Using Deep Learning to Predict Short Term Traffic Flow: A Systematic Literature Review

Usman Ali(✉) and Tariq Mahmood

Institute of Business Administration (IBA), Karachi, Pakistan
{uali,tmahmood}@iba.edu.pk

Abstract. This paper systematically reviews Deep Learning-based methods for traffic flow prediction. We extracted 26 articles using a concrete methodology and reviewed them from two perspectives: first, the deep learning architecture used; and second, the datasets and data dimensions incorporated. Recent big data explosion caused by sensors, IoV, IoT and GPS technology needs traffic analytics using deep architectures. This survey reveals that the LSTM (Long Short-Term Memory) Neural Networks are the most commonly used architecture for short term traffic flow prediction due to their inherent ability to handle sequential data. Among the datasets, PeMS is the most commonly used for traffic flow prediction task. Today, Intelligent Transport Systems (ITS) are not limited to temporal data; spatial dimension is also incorporated along with weather data, and traffic sentiments from twitter, Facebook and Instagram to get better results. In the authors' knowledge, this is the first deep learning review in ITS domain.

Keywords: Traffic flow prediction · Deep learning
Intelligent transport systems · Big data

1 Introduction

Traffic control and management is an important issue in urban transport networks. With the advancement in the economy and growth of automobile industry, the number of vehicles on roads is ever increasing. The density of traffic at a road segment may grow large due to road conditions, maintenance work, and weather conditions. This results in traffic jams that create bottlenecks in transportation systems and cause time and money losses for the travelers. To control the traffic flow and avoid congestion, intelligent transport systems have been in use for many years [1, 43, 44]. One focus of these systems is to avoid traffic jams by predicting the traffic flow at given road segments within a specified time interval thereby guiding travelers to avoid possibly congested road segments thus enabling traffic control authorities to control and manage traffic routes effectively.

Traffic flow prediction methods used so far can be categorized into parametric and non-parametric approaches. In parametric approach, the Auto Regressive Integrated Moving Average (ARIMA) model is a well-known framework and benchmark for short term traffic flow prediction [1]. Parametric models predict accurately if the traffic follows regular variations, but the accuracy drops when traffic is irregular due to some

T. Kováčiková et al. (Eds.): INTSYS 2017, LNICST 222, pp. 90–101, 2018.
https://doi.org/10.1007/978-3-319-93710-6_11

unexpected incident on the road, or when prediction is required over a longer interval, e.g., 30 min or more. [2]. These problems shifted research towards non-parametric approaches like non-parametric regression, Artificial Neural Networks, Support Vector Machines (SVM) and Probabilistic methods. These methods being able to model the non-linearity and spatio-temporal relationship in the traffic data, gave better results than parametric approaches. However, they require prior knowledge and considerable effort to extract features and perform pre-processing. With the increasing traffic density and deployment of sensors and cameras, the traffic data has entered big data paradigm. This data explosion causes a problem famously labeled as the curse of dimensionality [45, 46], which traditional parametric approaches are unable to handle effectively. To process and handle big traffic data, deep learning has been recently adapted to learn deep correlations within data; without any or a little prior knowledge and need of hand engineered features [3].

This paper reviews deep learning approaches for short term urban traffic flow prediction. Our objective is to get an insight into the most effective and prevailing deep learning techniques used for short term traffic flow prediction, challenges and related problems, and directions of future research. We pose five research questions: (1) *which deep learning architectures are being used for traffic flow prediction*? (Answered in Sect. 4), (2) *which deep learning method is best suited for traffic flow prediction and most widely used*? (Answered in Sects. 4.1 and 4.2), (3) *which datasets are widely used for traffic flow prediction*? (Answered in Sect. 4.3), (4) *How more data dimensions can be added to achieve better accuracy in traffic flow prediction* (Answered in Sect. 4.3); and (5) *what are the future research directions for improving short term traffic flow prediction*? (Answered in Sect. 5).

2 Research Methodology

We follow a systematic approach [4] to review the literature by framing the research questions (done above), identifying relevant work, studying and assessing the quality of studies, summarizing and interpreting the findings, and suggesting the future work.

2.1 Identifying Relevant Work

To perform the electronic search, we formulated and executed the following six queries on all relevant digital sources, specifically IEEE, Springer, Google Scholar, Elsevier and ACM. *Query 1*: “Deep Learning” AND “Traffic Flow Prediction”, *Query 2*: “Traffic Flow Prediction” AND “Deep Neural Networks”, *Query 3*: “Deep Convolutional Neural Networks” AND “Traffic Flow Prediction”, *Query 4*: “Deep Learning” AND “Urban Traffic Flow Prediction”, *Query 5*: “Traffic Flow Forecasting” AND “Deep Neural Networks” and *Query 6*: “Traffic Flow Forecasting” AND “Deep Learning”.

We used Boolean AND operator to search for both set of keywords in each query. We used “prediction” as well as “forecast” as they are used interchangeably. We also experimented with other keywords, e.g., “Short Term Traffic Flow Prediction” and “Urban Traffic Flow Prediction” with no difference in results from our six queries.

We explicitly searched for Deep Convolutional Neural Networks to filter the rare applications of these networks in traffic flow prediction. We summarized our results in Table 1.

Table 1. No. of publications retrieved after applying necessary filters.

Search query	Search engine					
	IEEE	Springer	Google	Elsevier	ACM	Total
Query 1	10	10	205	58	1	**284**
Query 2	419	1	33	38	15	**506**
Query 3	9	14	48	82	281	**434**
Query 4	16	38	13	155	619	**841**
Query 5	195	3	29	122	281	**630**
Query 6	11	5	103	163	497	**779**
Total	**660**	**71**	**431**	**618**	**1694**	**3474**

2.2 Retrieving Relevant Work

We used Mendeley tool to manage the search results of our queries. Due to broad nature of queries, many retrieved articles were not relevant to deep learning based traffic flow prediction. As most of the deep learning literature has been published after year 2000, no time filter was applied in queries. To filter the retrieved results and retain only the relevant literature, we used a top down approach. In first step, we manually read the title of each article and imported only those into Mendeley that had relevant titles to traffic forecast and neural networks. This gave 185 articles from IEEE, 131 from Springer, 82 from Google Scholar, 50 from Elsevier and 13 from ACM (total 461 articles). This included a lot of articles related to predictions based on shallow neural networks. Being deep learning the focus, we excluded shallow neural network based articles by reading title, abstract and keywords, which gave us 23 deep learning based articles. In third step, we reviewed references of these and found 3 other articles not extracted by our queries. Hence, we reviewed 26 articles published in journals, conferences and library catalogues. The above process is summarized in Fig. 1.

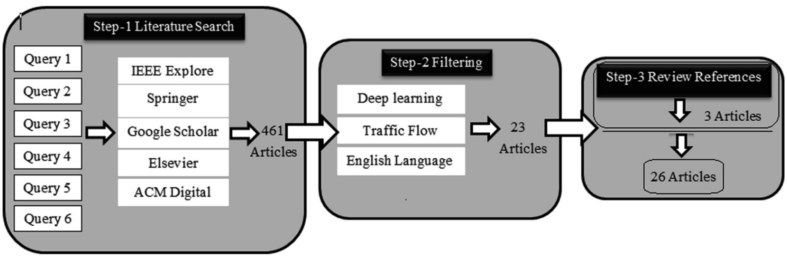

Fig. 1. Literature search flow chart.

We classified our 26 articles per source search engine in Fig. 2. Some articles were retrieved from more than one source; so, we counted them accordingly.

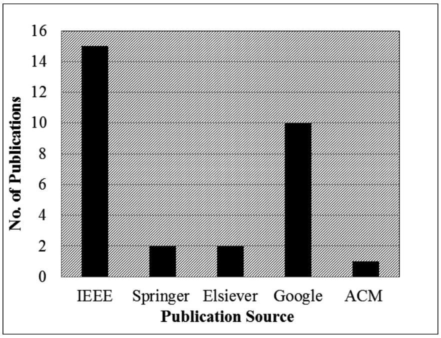

Fig. 2. Publications per source.

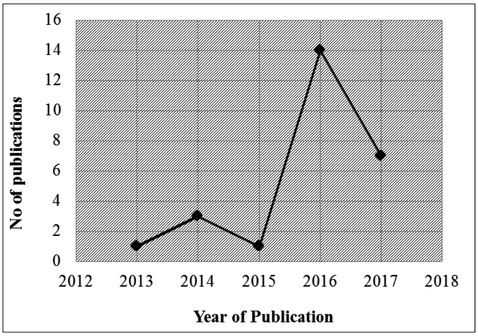

Fig. 3. Growth of research field.

3 Review Statistics

As deep learning is a new approach, all articles are published within the last five years. The distribution of publications per year is shown in Fig. 3. We classified the extracted publications per source. As shown in Table 2, about 46% publications appeared in conferences, 39% in journals and 15% in libraries and archives. Majority of the articles are published in IEEE-related journals and conferences.

Table 2. No. of publications per conference, journal and arvhive.

Source	Retrievals	Percentage
Conference proceedings	12	46%
Scientific journals	10	39%
Archives/libraries	4	15%

We reviewed the articles across the following seven criteria related to our research questions: (1) learning approach (LSTM, Stacked Auto Encoders, Convolutional Neural Networks, Deep Belief Networks), (2) learning algorithm (Back Propagation (BP), Backpropagation Through Time (BPTT), Greedy Layer wise training), (3) data sets, (4) simulation/actual traffic data, (5) time interval for prediction (15, 30, 45 or 60 min), (6) dimensions incorporated (temporal, spatiotemporal) and (7) additional information like online tweets, maps and sentiments. We summarize these findings in Table 3.

Table 3. Overall tabulated results of literature review.

First author	Year	Data location	Dataset	Prediction (Min)	Deep architecture	Algorithm	Data dimensions
Zheng Zhao [2]	2017	China	BTMB [34]	15,30,45,60	LSTM	Greedy Layerwise, Backpropagation (BP)	Spatiotemporal
Wenhao Huang [3]	2013	China, USA	PeMS [7], EESH	15,30,45,60	DBN	Greedy Layerwise, BP	Spatiotemporal
Wenhao Huang [5]	2014	China, USA	PeMS, EESH	15,30,45,60	DBN	BP, MTL	Spatiotemporal
Arief Koesdwiady [8]	2016	USA	PeMS, 16 NWS [35]	15	DBN	Greedy Layerwise, MTL	Spatiotemporal, Weather
Rida Soua [11]	2016	USA	PeMS, NWS, City Pulse [36]	15	DSET [39]	Mass Assignment Algorithm	Temporal, Weather, Twitter
Yuhan Jia [13]	2016	China	BTMB	2,10,30	DBN	Greedy Layerwise	Temporal
Yisheng Lv [14]	2014	USA	PeMS	15,30,45,60	SAE	Greedy Layerwise, BP	Spatiotemporal
Leelavathi [15]	2016	N/A	Simulation	15,30,45,60	SAE	Greedy Layerwise, BP	Temporal
Yanjie Duan [16]	2016	USA	PeMS	15	SAE	Backpropagation	Temporal
Hao-Fan Yang [17]	2016	UK	M6 Freeway	30,60,240,600	SAE(LM)	LM [40]	Spatiotemporal
Yanjie Duan [18]	2016	USA	PeMS	5,10,15,20,50	DSAE	Backpropagation	Spatiotemporal
Xiaolei Ma [19]	2015	China	Microwave	2	LSTM	Truncated BBTT	Temporal
Rose Yu [20]	2016	USA	Los Angeles Highway	5	LSTM	Backpropagation	Temporal
Yaun-yuan Chen [21]	2016	China	AMAP [37]	30	Stacked LSTM	BBTT	Spatio-Online Data
Hongxin Shao [22]	2016	USA	PeMS	15	LSTM	Linear Regression	Temporal
Rui Fu [23]	2016	USA	PeMS	5	LSTM, GRU	BPTT	Temporal
Xiaoguang Niu [24]	2014	China	GPS Data	15	RBM	SVM	Spatiotemporal
Xiaolei Ma [25]	2015	China	GPS Data	5,10,30,60	RBM RNN	Contrastive Divergence	Spatiotemporal
JingYuan Wang 26]	2016	China	GPS Data	5,30	eRCNN	BP, Transfer Learning	Spatiotemporal
Wu Yuankai [27]	2016	USA	PeMS	5	CNN, LSTM	Adamax Optimizer	Spatiotemporal
Felix Kunde [28]	2017	Dresden	VAMOS [38]	5,10,15,30,45	FFNN	Backpropagation	Spatiotemporal
Junbo Zhang [29]	2017	USA, China	GPS Data	30,60	St-ResNet	Deep Residual Learning	Spatiotemporal, Weather, Event
HongSuk Yi [30]	2017	Korea	OBD Data, GPS	cong/non-cong	Tenser Flow DNN	AdaGrad GD	Temporal
Xiaolei Ma [31]	2017	China	GPS Data	10, 20	CNN	Backpropagation	Spatiotemporal
Shiv Surva [32]	2016	Spain	TRANCOS	N/A	CNN	Backpropagation	Spatiotemporal
Nicholas G. Polan [33]	2017	USA	Chicago Interstate	5	Sparce DL	Backpropagation	Spatiotemporal

MTL: Multitask Learning, LM: Levenberg Marquardt, BBTT: Back Propagation Trough Time, N/A: Not Applicable

4 Research Findings

Huang [5] was the first researcher to use deep learning for traffic flow prediction in 2013. He used a Deep Belief Network [6] in the bottom and a regression layer in the top and achieved a 3% improvement over state-of-the-art. He similarly used a temporal Deep Belief Network with Multitask Learning (MTL) [3] on PeMS [7] dataset and achieved a 5% improvement. In 2016, Koesdwiady [8] used a DBN trained with SGD [9] on PeMS dataset and exploited spatio-temporal [10] dimensions of traffic data as well as weather conditions to predict the traffic flow. Soua et al. [11] used a DBN with Dampster Shafer theory [12] with mass assignment algorithm to predict traffic flow using spatio-temporal features of data as well as weather conditions and twitter sentiments about traffic. They used PeMS, NWS [35] and City Pulse [36] dataset for their work. In 2016, Yohan Jia used a temporal DBN with greedy layer wise unsupervised training [47] to predict the traffic flow on the second and third ring roads in Beijing [13]. Moreover, Yishang Lv in 2014 used Stacked Auto Encoders (SAE) with greedy layer wise unsupervised training on PeMS dataset to predict traffic flow [14]. In 2016, Leelavathi used SAE on spatio-temporal simulation data [15]. Yanjie Duan used SAE with back propagation on PeMS dataset [16]. Hao-Fan Yang used SAE Levenberg Marquardt (LM) [40] on spatio-temporal data of M-6 freeway (UK) with greedy layer wise training to predict traffic flow [17]. In 2016, Yanjie Duan used Denoising Stacked Auto Encoders (DSAE) with back propagation to forecast traffic flow using PeMS dataset [18].

Long Short-Term Memory (LSTM) is an extension of Recurrent Neural Networks capable of incorporating long temporal sequences greatly minimizing the vanishing gradient problem found in RNN with long sequences. In 2015, Xiaolei Ma and Zhimin Tao used LSTM to predict traffic speed using microwave detectors data in Beijing [19]. In 2016, Rose Yu used a Deep LSTM to predict traffic flow in extreme conditions using a real-world large-scale dataset in Los Angeles [20]. Yuan-yuan Chen used stacked LSTM with BBTT [41, 42] on PeMS dataset coupled with online data to predict traffic flow [21]. Also, Hongxin Shao used encoder-decoder based LSTM on PeMS dataset [22]. In 2016, Rui Fu used LSTM with Gated Recurrent Units (GRU) and BPTT on PeMS dataset [23]. In a most recent research paper in 2017, Zheng Zhao used LSTM on PeMS dataset and exploited spatio-temporal correlations using an ODC matrix to achieve state of the art results [2].

Some other related work includes the use of Restricted Boltzmann Machine (RBM), ST-ResNet, Tensor Flow based DNN, and Image based CNN and some hybrids of LSTM and CNN. Niu et al. used RBM for dimension reduction and fitting the non-linearity of traffic data distribution by minimizing the energy function [24], and Xiaolei Ma used a hybrid RNN-RBM to predict traffic congestion based on GPS data [25]. Moreover, Wang used Error feedback Recurrent Neural Network (eRCNN) with SGD and transfer learning on traffic data of Beijing city ring road 1 and 2 with spatio-temporal components [26]. Also, Yaunkai used a hybrid of LSTM and CNN with Adam optimizer on PeMS dataset [27] and Kunde used FFNN with SGD on spatio-temporal VAMOS data set of Dresden city [28]. In 2016, Zhang used St-ResNet on spatio-temporal, weather and event data using Beijing Taxicabs, NYC and Bike trajectories data for traffic flow prediction [29]. In 2017, Yi used Tensor Flow DNN to classify road segments as congested and non-congested [30]. Ma and Dai [31] and Tao and Babu [32] used CNN for traffic flow prediction while Polson used a sparse deep learning architecture using L1 regularization and a sequence of tanh layers to predict traffic flow and a sequence of tanh layers to predict traffic flows at two special events; a Chicago Bears football game and an extreme snow storm event [33]. In the next section, we describe the findings about deep architectures used for traffic flow prediction.

4.1 Deep Learning Architectures

Most prevailing deep learning techniques for traffic flow prediction are based on Long Short-Term Memory Networks (LSTM), Deep Belief Networks (DBN) and Auto Stack Encoders (SAE). Some hybrid techniques are also being used to achieve a high level of accuracy. Figure 4 shows the distribution of papers according to the used deep learning technique. Due to space limitations, we don't describe the deep learning architectures and algorithms in detail, but refer the reader to [41, 42].

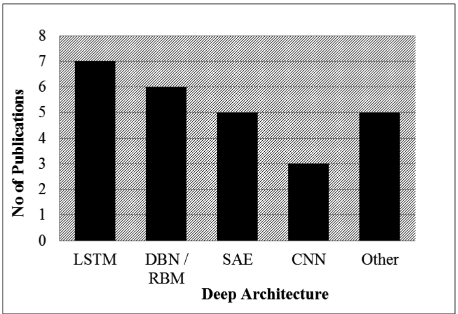

Fig. 4. Publications per DL architecture.

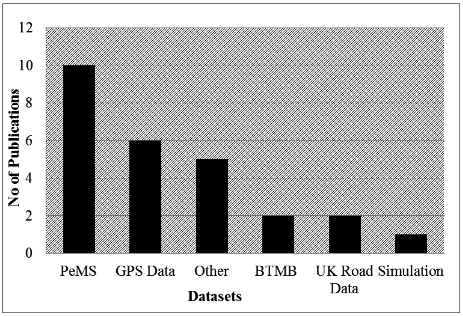

Fig. 5. Publications per dataset.

LSTMs are widely used due to the nature of traffic data. Traffic data is time series in nature; the traffic flow at a particular time on a particular location depends on the prior traffic flow at some earlier point on the same road. Traditional neural networks cannot model this situation, as all inputs and outputs are independent so adjacent layers are fully connected and there is no connection between the nodes of the same layer. Recurrent Neural Networks can model the sequential traffic flow by employing a feedback mechanism in hidden layer neurons from the previous state to current state. However, RNN suffer from vanishing gradients problem for long sequences. LSTM is an extension of RNNs that overcomes the problem of vanishing gradients [41, 42].

Deep Belief Networks (DBNs) are the second most widely used deep architecture. A DBN is formed by a stack of Restricted Bolztmann machines (RBMs) which are trained by using unsupervised greedy layerwise algorithm [47] followed by a supervised fine tuning using backpropagation [41]. As traffic flow at a point is a function of traffic flow at some earlier time at different linked roads, DBNs have an advantage of employing Multitask Learning (MTL) which allow learning several tasks together, e.g. incorporating traffic data of different roads from different related sources, which considerably improves performance [5].

Stacked Autoencoders (SAEs) are the third most widely used architecture which mostly incorporate greedy layerwise mechanism with backpropagation fine tuning [14, 15] or simply backpropagation [16, 18] for temporal as well as spatiotemporal traffic prediction.

4.2 Performance Comparison

This section provides a brief performance comparison of LSTMS, DBNs and SAEs in traffic forecast. Different accuracy measures used are Mean Absolute Error (MAE) [48], Mean Relative Error (MRE) [2], Root Mean Square Error (RMSE) [49], Mean Absolute Percentage Error (MAPE) [50], and precision [21] and accuracy [25]. We have made comparisons where same evaluation measure and time interval for prediction is used.

From survey results, it is found that DBNs are better at incorporating weather and sentiments data with spatiotemporal traffic data for traffic flow forecast [8, 11] with an average RMSE of 0.065 as compared to a temporal LSTM [22] with RMSE 2.51 and a temporal SAE [16] with 61.30 RMSE for 15 min prediction. A temporal DBN in [13] gives comparable results with MAPE 0.084 to a temporal LSTM in [20] with MAPE 0.081 for 30 min prediction and beats an SAE [15] with RMSE margin of 2.69. A spatiotemporal LSTM in [2] gives better MRE than a spatiotemporal SAE in [14] for 15 min prediction task. A spatiotemporal hybrid LSTM-CNN in [27] gives better MAE than a spatiotemporal DBN in [3] for 5 and 15 min prediction task respectively.

Table 4. Strength & weaknesses of deep learning architectures.

Architecture	Strengths	Weaknesses	Solution
LSTM	Best suited for temporal as well as spatiotemporal traffic data	Performance degrades in case of un expected events	Use a hybrid of Bayesian and LSTM network
DBN	Able to incorporate more dimensions like weather data and traffic sentiments	A proper integration model is needed to incorporate more dimensions	Use Multitask learning and incorporate more processing power
SAE	Can handle non-liner spatial and temporal data effectively. Can be trained greedy layerwise with supervised fine tuning	For higher accuracy, more auto encoders are needed requiring more processing time	Select a suitable training and fine tuning algorithm that gives good results with less autoencoders
CNN	Good for spatial traffic data. Surveillance cameras can be used to retrieve image data	Temporal features extraction from images is difficult	Use hybrid of CNN and LSTM to model spatiotemporal data

From the results it can be concluded that DBNs can incorporate more dimensions to give better results as compared to LSTMs and SAEs. When only temporal or spatiotemporal dimensions are used, LSTMS perform better than other two methods (Table 4).

4.3 Traffic Datasets

Our review shows that the PeMS dataset has been used most frequently for short term traffic flow prediction (Fig. 5). Apart from spatiotemporal traffic data, a number of traffic flow prediction techniques are based on weather data [8, 11, 29, 35], and online open data like twitter sentiments about traffic [11, 21, 36, 37]. Specifically, with the availability of more processing power, traffic flow prediction methods are not limited to only temporal domain; spatiotemporal features with weather conditions, sentiments from social websites and related events that affect traffic flow are also being incorporated to get a better prediction result.

PeMS has become a benchmark to test the accuracy of a given model. PeMS data is obtained from Caltrans Performance Management System. Data is collected from more than 40,000 loop detectors located on freeways spanned over all metropolitan areas of state of California. Data is collected after each 5 min in terms of number of vehicles passing through a particular loop detector and aggregated into 15-min periods as suggested by the Highway Capacity Manual.

With the advancement of GPS technology, IoT and IoV, traffic data can be directly obtained from vehicles. A GPS equipped vehicle records time and space information and thus travel speed can be directly measured [25]. If the average speed on a particular road segment is below some threshold, the road is considered to be congested. Images obtained from traffic monitoring cameras can also be used to extract spatiotemporal traffic data [27, 31, 32].

Traffic flow on a particular road is also affected by events like accidents, maintenance work, social events and weather conditions [8, 11, 29]. The spatiotemporal data needs to be coupled with such type of traffic influential information. The data from social websites like twitter, Facebook and Instagram can be used to incorporate the impact of a sudden or un-predicted event [11, 21].

5 Conclusion and Future Research Directions

This survey paper has systematically reviewed deep learning techniques for prediction of short term traffic flow. Deep learning architectures can model the non-linear behavior of traffic data and incorporate both temporal and spatial information supplemented with weather data and other traffic affecting events for traffic data analysis and prediction. At present, Long Short-Term Memory (LSTM), GRU LSTMs, Stacked Auto Encoders, Dual Stacked Auto Encoders, Restricted Boltzmann Machines, Deep Belief Networks, and Convolutional Neural Networks are the most commonly used deep learning techniques and give competitive results. As apparent from the survey results, LSTM and DBN are most widely used and CNN are less frequently used for traffic flow forecast.

To achieve better forecast results, we suggest three future research directions. *First*, a hybrid of Deep CNN and LSTM is proposed. CNNs are best at object detection and LSTM are good at handling sequential data. Spatio-temporal traffic data can be collected by already deployed surveillance cameras in the form of images and analyzed using image processing techniques in real time. Then an LSTM network can be used at the top to take this stream of data and forecast the traffic for the next time interval. *Second,* a hybrid of deep architectures can be used to model different types of traffic data. Apart from spatio-temporal data, useful information about traffic is available in the form sentiments on twitter, Facebook and Instagram and also in the form of text messages and telephone calls. LSTMS are good at categorizing text and speech and can be used for analyzing traffic sentiments in real time. LSTM architecture modeling traffic sentiments can be combined with other deep architectures like CNN or DBN modeling spatio-temporal traffic data. *Third,* there are some events like weather change, road accident, maintenance work or any other social event that affects the traffic flow. Among these, weather conditions, maintenance work and social events can be known in advance but road accidents can't be predicted. Even with weather conditions, maintenance work and the social event, there is always an uncertainty factor. To model this un-certainty, a bayesian network model can be used to incorporate the impact of such traffic affecting events in conjunction with other deep architectures.

References

1. Williams, B.M., Hoel, L.A.: Modeling and forecasting vehicular flow as a seasonal ARIMA process: theoretical basis and empirical results. J. Transp. Eng. **129**(6), 664–672 (2003)
2. Zhao, Z., Chen, W.: LSTM network: a deep learning approach for short-term traffic forecast. IET Intell. Transp. Syst. **11**(2), 68–75 (2017)
3. Huang, W., Hong, H., Li, M., Hu, W., Song, G., Xie, K.: Deep architecture for traffic flow prediction. In: ADMA, pp. 165–176 (2013)
4. Khalid, K.S., Kunz, R., Kleijn, J., Antes, G.: Five steps to conducting a systemic review. J. R. Soc. Med. **96**(3), 118–121 (2003)
5. Huang, W., Song, G., Hong, H., Xie, K.: Deep architecture for traffic flow prediction: deep belief networks with multitask learning. IET Intell. Transp. Syst. **15**(5), 2191–2201 (2014)
6. Hinton, G.E.: Scholarpedia (2009). www.scholarpedia.org/article/Deep_belief_networks
7. PeMS Data Source - Caltrans - State of California. http://www.dot.ca.gov/trafficops/mpr/source.html
8. Koesdwiady, A., Soua, R., Karray, F.: Improving traffic flow prediction with weather information in connected cars: a deep learning approach. IEEE Trans. Veh. Technol. **65**(12), 9508–9517 (2016)
9. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Lechevallier, Y., Saporta, G. (eds.) Proceedings of COMPSTAT'2010. Physica-Verlag HD, Heidelberg (2010)
10. Wei, L.-Y., Peng, W.-C., Lin, C.-S., Jung, C.-H.: Exploring spatio-temporal features for traffic estimation on road networks. In: Mamoulis, N., Seidl, T., Pedersen, T.B., Torp, K., Assent, I. (eds.) SSTD 2009. LNCS, vol. 5644, pp. 399–404. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02982-0_28
11. Soua, R., Koesdwiady, A., Karray, F.: Big-data-generated traffic flow prediction using deep learning and Dampster-Shafer theory. In: IEEE IJCNN, pp. 3195–3202 (2016)

12. Beynon, M., Curry, B., Morgan, P.: The Dampster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling. OMEGA **28**, 37–50 (2000)
13. Jia, Y., Wu, J., Du, Y.: Traffic speed prediction using deep learning method. In: 19th IEEE International Conference on Intelligent Transportation Systems, Brazil, pp. 1217–1222 (2016)
14. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.: Traffic flow prediction with big data: a deep learning approach. IEEE Trans. Intell. Transp. Syst. **16**(2), 865–873 (2014)
15. Leelavathi, M.V., Sahana Devi, K.J.: An architecture of deep learning method to predict traffic flow in big data. IJRET **5**(4), 461–468 (2016)
16. Duan, Y., Lv, Y., Wang, F.: Performance evaluation of the deep learning approach for traffic flow prediction at different times. In: IEEE International Conference on Service Operations and Logistics, and Informatics, Beijing, China, pp. 223–227 (2016)
17. Hao-Fan, Y., Dillon, S.: Optimized structure of the traffic flow forecasting model with a deep learning approach. IEEE Trans. Neural Netw. Learn. Syst. 1–11 (2016)
18. Duan, Y., Lv, Y., Liu, Y.-L., Wang, F.-Y.: An efficient realization of deep learning for traffic data imputation. Transp. Res. Part C Emerg. Technol. **72**, 168–181 (2016)
19. Ma, X., Tao, Z.: LSTM neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. Part C **54**, 187–197 (2015)
20. Yu, R., Li, Y.: Deep learning: a generic approach for extreme condition traffic forecasting. http://roseyu.com/Papers/sdm2017.pdf
21. Chen, Y.-y., Lv, Y., Li, Z., Wang, F.-Y.: Long short-term memory model for traffic congestion prediction with online open data. In: IEEE 19th International Conference on Intelligent Transportation Systems, pp. 132–137 (2016)
22. Shao, H., Soong, B.-H.: Traffic flow prediction with Long Short-Term Memory Networks (LSTMs). In: IEEE Region 10 Conference, pp. 2986–2989 (2016)
23. Fu, R., Zhang, Z., Li, L.: Using LSTM and GRU neural network methods for traffic flow prediction. In: 31st Youth Academic Annual Conference of Chinese Association of Automation, pp. 324–328 (2016)
24. Niu, X., Zhu, Y., Zhang, X.: DeepSense: a novel learning mechanism for traffic prediction with taxi GPS traces. In: 2014 IEEE Global Communications Conference, Austin, TX, pp. 2745–2750 (2014)
25. Ma, X., Yu, H.: Large scale transportation network congestion evolution prediction using deep learning theory. PLoS ONE. https://doi.org/10.1371/journal.pone.0119044
26. Wang, J., Gu, Q., Wu, J., Liu, G., Xiong, Z.: Traffic speed prediction and congestion source exploration: a deep learning method. In: IEEE 16th International Conference on Data Mining, pp. 499–508 (2016)
27. Wu, Y., Tan, H.: Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework (2016). http://arxiv.org/abs/1612.01022
28. Flex Kunde, F., Sauer, P.: Traffic prediction using a deep learning paradigm (2017). http://ceur-ws.org/Vol-1810/BIGQP_paper_03
29. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction (2016). http://arxiv.org/abs/1610.00081
30. Yi, H., Jung, H., Bae, S.: Deep Neural Networks for traffic flow prediction. In: IEEE International Conference on Big Data and Smart Computing, pp. 328–331 (2017)
31. Ma, X., Dai, Z.: Learning traffic as images: a deep convolution neural network for large-scale transportation network speed prediction (2017). http://arxiv.org/abs/1701.04245
32. Surya, S., Babu, R.V.: TraCount: a deep CNN for highly overlapping vehicle counting. In: 10th Indian Conference on Computer Vision, Graphics and Image Processing, p. 46. ACM (2016)

33. Polson, N.G., Sokolov, V.O.: Deep learning for short-term traffic flow prediction (2017). https://arxiv.org/pdf/1604.04527.pdf
34. Beijing Traffic Management Bureau. http://www.bjjtgl.gov.cn/english/Roads/index.html
35. National Weather service. http://www.weather.gov
36. City Pulse Dataset Collection. http://iot.ee.surrey.ac.uk:8080/
37. AMAP: A Web Based Service Provider in China. https://ditu.amap.com/,accessed:2016-09-01
38. https://tu-dresden.de/bu/verkehr/vis/vlp/forschung/forschungsprojekte
39. Shafer, G., et al.: A Mathematical Theory of Evidence, vol. 1. Princeton University Press, Princeton (1976)
40. Moré, J.J.: The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G. A. (ed.) Numerical Analysis. LNM, vol. 630. Springer, Heidelberg (1978)
41. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
42. Alpaydin, E.: Introduction to Machine Learning, 3rd edn. MIT Press, Cambridge (2014)
43. Intelligent Transportation System. https://en.wikipedia.org/wiki/Intelligent_transportation_system
44. Intelligent Transportation systems, Benefits, Costs, Deployment and Lessons Learned. https://ntl.bts.gov/lib/30000/30400/30466/14412.pdf
45. Bellman, R.E.: Dynamic Programming. Princeton University Press, Rand Corporation, Princeton (1957)
46. Verleysen, M., François, D.: The curse of dimensionality in data mining and time series prediction. In: Cabestany, J., Prieto, A., Sandoval, F. (eds.) IWANN 2005. LNCS, vol. 3512, pp. 758–770. Springer, Heidelberg (2005). https://doi.org/10.1007/11494669_93
47. Hinton, G.E., Osindero, S.: A fast learning algorithm for deep belief nets. Neural Comput. **18** (7), 1527–1554 (2006)
48. Hyndmana, R., Koehler, A.: Another look at measures of forecast accuracy. Int. J. Forecast. **22**(4), 679–688 (2006)
49. Chai, T., Draxler, R.: (RMSE) or (MAE)? –Arguments against avoiding RMSE in the literature
50. Tofallis, C.J.: A better measure of relative prediction accuracy for model selection and model estimation. Oper. Res. Soc. **66**(8), 1352–1362 (2015)