Traffic state prediction with machine learning algorithms for short-term and mid-term prediction time horizons

Arash Rasaizadi[1], Seyedehsan Seyedabrishami [2*]

[1]Ph.D. candidate in transportation planning engineering, Tarbiat Modares University, Tehran, Iran
[2]Assistant Professor of transportation planning engineering at the School of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran (seyedabrishami@modares.ac.ir)

ABSTRACT

Predicting traffic variables and informing the passengers and the transportation network operators is one way to manage the travel demand. By informing the future state of traffic through intelligent transportation systems, there is more readiness to avoid congestion. In this study, three machine learning algorithms, including support vector machine (SVM), random forest (RF), and long short-term memory (LSTM), were used to predict the hourly traffic state, consist of light, semi-heavy and heavy states, for Karaj to Chaloos rural road in the north of Iran. Predictor variables of mid-term models are calendar information, weather, and road blockage policies. In contrast, in short-term models, in addition to the mentioned variables, the observed traffic states in the past three to eight hours have been used, and these models can only predict the future of one and two hours. The results show that short-term LSTM is the most accurate traffic state predictive model with accuracy equal to 90.11%. Among the mid-term models, the LSTM model has predicted traffic state more accurately than SVM and RF, and its accuracy is equal to 82.07%. Also, LSTM has the highest values of f1 measure to predict light, semi-heavy, and heavy, which are equivalent to 0.86, 0.93, and 0.81, respectively. Also, the hour, holiday, and type of holiday variables and traffic state observed in 3 to 8 hours later variables have the greatest effect on increasing the accuracy of mid-term and short-term models, respectively.

KEYWORDS

Traffic state prediction, Support vector machine, Random forest, Long short-term memory, Intelligent transportation systems

* Corresponding Author: Email: seyedabrishami@modares.ac.ir

1. Introduction

Intelligent transportation systems (ITS) are one of the main components of traffic management systems. The result of the use of these systems is to achieve or maintain the balance between travel supply and demand with high efficiency and low cost. The advanced traveler information systems (ATIS) is one of the most important of subsets of ITS. In these systems, all the practical and available data of the transportation network will be given to the passenger to schedule their travels. This information can be used for the existing state of the transportation network but the effectiveness of this information can be more if the future of the transportation network is predicted and informed [1].

In this study, hourly traffic state including, light, semi-heavy and heavy is predicted using three methods based on machine learning algorithms including support vector machine (SVM), random forest (RF), and long short-term memory (LSTM). To predict traffic state, spatial-temporal influential predictor variables are extracted and added to the data set. Models have been divided into two groups: medium-term and short-term models. In the short-term model, in addition to the spatial-temporal variables, the traffic state observed in the past three to eight hours has also been used and these models only can predict traffic hours for one and two next hours. In the medium-term models, the previous observations of the traffic situation have not been used. Finally, the accuracy of these two sets of models is evaluated and compared. The traffic data of Karaj to Chaloos, a rural road in the north of Iran, is used to predict traffic state.

2. Data and methodology

This used traffic data is collected for one section of the Karaj-Chaloos rural road in the north of Iran by loop detectors. Data collection has been carried out from 21 March 2017 to 1۱ September 2019. Data is divided into three sections; observations of the first 22 months are used to train models, the next two months for validation, and the last six months for test the predictions. Table 1 describes the candidate features to predict traffic state in the dataset.

Table 1: Description of candidate features

Feature Name	Description
Season	Including spring, summer, fall, and winter
Solar month	Including 12 solar months
Lunar month	Including 12 lunar months
Day of a solar month	Including 29-31 days of a solar month
Day of a lunar month	Including 29-30 days of a lunar month
Time of day	Including 24 hours a day
6 hour before the holidays	Equal to 1 if it is 1 to 6 hour before holidays
6 hour after the holidays	Equal to 1 if it is 1 to 6 hour after holidays
Day or night	Including day and night
Number of holidays	The number of sequential holidays
Holidays	Includes 1 for holidays and 0 for other days
Holiday type	Type of holidays
Holiday in three days later	Equal to 1 if three days later is a holiday
Type of holidays in three days later	Including the holiday type of three days later if it is a holiday, otherwise equals 0.
Holiday in three days ago	Equal to 1 if three days ago is a holiday
Type of holidays in three days ago	Including the holiday type of three days ago if it is a holiday, otherwise equals 0.
Holiday in two days later	Equal to 1 if two days later is a holiday
Type of holidays in two days later	Including the holiday type of two days later if it is a holiday, otherwise equals 0.
Holiday in two days ago	Equal to 1 if two days ago is a holiday
Type of holidays in two days ago	Including the holiday type of two days ago if it is a holiday, otherwise equals 0.
Holiday in a day later	Equal to 1 if a day later is a holiday
Type of holidays in a day later	Including the holiday type of a day later if it is a holiday, otherwise equals 0.
Holiday in a day ago	Equal to 1 if a day ago is a holiday
Type of holidays in a day ago	Including the holiday type of a day ago if it is a holiday, otherwise equals 0.
Weather condition	Including sunny, rainy, and snowy

Used machine learning algorithms to predict traffic state are support vector machine (SVM), random forest (RF), and long short-term memory (LSTM).

In SVM, the support vectors are a set of points that characterize the boundaries of classes in an n-dimensional space. The SVM determines class boundaries which leads to the best classification and separation of data by embedding support vectors [2]. The RF consists of a large number of decision trees. In this model, the training data is divided between decision tree models. Predictions are made for each decision tree. The average of predictions is determined as the RF's final prediction [3]. The LSTM networks are a special type of recurrent neural network (RNN) capable of learning long-term dependencies. Traditional RNNs are not able to train the time series with long time lags. LSTM can address this issue by incorporating memory units and learning when to forget previous memories and update memories [4].

3. Results and discussion

Table 2 shows the accuracy of models for train and test datasets.

Table 2: Accuracy of models for train and test datasets

Model	Accuracy of short-term models (%)		Accuracy of medium-term models (%)	
	Train	Test	Train	Test
SVM	90.82	87.74	83.82	76.63
FR	92.45	87.51	84.13	79.04
LSTM	96.88	90.11	88.69	82.07

For both short-term and medium-term models, the LSTM has the highest accuracy. Also, short-term models are more accurate compared to medium-term models.

Tables 3 and 4 show the performance of models to predict each traffic state in short-term and medium-term models, respectively.

Table 3: Precision, recall and F1 in short-term models

Model	State	Precision	Recall	F_1
	Light	0.81	0.76	0.78
SVM	Semi-heavy	0.88	0.92	0.9
	Heavy	0.74	0.67	0.7
	Light	0.84	0.83	0.83
RF	Semi-heavy	0.91	0.92	0.91
	Heavy	0.79	0.77	0.78
	Light	0.86	0.86	0.86
LSTM	Semi-heavy	0.93	0.94	0.93
	Heavy	0.84	0.79	0.81

Table 4: Precision, recall and F1 in medium-term models

Model	State	Precision	Recall	F_1
	Light	0.68	0.69	0.69
SVM	Semi-heavy	0.83	0.85	0.84
	Heavy	0.63	0.52	0.57
RF	Light	0.7	0.76	0.73
	Semi-heavy	0.86	0.84	0.85
	Heavy	0.66	0.62	0.64
	Light	0.75	0.8	0.77
LSTM	Semi-heavy	0.88	0.86	0.87
	Heavy	0.69	0.64	0.67

Tables 3 and 4 shows that in term of F_1, LSTM predicts all of the traffic states more accurate then SVM and RF. Other findings are as follow:

The use of the hour and holiday-related variables have the greatest effect on increasing the accuracy of the medium-term models. Using observations from three to eight hours ago has the greatest effect on increasing the accuracy of the short-term model. The effective factors on traffic patterns are constant and are not a function of the used method. There is not a variable that removing it causes an increase in the accuracy of the predictions.

4. Conclusion

Traffic state is a qualitative traffic parameter that shows the performance of the road and is more understandable for travelers. After informing predicted traffic state to travelers and transportation agencies through ATIS more sustainable transportation system can be expected. This paper aims to predict the traffic state by using three machine learning techniques, SVM, RF, and LSTM. Traffic data of one section of Karaj to Chaloos is used to train and test models. Results show that LSTM outperforms SVM and RF, in both short-term and medium-term prediction time horizons. In general, short-term models are more accurate compared to medium-term models but the prediction time horizon of short-term models is limited to one and two next hours.

5. References

[1] Polydoropoulou, A., and Ben-Akiva, M., 2018. “The effect of advanced traveller information systems (ATIS) on travellers’ behaviour”, pp. 317-352, Routledge.

[2] Yang, J., Deng, J., Li, S., and Hao, Y., 2017. “Improved traffic detection with support vector machine based on restricted Boltzmann machine”, *Soft Computing*, 21(11), pp. 3101-3112.

[3] Probst, P., Wright, M. N., and Boulesteix, A. L., 2019. “Hyperparameters and tuning strategies for random forest”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), 1301.

[4] Sherstinsky, A., 2020. “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network”, *Physica D: Nonlinear Phenomena*, 404, 132306.