Imperial College London

Short-Term Traffic Prediction: Modelling Temporal-Spatial Features in Local Highway Networks with Deep Neural Network

Jingqing Zhang,[1] Chao Wu,[1] and Yike Guo.[1]

[1] Data Science Institute, Imperial College London.

{ jingqing.zhang15, chao.wu, y.guo } @ imperial.ac.uk

Abstract

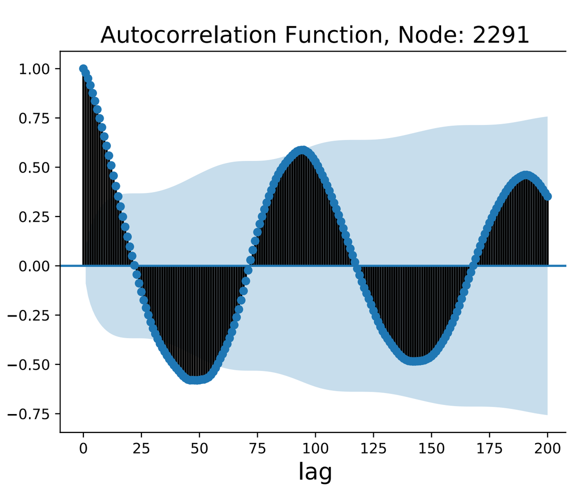
Both temporal and spatial features provide significant implications for short-term traffic volume prediction. The problem is challenging due to various non-linear temporal dynamics at different locations, complicated spatial dependencies and difficulty for longer-step ahead forecasting. We propose two deep learning models, CNN LSTM with attention mechanism (CNN-LSTM-Attn) and Temporal-Spatial-LSTM (TS-LSTM) to incorporate temporal and spatial correlations. Experiments show that both models outperform baselines on the Highways England dataset and the CNN-LSTM-Attn achieves lowest MAPE 9.26% on 2-hour traffic volume prediction. We also evaluate the CNN-LSTM-Attn on the KDDCUP17 dataset and our model defeats the model that got first place in the competition with lower MAPE 10.48%. Our models achieve 2-hour forecasting, which is longer than previous literature, with outstanding accuracy and robustness.

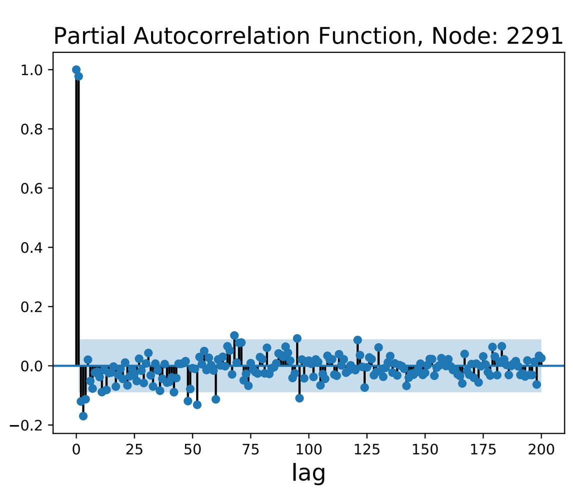
Dataset

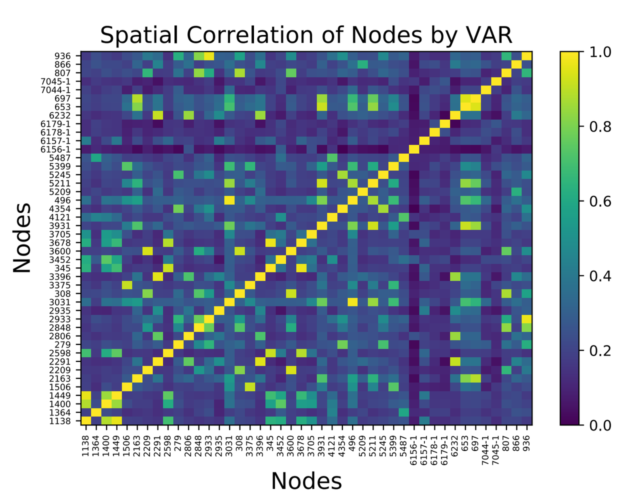
		Start	End	Full Batch Size
Highways England dataset	Train	01/01/2016	30/09/2016	1,136,290
	Valid	01/10/2016	31/12/2016	387,860
	Test	01/01/2017	30/06/2017	814,292
KDDCUP 17 dataset	Train	19/09/2016	10/10/2016	3456
	Valid	11/10/2016	17/10/2016	
	Test	18/10/2016	24/10/2016	14

Data Analysis

	Traffic Volume	Start of Peak Hours	End of Peak Hours
Workdays	0.00%	07:38	19:28
Weekends	10.87% less	11:03	19:56
Bank Holidays	14.76% less	11:24	19:47
School Holidays	0.92% more	07:44	19:17
Christmas Day	45.06% less	10:59	20:48
Christmas Period	20.02% less	11:26	19:05
New Year's Day	35.14% less	12:32	19:16







Problem Definition

Given a sequence of traffic volume observations $\{x_1, x_2, \dots x_t\}$ with a model $\mathcal{M}$ and its parameters θ

$$\tilde{x}_{t+1}, \tilde{x}_{t+2}, \dots, \tilde{x}_{t+H} = \underset{x_{t+1}, x_{t+2}, \dots, x_{t+H}}{\operatorname{argmax}} P(x_{t+1}, x_{t+2}, \dots, x_{t+H} | x_1, x_2, \dots x_t, \mathcal{M}, \theta)$$

If considering spatial correlations, $X_t = [x_t^1; x_t^2; \dots; x_t^N]$ concates N nodes

$$\tilde{X}_{t+1}, \tilde{X}_{t+2}, \dots, \tilde{X}_{t+H} = \underset{X_{t+1}, X_{t+2}, \dots, X_{t+H}}{\operatorname{argmax}} P(X_{t+1}, X_{t+2}, \dots, X_{t+H} | X_1, X_2, \dots X_t, \mathcal{M}, \theta)$$

REFERENCES

[1] Dong, H., Supratak, A., Mai, L., Liu, F., Oehmichen, A., Yu, S., and Guo, Y. (2017a). TensorLayer: A Versatile Library for Efficient Deep Learning Development. ACM Multimedia.

[2] Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT Press.

[3] Lippi, M., Bertini, M., and Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. IEEE Transactions on Intelligent Transportation Systems.

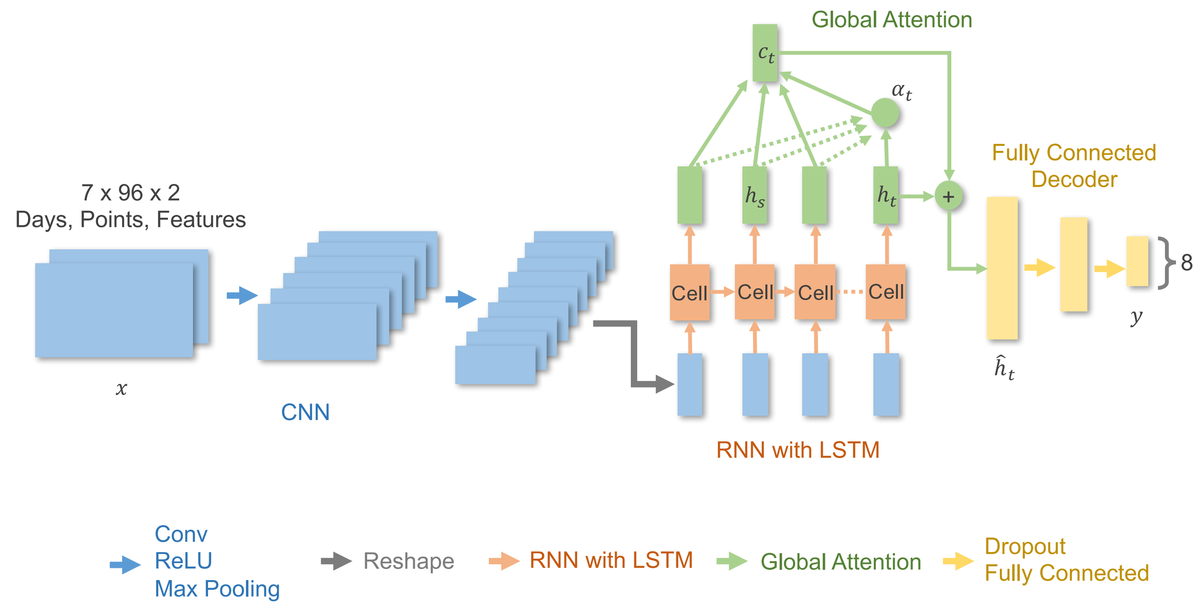
[4] Luong, M.-T., Pham, H., and Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.

[5] Santoro, A., Raposo, D., Barrett, D. G., Malinowski, M., Pascanu, R., Battaglia, P., and Lillicrap, T. (2017). A simple neural network module for relational reasoning. arXiv preprint arXiv:1706.01427.

* More information about the project: https://jingqingz.github.io/projects/2017-09-EnglandHighway

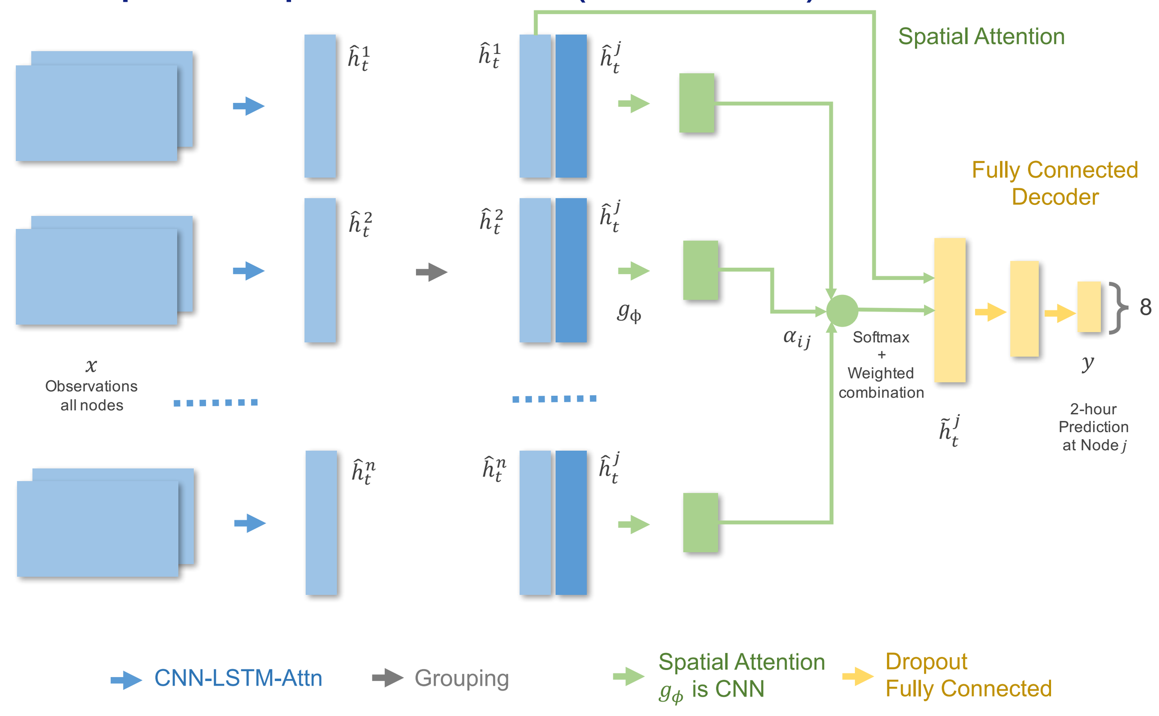
Methodologies

CNN-LSTM with Attention (CNN-LSTM-Attn)



- CNN Embedding
- RNN with LSTM
- Global Attention
  - $\alpha_t(s) = \frac{\exp(score(h_s, h_t))}{\sum_{s'} \exp(score(h_{s'}, h_t))}$
  - $score(h_s, h_t) = h_s^T W_\alpha h_t$
  - $c_t = \sum_s \alpha_t(s) h_s$
  - $\hat{h}_t = [c_t; h_t]$
- Fully Connected Decoder
  - Predict 2-hour together
- One model for the whole highway network

Temporal-Spatial-LSTM (TS-LSTM)

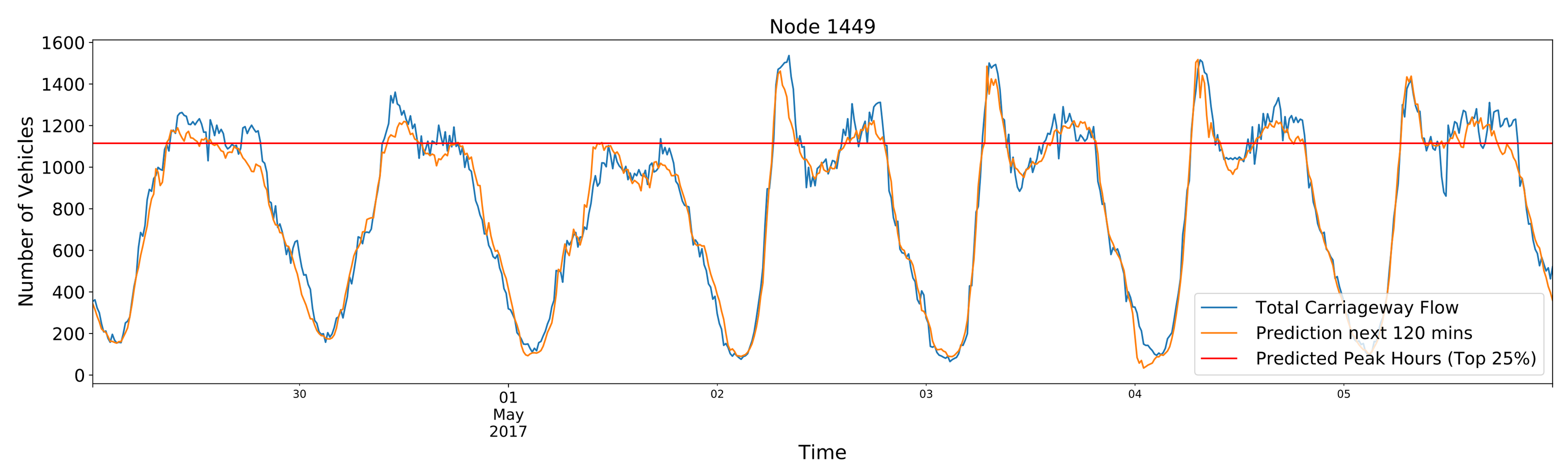


- CNN-LSTM-Attn
  - Applied on all locations
- Spatial Relational Module
  - For a specific location j
  - $\alpha_{ij} = \frac{\exp(g_\phi(\hat{h}_t^i, \hat{h}_t^j))}{\sum_{i'} \exp(g_\phi(\hat{h}_t^{i'}, \hat{h}_t^j))}$
  - $\tilde{h}_t^j = \sum_i \alpha_{ij} \hat{h}_t^i$
- Fully Connected Decoder
  - Predict 2-hour together
- Prior Knowledge not necessary

Experiments

Highways England dataset: 2-hour prediction with error rate (MAPE)

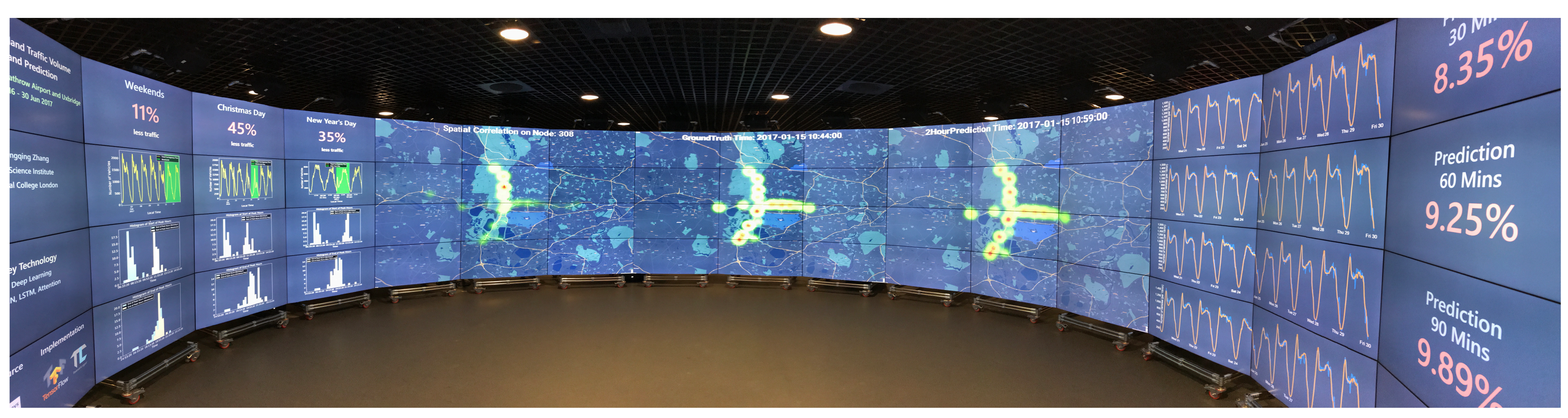
Prediction	SARIMA	VAR	GP	SVR	LSTM	CNN-LSTM-Attn	TS-LSTM
15-min		8.99	12.15	9.48	7.73	**7.54**	8.53
30-min		12.06	19.16	13.23	8.77	**8.35**	8.50
45-min		15.09	25.47	17.01	9.51	**8.88**	9.17
60-min		18.37	31.17	20.94	10.02	**9.25**	9.86
75-min		21.93	36.27	25.14	10.42	**9.60**	10.33
90-min		25.74	40.87	29.47	10.74	**9.89**	10.59
105-min		29.77	45.03	33.97	11.04	**10.12**	11.05
120-min		33.93	48.82	38.64	11.54	**10.44**	12.07
Overall	22.00	20.73	32.37	23.48	9.97	**9.26**	10.01



KDDCUP 17 dataset: 2-hour prediction with error rate (MAPE)

	CNN-LSTM-Attn	The-Model-Got-1st-Place
MAPE	**10.48**	12.03

Visualisation: Global Data Observatory



ACKNOWLEDGEMENTS

Jingqing Zhang would like to thank LexisNexis Corporation for their kindly financial support. We are grateful to Mr. Simiao Yu, Mr. Hao Dong, Mr. Akara Supratak, Dr. David Birch, Dr. Miguel Molina-Solana, Dr. Fangde Liu, Dr. Kai Sun, Mr. Yuanhan Mo, Mr. Xiaoping Fan, Ms. Pan Wang for their kind advice.