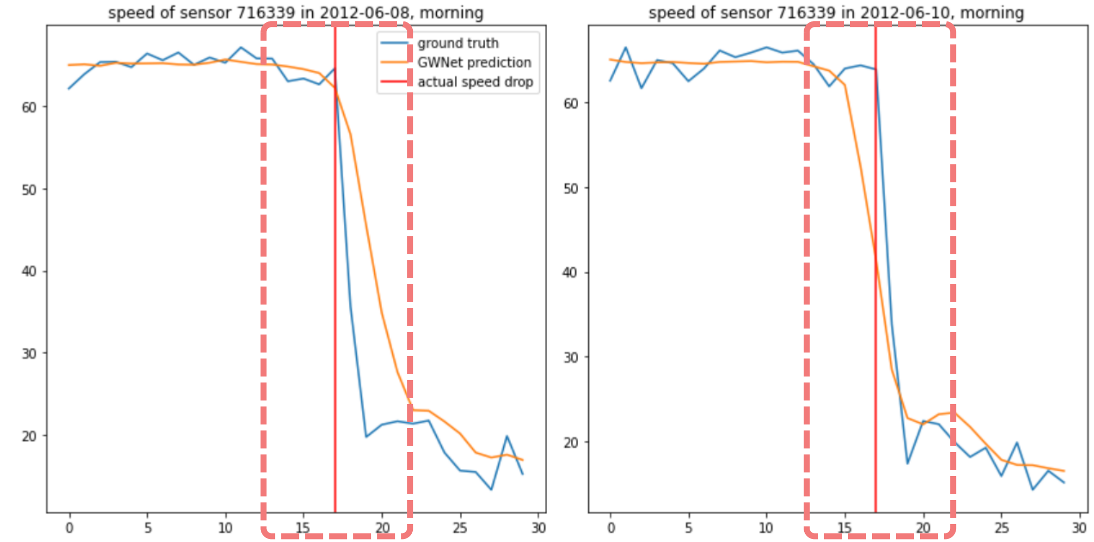


Table 2: Ablation study result. Note that 'Ours' means PM-MemNet.

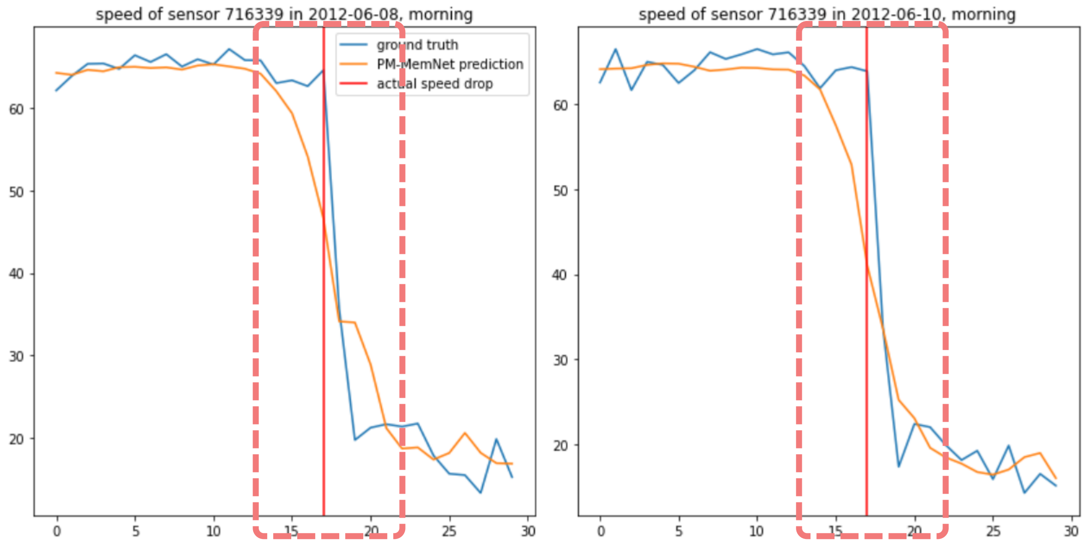
Dataset	T	Metric	Ours	SimpleMem	CNN Decoder	RNN Decoder	Ours (L=1)	Ours ($\lvert\mathbb{P}\rvert = k$)	Ours ($\lvert\mathbb{P}\rvert >> 1000$)
NAVER-Seoul	15min	MAE	4.57	5.72	**4.56**	4.67	4.72	4.66	4.59
		MAPE	14.43	18.18	**14.40**	14.84	14.98	14.77	14.48
		RMSE	6.72	8.79	**6.71**	6.83	6.87	6.89	6.73
	30min	MAE	**5.04**	5.84	5.06	5.19	5.22	5.21	5.09
		MAPE	**16.34**	18.86	16.36	16.87	16.97	16.91	16.41
		RMSE	**7.86**	9.24	7.90	8.02	8.03	8.18	7.97
	60min	MAE	**5.24**	6.38	5.32	5.47	5.52	5.53	5.36
		MAPE	**16.94**	21.42	17.19	17.91	17.97	18.05	17.19
		RMSE	**8.39**	10.08	8.51	8.69	8.70	8.92	8.67
	90min	MAE	**5.40**	6.95	5.55	5.70	5.72	5.74	5.52
		MAPE	**17.44**	23.89	17.99	18.73	18.63	18.73	17.76
		RMSE	**8.68**	10.88	8.82	9.10	9.05	9.31	8.94
METR-LA	15min	MAE	2.66	3.01	**2.63**	2.68	2.68	2.67	2.68
		MAPE	7.06	8.03	**6.98**	7.10	7.11	7.09	7.13
		RMSE	**5.28**	5.94	5.32	5.31	5.31	5.35	5.31
	30min	MAE	3.02	3.27	**3.01**	3.06	3.06	3.06	3.04
		MAPE	8.49	9.20	**8.46**	8.56	8.59	8.67	8.51
		RMSE	6.28	6.68	6.36	6.32	**6.27**	6.36	6.32
	60min	MAE	**3.40**	3.72	3.41	3.46	3.47	3.49	3.45
		MAPE	**9.88**	10.94	**9.88**	10.02	10.07	10.34	9.87
		RMSE	**7.24**	7.70	7.28	7.31	7.25	7.39	7.32
	90min	MAE	**3.64**	4.09	3.65	3.71	3.73	3.75	3.69
		MAPE	**10.74**	12.25	10.85	10.87	10.98	11.30	10.63
		RMSE	7.74	8.38	**7.71**	7.81	7.75	7.91	7.81

Experiments: Qualitative Evaluation

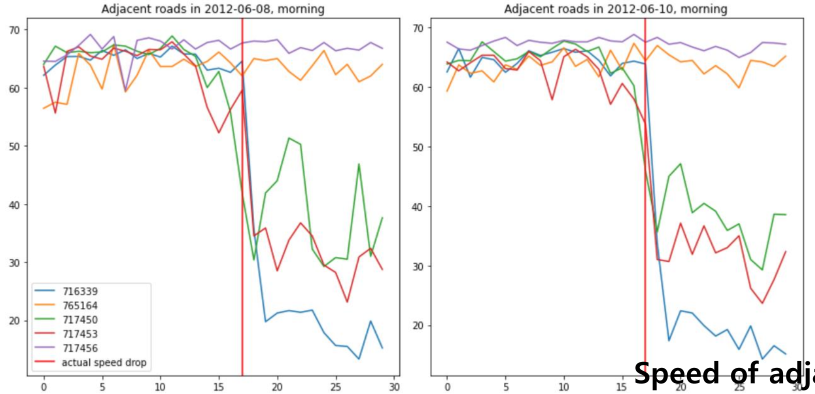
- In contrast to GWNet, PM-MemNet shows much **robust prediction**
- Because of fluctuation, **GWNet forecasts differently** even in **similar circumstances**



GWNet Prediction Results



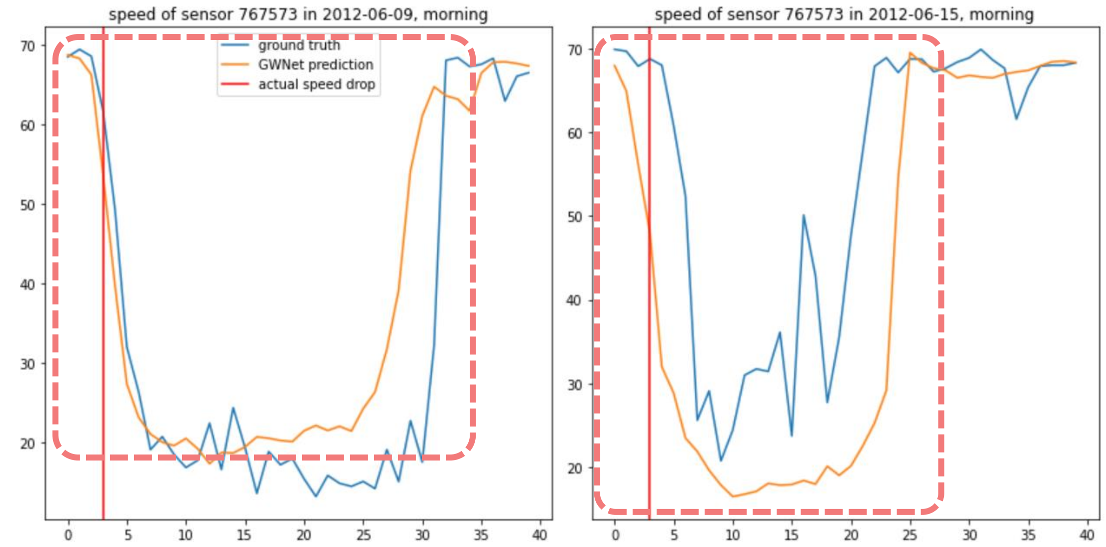
PM-MemNet Prediction Results



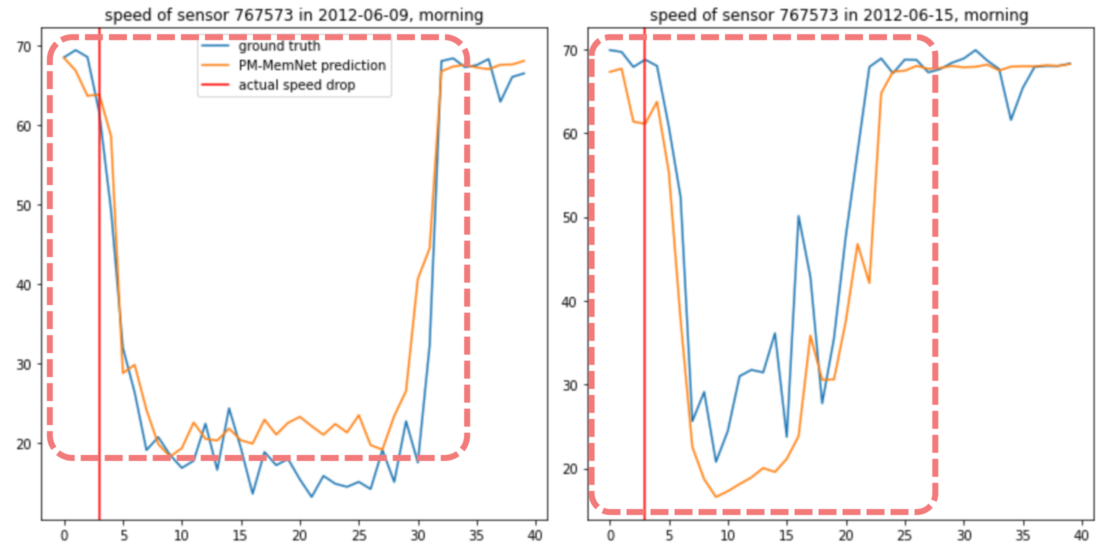
Speed of adjacent roads

Experiments: Qualitative Evaluation

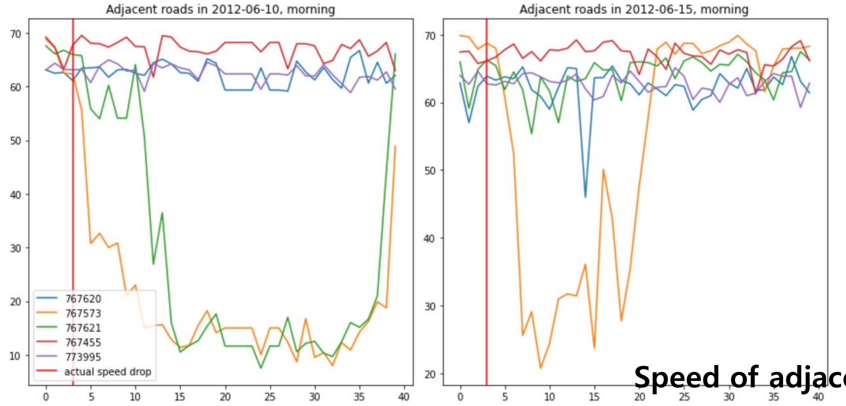
- GWNet cannot deal with situation in right graph
- PM-MemNet shows better adaptation in both situations



GWNet Prediction Results



PM-MemNet Prediction Results



Speed of adjacent roads

Summary

- **Traffic forecasting model can achieve benefits in both robustness and situation-aware forecasting with pattern-matching**
- **We suggests new perspective for traffic forecasting**
  - Reformulate forecasting problem into **pattern matching problem**
  - Design a novel traffic forecasting model **PM-MemNet**
  - Have proven **small set of patterns** can represents **core features**
  - Suggest many possible **research directions for our problem setting**

- **More details can be found:**
  - Paper: https://arxiv.org/abs/2110.10380
  - Code and NAVER-Seoul dataset: https://github.com/HyunWookL/PM-MemNet