Spatial-Temporal Hybrid Deep Neural Networks for Early Congestion Detection

Extended Abstract for hEART 2018

Lin Zhu, Fangce Guo, Rajesh Krishnan, John W. Polak
Centre for Transport Studies, Imperial College London, London, UK

INTRODUCTION

In recent years, increasing attention has focused on the detailed analysis of the spatio-temporal patterns of traffic with the objective of identifying the occurrence and nature of traffic congestion (Zhang et al., 2016). Such work has focused on both modelling congestion formation and propagation (Saeedmanesh and Geroliminis, 2017), as well as clustering traffic congestion with emerging traffic data in urban networks (Anbaroglu et al., 2014). Nevertheless, to the best of our knowledge, little research has been conducted in the early detection of the onset of congestion and how the capability for such early detection might be most effectively used to prevent or mitigate the growth of congestion.

Recurrent Congestion (RC) which usually occurs at peak hours typically with a daily pattern and is caused by excess travel demand relative to the network capacity (Anbaroglu et al., 2014). Reliable early detection of incipient RC can improve the network management response by enabling drivers to divert at an earlier stage in their journey and y giving traffic managers extra time to develop appropriate traffic signal plans and other control measures (Zhang et al., 2014).

One of the main characteristic of urban traffic is that congestion is spatially correlated among adjacent roads and it can spread or dissipate with variable propagation and dissolution rates in both time and space (Saeedmanesh and Geroliminis, 2017). The explicit recognition of this spatiotemporal correlation structure adds complexity to the problem. In this paper, we argue that if properly understood it also enables us to better understand how to (a) diagnose the onset of congestion (b) characterise the temporal evolution of congestion patterns and (c) proactively develop strategies to mitigate urban traffic congestion problems.

The aim of this paper is therefore to develop a hybrid deep learning-based detection methodology that can provide a spatiotemporal understanding of urban congestion and accurate early detection of RC in urban networks. The contributions of this paper are: (1) it presents a novel hybrid deep learning based early detection method that exploits the spatio-temporal structure of urban traffic; (2) the methods developed are capable to detecting multiple levels or severities of congestion with reasonable accuracy in the early stage; (3) the performance of a number of RC early detection methods are compared and evaluated with respect to the size of time windows (i.e., how many prior time steps should be used as inputs) and number of time steps ahead (how early can it detect).

BACKGROUND

The existing literature provides various RC prediction methods and techniques that have been employed to examine traffic data in urban networks. These RC methods consist of two categories, i.e., statistical methods and machine learning methods. Over the last few years, most research studies have attempted to statistically estimate or predict traffic congestion by using traffic flow theory, including car-following model (Gazis et al., 1959) and cell transmission model (Daganzo, 1994) which are quite difficult to be applied to real-case scenarios without optimised environment on a large scale. In addition, researchers have conducted substantial studies on motorways or highways (Yildirimoglu and Geroliminis, 2013), while the application of RC detection in the context of urban networks has been more limited, because of their greater topological and control complexity and their vulnerability to a wider range of sources of interruption (Anbaroglu et al., 2014).

On the other hand, machine learning based RC prediction methods aim to estimate or predict traffic variables, such as traffic flow, travel time and speed in a short-time window as these variables are adopted as important factors of Level of Service (LoS) (Botzow, 1974). For instance, Yu *et al.* (2016) used Back Propagation Neural Networks (BPNN) to detect traffic congestion based on occupancy rate variables and proved that the BPNN was capable of detecting traffic congestion with stable performance. An *et al.* (2016) proposed a three-step RC detection procedure with traffic speed input from GPS-equipped taxi. However, most studies related to RC tend to focus on analysis and prediction of traffic recurrent congestion once it has occurred and over relatively limited prediction horizons, typically a 5-minute or 15-minute time window. This short-term perspective provides little time for traffic operators to proactively formulate and deploy management plans and also can fail to accommodate the communications latency of various traffic sensor devices.

Little attention has been paid to the development of methods that can provide a significant early warning of the formation of congestion and the characteristics of its spatio-temporal evolution. Such a system would substantially reduce the time-constraints affecting traffic operators and provide them with an RC index for proactive reaction. Given the limitation of existing literature mentioned above, new techniques are required to investigate the detection of RC on a large scale network by considering both spatial and temporal correlations.

Convolutional Neural Networks (CNN), initially introduced in 1980 as a derivative of conventional multilayer neural networks, are fundamentally supported by neuron science (Fukushima, 1980). In addition to the fully connected layers found in conventional multilayers, a CNN includes convolutional layers and pooling layers where the convoluted layers are locally connected, and parameters are significantly reduced in the pooling process. The locally-connected convoluted layers enable a CNN to capture complex spatial correlation problems (Krizhevsky et al., 2012), while reducing parameters in the pooling layer which makes a CNN potentially applicable to large-scale traffic network problems (Karpathy et al., 2014). Recently, CNN was used to directly capture spatial traffic features and correlations in the urban traffic network as a whole on a large scale network because of its capability to learn spatial correlations (Ma et al., 2017). However, a common issue of CNN is inefficient to learn the temporal information with time series inputs. The Long Short-Term Memory (LSTM), firstly proposed with the concept of gated recurrent units by Hochreiter and Schmidhuber (1997), has become an effective choice for analysing the sequential data. Intuitively, more information is needed to decide how to integrate the previous information into current decision, so the closely recent information before decision time step t may be not enough and information further back is necessary. The LSTM introduces to connect previous relative information between data points with a large lag and handle long-term dependencies, thus exhibits superior capability for time series analysis (Wu and Tan, 2016). Traditional Recurrent Neural Networks (RNNs) mainly have two issues when dealing with short-term prediction: 1) poor performance with long time spans 2) difficult to find optimal time window size or lags (Ma et al., 2015). LSTM is one of the more practical ways to address these limitations of RNNs, thus LSTM is proposed to capture the temporal information in traffic data. The combination of CNN and LSTM has advantages of extracting spatial information and temporal correlations by using CNN and LSTM respectively. Therefore, in this paper, we propose a novel deep neural network combined CNN and LSTM to address the gaps mentioned above.

METHODOLOGY

This paper proposes a novel method for the automated early detection and alerting of road network congestion, operating over time windows ranging from half an hour to three hours and early detect the congestion states before 1 to 8 time steps ahead. The whole methodology procedure consists of two steps, namely label generation and early detection, where the label is generated using Expectation Maximisation algorithm (Dempster et al., 1977) into congested traffic state and uncongested state, and early detection model consists of a CNN, an LSTM and two fully-connected layers. The methodology framework is shown in Figure 1.

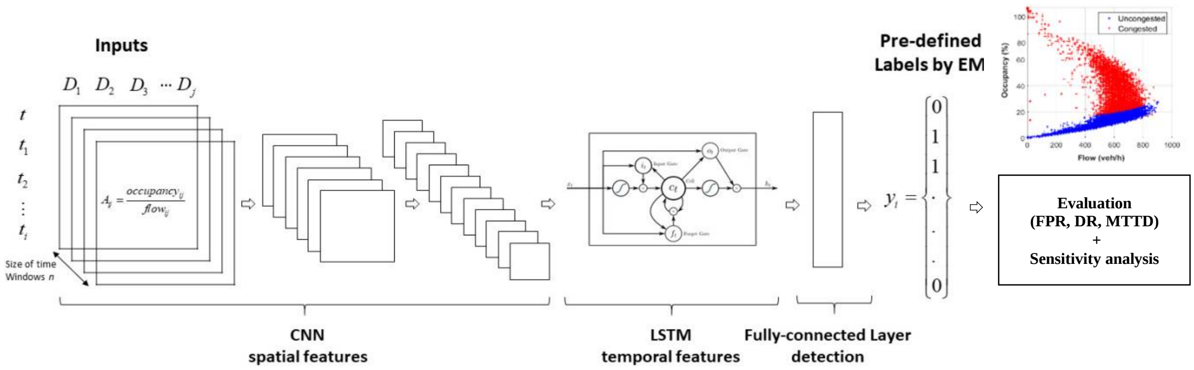


Figure 1. Methodology Framework

EM Label Generation

In the first step, an unsupervised learning method, the Expectation Maximisation (EM) algorithm, is used to classify ("generate labels" in the parlance of the machine learning community) the traffic state according to different levels of severity of congestion. With the assumption that traffic states can be clustered into two regimes by uncongested $(Z_i = 0)$ and congested $(Z_i = 1)$ which follow a Gaussian distribution with $p(\boldsymbol{\alpha})|_{\boldsymbol{z}=0} \sim \mathcal{N}(\mu_0, {\sigma_0}^2)$ and $p(\boldsymbol{\alpha})|_{\boldsymbol{z}=1} \sim \mathcal{N}(\mu_1, {\sigma_1}^2)$ respectively. Then, Gaussian mixture model can be defined by $p(\boldsymbol{\alpha}|\boldsymbol{\Theta}) = \sum_k \gamma_k p_k\,(\boldsymbol{\alpha}|\theta_k)$ where each p_k is a Gaussian distribution function parameterised by θ_k, where $\theta_k = (\,\mu_k, \sigma_k^2)$ and $k = \{0,1\}$. Then by using Bayesian theory and Maximum Likelihood estimation theory, the unknown parameter set $\boldsymbol{\Theta}$ can be estimated with an iteration of log-likelihood expectation step and maximisation step. The EM algorithm has been proved to be an effective and transferable probabilistic traffic state classifier which can capture the latent features of the underlying distribution (Han et al., 2010). It can be viewed as a form of unsupervised learning method in which this context is used to generate labels that can in turn be used to drive a more sophisticated CNN-LSTM-based supervised deep learning method, which is used to detect the early onset of congestion.

Matrix Transformation

In the second step, these labels are used as the input to a CNN-LSTM-based hybrid deep learning early detection model, including a CNN, an LSTM and two fully-connected layers which are designed to capture spatial features and temporal features, and early RC detect respectively. In order to implement the CNN, the time-series traffic flow data have been converted into a 3D time-space feature space where *x-axis*, *y-axis and z-axis* represent time, space and time windows, i.e., the number of time lags, respectively as matrix inputs for a CNN network.

CNN-LSTM Modelling

After converting into a 3D time-space feature space, we define the architecture of the CNN model to extract spatial features 3D time-space matrix. The convolutional layer serves as a detection filter to scan an input with a weighting function w. The convolutional operation can be defined as:

$$S(i,j) = (I \times K)(i,j) = \sum_m \sum_n I(m,n) K(i-m, j-n) \quad (1)$$

where $I(m,n)$ defines the input, and $K(m,n)$ is the convolutional layer kernel or feature map to measure the similarities among input and output a heatmap $S(i,j)$ which represents the region of interest. The following layers gradually detect the more abstracted features and reduce the size of inputs. CNN is capable of capturing patterns in local regions and these abstracted patterns is then fed into an LSTM model.

In addition to RNNs, the key idea of LSTM is the memory cell in hidden layers where errors can flow back forever and make error flow tend to decay exponentially in the whole process from an input gate i_t , a self-recurrent connection neuron c_t, a forget gate f_t to an output gate o_t (Gers, 2001). The mathematical equation detailed this whole LSTM process (i_t, c_t, f_t, o_t) with activation function $\int$ will not be covered in this extended abstract due to the extensive derivatives and readers may refer to Gers's research (2001) for more details. The final set of layers is composed of dropout (Srivastava et al., 2014) and fully-connected hidden layers which make a specific classification based on all features detected by previous layers.

Other Competitors and Performance Evaluation

Other conventional machine learning methods, such as Multilayer Perceptron (MLP) and Random Forest (RF) are used to compare with the proposed CNN-LSTM method in the early detection part. In order to comprehensively evaluate the performance of the proposed models, three evaluation indexes from machine learning classification problems are used. They are False Positive Rate (FPR), Detection Rate (DR) and Mean Time to Detect (MTTD).

RESULTS AND DISCUSSION

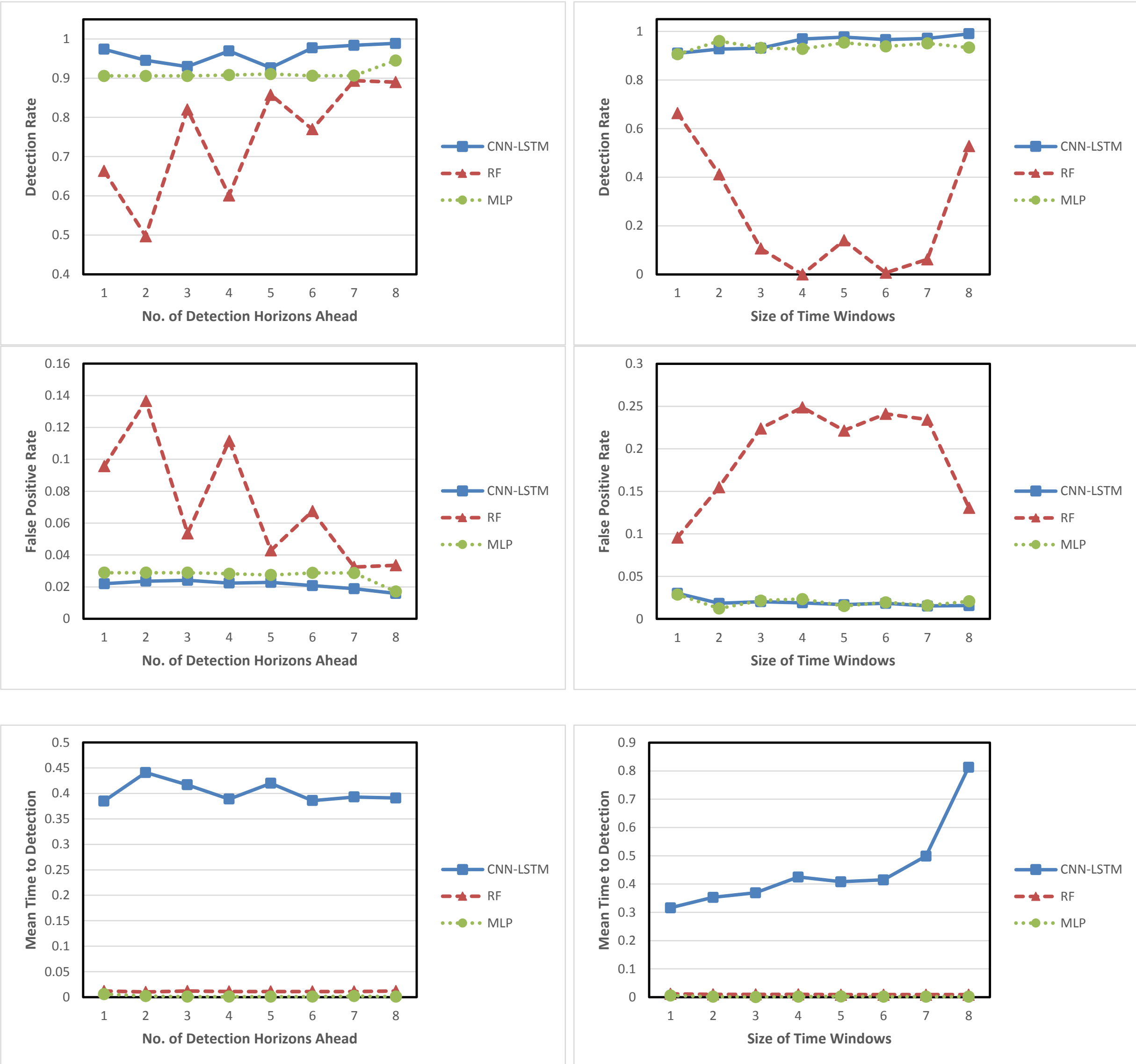
The proposed detection methods are tested using traffic flow and occupancy data from the City of Bath. The case study consists of two main corridors with 18 detectors and 15 detectors respectively. The traffic data which has been pre-processed using the DSA algorithm (Chen, 2003) before feeding into the model covers two years from June 2015 to June 2017 in 15-minute time intervals. The experiment will start with binary labels and find the accuracy corresponding to different size of time windows and number of time steps ahead, and then expand to multiple labels in the future studies.

The preliminary results based on binary labels are as follows. After classifying the traffic states into two regimes based on traffic occupancy and traffic flow by using the EM algorithm, time series data are transformed into matrix data using the transformation method introduced in methodology section. Stochastic gradient methods are used to minimise the loss function and update the weights and bias step by step for MLP and RF. As a result, the MLP is set up with a hidden layers size of (10, 2) and regularisation parameter of 0.00001, while RF is configured to generate 10 decision tree with the depth of 3. The settings of CNN-LSTM net consist of 5 learned layers including 3 convolutional layers with kernel size of 3×3, each followed by a max polling layers with kernel size of 2×2, single layer LSTM as cell with 256 units of the hidden state and two 2 fully-connected layers with 64 and 48 units respectively.

The detailed example of a comparison of CNN, MLP and RF with binary labels, time window size of 3 and 2 time steps ahead on Lower Bristol Road is shown in Table 1. As shown in Table 1, both proposed CNN-LSTM model and conventional machine learning techniques perform well in the early detection method, with low FPR and high DR and Precision. Among three methods, CNN-LSTM slightly outperforms than the other two methods in terms of FPR and DR but suffers from long time for early detection.

Table 1. Performance of Early Detection (No. time window = 3 and No. time steps ahead =2)

	Detection Rate (DR)	False Positive Rate (FPR)	MTTD
CNN-LSTM	97.04%	2.08%	0.383 sec
MLP	94.01%	2.20%	0.001 sec
RF	63.64%	10.52%	0.009 sec



(a) (b)

Figure 2. Sensitivity Analysis in terms of (a) number of detection horizons ahead (b) size of time windows

After analysing performance indexes, sensitivity analysis in terms of the number of detection horizons ahead and size of time windows is conducted to study the impact of these factors to the performance. The result shows CNN-LSTM are quite insensitive to detection horizons but the performance increase with larger size of time windows, which means that CNN-LSTM may result most reliable early detection providing that at least half an hour time window is given. In the next step, more experiments with multiple labels will be conducted to examine the performance of the proposed methods.

CONCLUSION

In this paper, an early detection model based on the combination of unsupervised learning and supervised learning is presented. The proposed model can be used for prediction problems in large urban networks where spatial and temporal correlations are significant factors for the prediction. In order to evaluate the performance of the new approach, we tested it to the prediction of traffic congestion with real traffic data collected in the City of Bath, and compared it with benchmarks, i.e., MLP and RF, in terms of detection rate, false positive rate and mean time to detection. Preliminary results are promising and have demonstrated that CNN-LSTM-based early detection model may be superior to conventional machine learning methods especially with larger size of time window. The longer time window can be practically extracted from the historical dataset and improve detection accuracy of real time applications with different levels of latency. The sensitivity of size of time windows and time steps is varied among three methods, while MLP is insensitive to the number of time steps ahead, which indicates that MLP may accurately detect recurrent congestion even with long time gaps. More experiments with multiple labels will be conducted to examine the performance in the future study.

REFERENCES

An, S., Yang, H., Wang, J., Cui, N., Cui, J., 2016. Mining urban recurrent congestion evolution patterns from GPS-equipped vehicle mobility data. Inf. Sci. (Ny). 373, 515–526.

Anbaroglu, B., Heydecker, B., Cheng, T., 2014. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. Transp. Res. Part C Emerg. Technol. 48, 47–65.

Botzow, H., 1974. Level-of-service concept for evaluating public transport. Transp. Res. Rec. 519, 73–84.

Chen, C., 2003. Detecting Errors And Imputing Missing Data For Single Loop Surveillance Systems. Transp. Res. Rec. J. Transp. Res. Board 1855, 160–167.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp. Res. Part B 28, 269–287.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. R. Stat. Soc. Ser. B 39, 1–38.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36, 193–202.

Gazis, D.C., Herman, R., Potts, R.B., 1959. Car-Following Theory of Steady-State Traffic Flow. Oper. Res. 7, 499–505.

Gers, F., 2001. Long Short-Term Memory in Recurrent Neural Networks. Lausanne, EPFL 2366, 102.

Han, J., Krishnan, R., Polak, J.W., Barria, J., 2010. A New Method for Probabilistic Traffic State Identification Using Loop Detector Data: Theory and Empirical Results, in: UTSG. Plymouth, UK.

Hochreiter, S., Urgen Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Li, F.F., 2014. Large-scale video classification with convolutional neural networks. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 1725–1732.

Krizhevsky, A., Sutskever, I., Geoffrey E., H., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst. 25 1097–1105.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., Wang, Y., 2017. Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17, 818.

Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C Emerg. Technol. 54, 187–197.

Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transp. Res. Part B Methodol. 105, 193–211.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Wu, Y., Tan, H., 2016. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. ArXiv 1–14.

Yildirimoglu, M., Geroliminis, N., 2013. Experienced travel time prediction for congested freeways. Transp. Res. Part B Methodol. 53, 45–63.

Yu, X., Xiong, S., He, Y., Wong, W.E., Zhao, Y., 2016. Research on campus traffic congestion detection using BP neural network and Markov model. J. Inf. Secur. Appl. 31, 54–60.

Zhang, X., Onieva, E., Perallos, A., Osaba, E., Lee, V.C.S., 2014. Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Transp. Res. Part C Emerg. Technol. 43, 127–142.

Zhang, Z., He, Q., Tong, H., Gou, J., Li, X., 2016. Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network. Transp. Res. Part C Emerg. Technol. 71, 284–302.