Traffic Flow Prediction via Convolutional Deep Learning

Stefano Bilotta, Enrico Collini, Paolo Nesi, Gianni Pantaleo
University of Florence, Distributed Systems and Internet Technology, DISIT Lab
https://www.disit.org , Https://www.snap4city.org, corresponding: paolo.nesi@unifi.it

Traffic flow measuring is central for intelligent transportation systems (ITS). According to recent technologies, real-time traffic flow data can be measured, collected and exploited. The knowledge of real-time traffic flow data enables the development of a large number of services such as congestion detection and reduction; computing of origin-destination matrices; incident management; optimization of existing infrastructures of public transport; dynamic network traffic control; improved information services (e.g., traffic information, dynamic route guidance, road digital signage, planned routing); plan for future investments on mobility solutions; reducing fuel consumption and emissions of both CO2 and NO2 that strongly depend on fossil combustion (and thus on traffic, as well); predicting NOX. This paper presents a solution to compute short-term traffic flow sensors predictions up to 1 hour in advance, with a resolution of 10 minutes. The proposed results are innovative since the solution proposed:

- overcomes the state-of-the-art solutions in terms of precision and it is based on a never used architecture for the purpose:
- clarifies which are the features actually relevant (historical, seasonality, weather, pollutant, etc.) in prediction computation, thus providing errors in all possible feature combinations for a large range of different machines and deep learning techniques vs the proposed solution, aiming at covering all cases reported in the literature for total of 512 combinations,
- has been validated in a complex urban network of a real-world road structure, which is an aspect totally different from most solutions only tested on high-speed roads which have less noisy and quite regular traffic flow conditions.

The data related to traffic measurements have been divided into categories. The *Traffic* category includes the TrafficFlow metric at the observation time that refers to the number of vehicles detected by sensors, while *TrafPlus includes* other measures coming from traffic sensors, such as: the vehicles' AverageSpeed (km/h) and the Concentration which is a punctual measure expressed in percentage. The *DateTime* category includes the timeOfTheDay metric, encoded with a number that ranges from 1 to 144, since traffic flow data are measured and collected every 10 minutes. Typically, these values are used to also consider the data seasonality that may have different trends, e.g., working days with respect to weekends. Usually, the trend related to the vehicle number is similar on the same day of the week (e.g., Monday of current week with respect to Mondays in past weeks). Features related to the *Seasonality* are dayOfTheYear, dayOfTheWeek, Weekend, and Year.

EXPERIMENTAL RESULTS

According to data, the identified challenge was not only to find the best architecture to predict the traffic flow with a resolution of 10 minutes for the next hour, but also to discover the most informative set of features for the analysed models.

TABLE I **Overview of the features used in the short-term prediction models.**

Category	Feature	Description
Traffic Trafplus	*Ytaffic Flow*	Real number of vehicles recorded every 10 minutes
	AverageSpeed	Average speed of vehicles (Km/h)
	Concentration	Number of vehicles in terms of road occupancy (%)
DateTime	*timeOfTheDay*	Time of the day {1, 144}
	dayOfTheYear	Day of the year {1, 366}
seasonality	*dayOfTheWeek*	Day of the week {1,7}
	Weekend	0 for working days, 1 else
	Year	The year of the observation
Temporal	Previous observation's difference of the previous week (*dP*)	the difference between the number of vehicles in the observation day (d) at the time slot t and the number of available vehicles during the previous time slot (t-1) of the previous day (d-1)
	Subsequent observation's difference of the previous week (*dS*)	the difference between the number of vehicles in the observation day (d) at the time slot t and the number of vehicles during the successive time slot (t+1) of the previous day (d-1).
	Previous week observation (*PwVF*)	the number of vehicles of the previous week (d-7) in the same time slot (t).
Weather	*Air Temperature*	City temperature one hour earlier than *Time* (°C)
	Humidity	City humidity one hour earlier than *Time (%)*
	Pressure	City pressure one hour earlier than *Time (millibar mb)*
	Wind Speed	City wind speed one hour earlier than *Time (KM/h)*
AirPoll	*CO*	Concentration of CO one hour earlier than *Time*
	NO2	Concentration of NO2 one hour earlier than *Time*
	O3	Concentration of O3 one hour earlier than *Time*
	PM10	Concentration of PM10 one hour earlier than *Time*
	PM2.5	Concentration of PM2.5 one hour earlier than *Time*

In order to better understand the influence of each feature category, we have collected a large set of features, as described in **Table** I: traffic, datetime, seasonality, temporal, weather and air pollutants. Assuming the *Traffic* category mandatory as input for the construction of any predictive model, the number of combinations of the other 6 feature categories reaches 64. Thus, we have trained, tested, and validated all 64 combinations vs the traditional ML and deep learning and CONV-BI-LSTM) (see **Table** II), which included also the ones used in literature. The aim was to identify the best model, and at the same time to understand which are the most relevant features.

TABLE II -- THE **MAPE** ESTIMATED FOR **64** COMBINATIONS OF FEATURES FOR ALL THE IDENTIFIED TECHNIQUES AS THE MEDIAN VALUE ON THE SENSORS IN THE **3** CLUSTERS DESCRIBED ABOVE. THE ORDER IS BASED ON THE COMBINATION OF FEATURES. IN BOLD, BEST RESULTS/CONFIGURATIONS. IN BOLD WITH CITATION: RESULTS OBTAINED TAKING INTO ACCOUNT SOLUTIONS FROM THE STATE OF THE ART. PLEASE NOTE THAT **CONV-BI-LSTM** OVERCOMES ALL OF THEM IN THE SAME FEATURE CONDITIONS.

ID	Features adopted in the model						Median value of MAPE for prediction results by technique								min
	Date time	Traf plus	Temp oral	Season ality	Airpoll	weath er	RF	XGBO OST	DNN	LSTM	BI-LSTM	Autoencod er BI-LSTM	Attention CONV-LSTM	CONV-BI-LSTM	
C1	Y	Y	Y	Y	Y	Y	29.342	34.552	42.754	49.407	34.865	34,708	37,059	31.365	29.342
C2	Y	Y	Y	Y	Y	N	29.682	35.545	43.400	49.832	35.870	35,707	39,506	35.613	29.682
C3	Y	Y	Y	Y	N	Y	28.782	34.441	35.465	36.824	31.555	32,998	33,179	30.894	28.782
C4	Y	Y	Y	Y	N	N	30.935	35.373	38.942	35.383	30.564	32,969	35,713	32.485	30.564
C5	Y	Y	Y	N	Y	Y	29.776	34.469	33.425	42.301	39.865	37,167	35,161	36.897	29.776
C6	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**29.598**	**35.547**	**33.865**	**36.792**	**35.097**	35,322	29,923	**25.981**	**25.981**
C7	Y	Y	Y	N	N	Y	29.421	33.711	31.377	34.736	40.510	37,110	30,741	30.106	29.421
C8	Y	Y	Y	N	N	N	31.245	34.414	32.026	37.823	40.662	37,538	31,263	30.500	30.500
C9	Y	Y	N	Y	Y	Y	29.626	36.919	42.187	**37.068 [3]**	34.297	35,608	36,651	**31.115**	29.626
C10	Y	Y	N	Y	Y	N	29.964	35.802	47.201	41.334	34.743	35,272	40,658	34.116	29.964
C14	Y	Y	N	N	Y	N	29.764	36.374	36.203	43.510	35.744	36,059	33,015	29.827	29.764
C15	Y	Y	N	N	N	Y	29.972	35.423	31.526	46.201	37.209	36,316	32,919	34.313	29.972
C16	Y	Y	N	N	N	N	**30.960 [1]**	34.235	30.338	**37.068 [2]**	**38.082 [4]**	**34,235[5]**	**29,455[6]**	**28.573**	28.573
C17	Y	N	Y	Y	Y	Y	29.281	34.503	72.909	64.557	48.685	41,594	51,026	29.144	29.144
C18	Y	N	Y	Y	Y	N	30.184	35.350	59.458	68.127	46.874	41,112	44,810	30.163	30.163
C27	Y	N	N	Y	N	Y	28.986	35.218	57.938	50.333	59.419	47,318	43,298	28.658	28.658
C28	**Y**	**N**	**N**	**Y**	**N**	**N**	**31.068**	**35.878**	**66.634**	**50.957**	**55.096**	45,487	47,097	**27.561**	**27.561**
C29	Y	N	N	N	Y	Y	29.301	37.532	38.325	40.677	50.303	43,917	35,554	32.784	29.301
C30	Y	N	N	N	Y	N	29.323	37.284	37.149	48.801	55.064	46,174	34,721	32.294	29.323
C31	Y	N	N	N	N	Y	29.964	36.331	34.638	56.157	45.016	40,673	35,293	35.949	29.964
C32	**Y**	**N**	**N**	**N**	**N**	**N**	**29.281**	**34.574**	**33.028**	**57.961**	**44.977**	39,775	29,320	**25.612**	**25.612**

Not all combinations have been reported in the table for the lack of space. The best results have been achieved by the predictive models presenting a convolutional layer which efficiently extracts local features in noisy data (**Table** VI rows C32 – C6 – C28).

While a bidirectional approach improved, in most cases, the classic LSTM by performing an additional training on reversed order input dataset, which seems to lead to a better understanding of the underlying context also in time series data [7]. CONV-BI-LSTM approach worked in a quite satisfactory manner without considering Weather and Airpoll features, whereas RF generically benefits from such presence. Moreover, other features positively could contribute to the precision in terms of MAPE, but the impact of categories Trafplus, Temporal, Seasonality is not as evident on **Table** II as for the Weather and Datetime.

Data missing is an inevitable problem when dealing with real-world IoT sensor networks and of course, the traffic data from the real traffic system scenario of this study are affected by this problem. Traffic sensors may suffer of problems such as detector malfunction and communication failure, while there could be also some problems during the data acquisition process. All these problems can affect the monitoring of traffic and may constrain the predictive capability of the predictive models at runtime. The approaches of data imputation for producing surrogate data may help in creating dense data in training and execution [63], while actual data are preferrable. Therefore, in training, we overcome the occurrence of missing data cases by considering only complete samples/sequences according to the architecture.

This paper proposes a solution and an approach for short-term traffic flow prediction by using traditional machine learning as RF and XGBOOST and comparing them with deep learning techniques as DNN, LSTM, BI-LSTM, Attention based CONV-LSTM, Autoencoder BI-LSTM, and the proposed CONV-BI-LSTM. In the paper a comparative analysis has been performed, taking into account a large number of solutions and features, thus analysing the precision of such different techniques. Best solution turned out to be the solution, namely CONV-BI-LSTM which in most cases produced better results with respect to other solutions already in the state of the art and even better results could be obtained with the proposed feature combination.

REFERENCES

[14] [1] Zheng, C. H. E. N. G., & Xian-fu, C. H. E. N. (2016). The model of short-term traffic flow prediction based on the random forest. Microcomputers and Its Applications, 35(10), 46-49.

[23] [2] Y. Tian and L. Pan, "Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network," 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), 2015, pp. 153-158, doi: 10.1109/SmartCity.2015.63.

[38] [3] Awan, Faraz & Minerva, Roberto & Crespi, Noel. (2020). Improving Road Traffic Forecasting Using Air Pollution and Atmospheric Data: Experiments Based on LSTM Recurrent Neural Networks. Sensors. 20. 3749. 10.3390/s20133749.

[39] [4] R.L. Abduljabbar, H. Dia and P.W. Tsai (2021). Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction. *Journal of Advanced Transportation, 2021, 5589075.*

[45] [5] A. Essien, I. Petrounias, P. Sampaio, and S. Sampaio, “A deep-learning model for urban traffic flow prediction with traffic events mined from twitter,” World Wide Web, vol. 24, no. 4, pp. 1345–1368, Jul. 2021, doi: 10.1007/s11280-020-00800-3.

[46] [6] S. Verma, “An Attention Based Deep Learning Model For Traffic Flow Prediction Using Spatio Temporal Features Towards Sustainable Smart City,” International Journal of Communication Systems, vol. 34, Feb. 2021, doi: 10.1002/dac.4609.

[61] [7] S. Siami-Namini, N. Tavakoli and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 3285-3292, doi: 10.1109/BigData47090.2019.900599