AUT Journal of Mathematics and Computing

AUT J. Math. Com., 3(2) (2022) 129-135
DOI: 10.22060/AJMC.2022.21395.1089

Original Article

Deep learning model for express lane traffic forecasting

Farzad Karami[a], Shahram Bohluli[b], Chao Huang[c], Nassim Sohaee*[d]

[a] *Amazon Inc., Austin, Texas, USA*
[b] *Gradient Systematics LLC., Dallas, Texas, USA*
[c] *Modern Mobility Partners LLC. , Atlanta, Georgia, USA*
[d] *Department of Information Technology and Decision Science, University of North Texas, Denton, Texas, USA*

ABSTRACT: Traffic forecasting plays a crucial role in the effective operation of managed lanes, as traffic demand and revenue are relatively volatile given parallel competition from adjacent, toll-free general purpose lanes. This paper proposes a deep learning framework to forecast short-term traffic volumes and speeds on managed lanes. A network of convolutional neural networks (CNN) was used to detect spatial features. Volume and speed were converted into heatmaps feeding into the CNN layers and temporal relationships were detected by a recurrent neural network (RNN) layer. A dense layer was used for the final prediction. Six months of historical volume and speed data on the I-580 Express Lanes in California, United States were utilized in this case study. Computational results confirm the effectiveness of the proposed data-driven deep learning framework in forecasting short-term traffic volumes and speeds on managed lanes.

Review History:
Received:31 May 2022
Accepted:24 August 2022
Available Online:01 September 2022

Keywords:
Traffic forecast
Convolutional neural netwrok
Toll management

AMS Subject Classification (2010):
68T07; 90B20

(Dedicated to Professor S. Mehdi Tashakkori Hashemi)

1. Introduction

Over the past decade, transportation agencies throughout the United States have implemented a variety of innovations to meet the mobility needs of a growing population and economy. These innovations include variable tolling, dynamic tolling and managed lanes such as High Occupancy Vehicle (HOV) lanes and High Occupancy Toll (HOT) lanes. Traffic forecasting for managed lanes and toll roads has traditionally relied on conventional travel demand modeling techniques, such as four-step and activity-based transportation models [8, 19, 10, 11, 9, 24, 16, 20, 23, 28]. Predicting a traveler's choice between a toll-paying route and free alternatives is typically based on expected utility theories [17, 18, 1, 2, 3, 4, 5, 12, 15]. This traditional methodology has been applied widely for decades and has many advantages, including relatively high prediction accuracy and flexibility for scenario analysis. The travel demand modeling approach is usually applied for long-term planning purposes over 10, 20, or 30+ years. One objective of developing travel demand models (TDM) is to provide revenue forecasts to support operating agencies in developing an expenditure plan in accordance. However, significant effort is required to calibrate every step in the modeling process. For example, a route choice model predicting the choice of a tolled facility or lane versus toll-free alternatives typically requires discrete choice data from a stated preference survey [1, 4, 5]. In addition,

*Corresponding author.
E-mail addresses: farzad.karami@utdallas.edu, sbohluli@gmail.com, chao@modernmobilitypartners.com, nassim.sohaee@unt.edu

TDM results are highly influenced by demographic assumptions and the derived distribution of value of time savings (VOT). A complete, well-calibrated TDM typically requires several months of teamwork and costs hundreds of thousands in tax dollars. In the revolutionary field of artificial intelligence, machine learning (ML) and deep learning are rapidly emerging as an alternative to traditional TDMs for traffic forecasting. ML algorithms "learn" how to perform important tasks by generalizing from provided examples. ML is a data-driven technique; as more data become available, more ambitious problems can be tackled. As a result, ML is widely used in computer science, stock marker forecasting and traffic flow prediction. Most of the basic machine learning models become progressively better in task they perform, but they need human intervention to improve the performance. If an ML algorithm returns an inaccurate prediction, then an engineer needs to step in and make adjustments. A subset of ML models, called deep learning models have the capability of improving the performance of the model by learning in-depth, complex, and non-trivial behaviors in data. A deep learning model is designed to continually analyze data with a logic structure similar to how a human brain would draw conclusions. To achieve this, deep learning uses a layered structure of algorithms called an artificial neural network (ANN). The design of an ANN is inspired by the biological neural network of the human brain. This makes for machine intelligence that's far more capable than that of standard machine learning models. The present study proposes a method for predicting mid-term (i.e., several weeks or months in the future) traffic volumes on managed lanes via deep learning with historical traffic data. Using 6 months of historical volume and speed data on the I-580 Express Lanes in California, United States, this case study implemented a network of convolutional neural networks (CNN) to detect spatial features in the data (i.e., the relationship between speed and volume among the lanes and measurement locations). CNN layers are type of ANN designed to learn form inputs of image format. With 17×6 heatmaps for each speed and volume represented by the CNN layers, the data are fed into a recurrent neural network (RNN) layer to detect temporal relationships. Considering the time series nature of traffic forecasting, RNN helps to learn from the traffic structure of past time windows. The data are then passed into a dense layer for prediction. The loss function is Mean Squared Error and the chosen metric is Mean Absolute Error. The Keras library in Python is used and the script is run on a GPU system. Deep learning algorithms have been used recently in other fields such as electricity load forecasting and short-term (i.e., next few minutes to hours) traffic flow prediction. However, the present study represents one of the first attempts to implement this methodology to forecast traffic and speeds on managed lanes for next few months. The remainder of this paper presents a review of the relevant literature, the methodology of the machine learning algorithm and a description of the utilized data, followed by the model results, conclusions and recommendations for future work.

2. Literature Review

In recent years, researchers have begun developing and implementing ML methods for the purposes of traffic forecasting. Lv et al. [22] used a novel deep learning method to predict traffic flows based on deep architecture models with big traffic data. Their model considers spatial and temporal correlations inherently, and a stacked autoencoder model is used to learn generic traffic flow features. Their model was trained in a greedy layer-wise fashion, and experiments demonstrate that their method for traffic flow prediction has superior performance to tradition approaches. By identifying trends over a longer period and offering a more comprehensive summary, Boyer et al. [7] expanded on and advanced previous work that used ML models to characterize transportation research. Using Google's passive origin-demand (OD) data for the San Francisco Bay Area, Sana et al. [26] applied several ML models to predict hourly OD. Several ML techniques (support vector regression, K Nearest Neighbor, Random Forest and Neural Networks) were explored. They concluded that ML techniques might be suitable for use with passive data sources if prediction accuracy is the primary goal, assuming the algorithm is properly developed. The drawback of such a model is that outputs and sensitivities are hard to interpret given the lack of visible mathematical formulas. Aiming at improving the identification and prediction of congestion formation, Elfar et al. [14] tested three ML techniques for short-term traffic congestion prediction—logistic regression, random forests and neural networks—using vehicle trajectories available through connected vehicles technology. They applied the Next Generation Simulation (NGSIM) program, which is backed by the FHWA, to replicate the provided vehicle trajectories. Two types of predictive models were developed in their study: (1) offline models, which are calibrated based on historical data and are then updated (i.e., re-trained), and (2) online models, which are calibrated using historical data and updated regularly using real-time information. Results indicated that the accuracy of the models built in their study to predict the congested traffic state can be as high as 97Using travel time data collected from a real-world road network, Luan et al. [21] investigated the transferability of different ML methods in short-term traffic prediction. The experiments show that it is possible to transfer ML models trained on a link to other links under certain conditions based on the similarity of observable factors in the training and target links. In 2018, Poddar et al. [25] utilized five state-of-the-art naive algorithms (Naïve Bayes, k-NN, decision tree, random forest and support vector machine) to identify congestion by extracting images from CCTV cameras installed at different locations in the state of Iowa. Applying these algorithms to the test dataset, the support vector machine achieved the highest f1-score. The researchers also noted that the presence of glare or other inhibitions in the camera during the daytime can lead to misclassifications of the traffic state. However, the results were within 5Dharia and Adeli [13] investigated a neural network model for forecasting freeway link travel times using the counter propagation neural (CPN) network. They found that the proposed freeway link travel forecasting model is particularly suitable for real-time advanced travel information and management systems.

3. Methods

The task of time series forecasting usually involves autoregression (AR) methods. An AR model can predict future behavior based on past behavior when there is a correlation between 1) values in a time series and 2) the values that precede and succeed them. For more than four decades, Box and Jenkins' ARIMA model [6] has been widely used as a standard AR model to forecast time series data. Despite the popularity of this model, there are some major drawbacks such as being limited to analyzing linear data. The ARIMA does not yield reliable forecasts when the underlying dataset exhibits non-linear behavior. In 1979, Hochreiter and Schmidhuber [27] introduced another technique for time series forecasting: Recurrent Neural Networks (RNN). Deep learning methods, including RNN, can identify the structure and pattern of data—such as non-linearity and complexity—in time series forecasting. Long Short-Term Memory (LSTM) is a type of RNN with the capability of remembering values from earlier stages for the purpose of future use. LSTM models can provide reliable forecasting if enough historical data are provided to determine the dataset's underlying structure. However, many real-world scenarios lack rich historical data, such as managed lanes that have recently become operational. Forecasting 18 months of traffic and revenue is one of the top on-demand requests by managed lane operators. For these purposes, the lack of an adequately-sized training dataset usually results in extreme over/underestimation in traffic forecasting. In order to get an insight into available data we can convert it into an image. The heatmap function can be used to create gradients of color: small values will be colored in light blue and large values in dark blue. Then data is represented as a scatterplot. In this data transformation we can change the size of the plot according to the requirement of the project. Large time windows will create more detailed images and small time windows will make coarse and pixelized images. Considering this data preprocessing step, there is a flexibility in the size of the training dataset. The present study attempts to address this issue by creating a neural network that, in addition to the RNN and LSTM layers, incorporates other effective and widely used deep learning layers such as convolutional layers. Convolutional neural networks (CNN) were initially invented for image recognition. The combination of RNN and CNN creates a system that can understand images and provide a temporal context for individual images. In this study, we first transformed historical toll, volume, and speed data into a set of visual images for use as an input dataset for a CNN model. We then added LSTM layers to forecast future using past events. Convolutional layers help to learn the complex structure of data, and the LSTM layer remembers these complex structures for future use. In this model we have created input images using end to end volume and speed data along each lane for a predefined number of timesteps (i.e., the span of time in the dataset).

4. Modeling Procedure

The model in this study is based on the spatial relationships between measurement points as well as the time dependencies between timesteps. To capture these dependencies for the benefit of forecasting, a deep learning model was developed. Each sample point consists of a heatmap with two dimensions: volume and speed. Regarding the network, a stack of four layers is used to accomplish the forecasting task. The first four layers consist of convolutional networks to capture the spatial dependencies at each heatmap (i.e., samples at each timestep). Then the extracted features were fed into an LSTM layer for extracting the sequential dependencies. Finally, the processed data were input to a dense layer for the final determination of the output values. The data were reformatted into pairs of timesteps and heatmaps. Each heatmap is a matrix of 17×6, in which 17 is the number of data gathering points distributed along the corridor and 6 represents the number of lanes. The data were prepared to be fed into the developed network in the following manner: Each sample was chosen to contain a predefined number of heatmaps, which is referred to as the "training steps number," and the output is defined as the heatmaps a certain number of steps ahead of the last training sample. The output can prosses several timesteps or a single timestep. Twenty percent of the sample was used as validation dataset; results are reported for both the validation and training datasets. The effect of night traffic (between 8 PM and 5 AM of the next day) is not considered in this model, since the traffic at nights has a negligible correlation with daytime traffic and the express lanes are open to traffic but toll-free during 8 PM to 5 AM.

Figures 2 illustrate the heatmaps for the eastbound and westbound directions of the I-580 corridor. Figure 3 illustrates the architecture of the deep learning model. N represents the number of timesteps included in each training sample and d represents the steps shifted forward for the prediction.

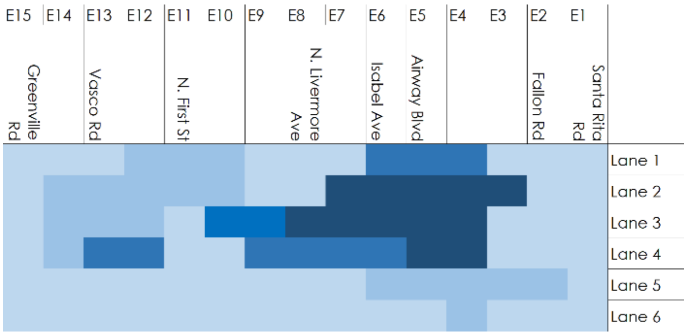

Figure 1: Example of heatmaps used for the spatial presentation of the data for eastbound.

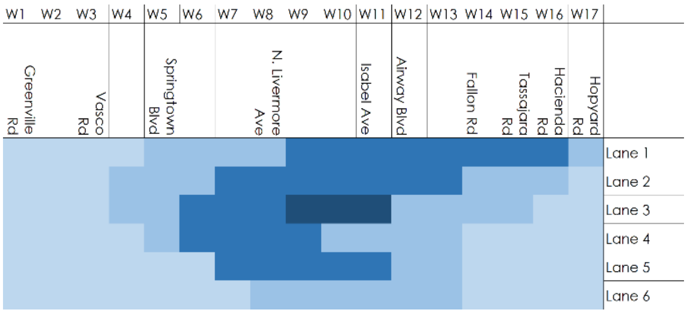

Figure 2: Example of heatmaps used for the spatial presentation of the data for westbound.

5. Data Description

In May 2010, the California Department of Transportation (i.e., Caltrans) developed the I-580 Corridor System Management Plan (CSMP). The I-580 CSMP focused on the "supply side" of the congestion problem and identified existing and future operational and capacity needs in the corridor. Various measures has been taken into consideration to reduce travel demand by commuters and commercial trucking in the referred corridor. Study findings and policy directives will be presented for consideration by local and regional agencies during future general plan and regional transportation plan updates respectively. One of the major action in this regard was considering one express lane in each direction.

The I-580 Express Lanes became operational in February 2016. The I-580 Express Lanes span approximately 12 miles from Greenville Road to just before the I-680 overpass in Dublin in the westbound direction, and approximately 10 miles from Hacienda Drive in Pleasanton to Greenville Road in Livermore in the eastbound direction (Figure 4).

I-580 has one express lane in the westbound direction; in the eastbound direction, I-580 has one express lane spanning the entire 12-mile corridor and two express lanes from the toll plaza near Fallon Road to N. First Street. The westbound express lane has been divided into eight toll zones, whereas the eastbound express lanes have been divided into seven toll zones. Each toll zone may include multiple toll plazas. In the westbound direction, a single

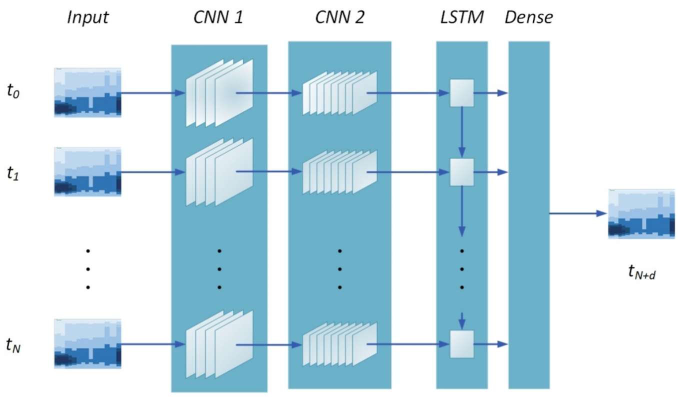

Figure 3: Architecture of the CNN Model

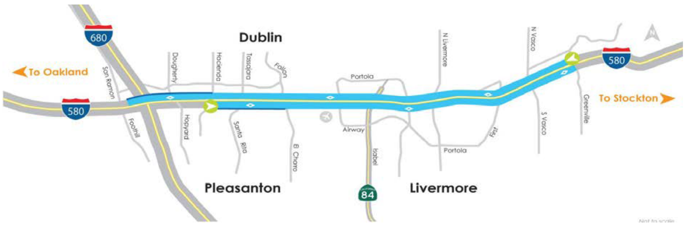

Figure 4: I-580 Express Lanes Location Map

express lane begins just west of Greenville Road, adjacent to four general-purpose (GP) lanes. The express lane is continuously accessible from the GP lanes until Hacienda Road, where it becomes a buffered lane until it ends just east of San Ramon Road/Foothill Road. The buffer is a double white stripe; no ingress/egress is permitted within the buffered section. In the eastbound direction, the express lane begins at Hacienda Drive as a single, buffered lane adjacent to four GP lanes. Just west of Fallon Road, a second express lane is added, and the lanes are opened for continuous ingress/egress, still adjacent to four GP lanes. The second express lane becomes a GP lane just west of Vasco Road; the remaining express lane becomes a GP lane just west of Greenville Road. The obtained volume and speed data are summarized below:

- The dataset contains bidirectional speed and volume data for each tolling zone from 5 AM to 8 PM over a span of 6 months from July 1 to December 31, 2018.
- The speeds are averaged, and the volumes are aggregated to 15-minute.
- The data include a 7880*17*6 tensor (number of timesteps, number of measurement locations, and number of lanes).
- The aim is to predict a timestep by its predecessor timesteps as input
- A timestep of 985 (15-minutes speed and volume data) was chosen, which is roughly equal to 10 days of measurement.
- A window of 985 timesteps was shifted forward with a single timestep at each time and the data for the half of the last timestep taken as the output.
- The final training matrices for volume and average speed are as follows: (6895, 985, 17, 6)
- Data for July to mid November 2018 is used for training and the remainder of 2018 for validation.

6. Results

The model was trained for eastbound and westbound directions separately. The values were normalized between 0 and 1 for all cases. The MAE is reported as the percentage of the maximum value. For a showcase, the prediction was conducted for 5 days ahead of the current time. The validation MAE across all lanes shows an accuracy of around 0.07. Considering the simple average of volume as baseline, MAE is 0.11. The results show a significant in the volume prediction. However, the performance of the model in speed prediction is same as the baseline model. The reason for that is the speed on the express lane is controlled to stay within a certain range and the variation around the average value is relatively small. Therefore, the average of the speed does not change considerably day by day and the baseline approach gives the best estimation. Table 1 shows the final loss function and the mean squared error for each case.

7. Conclusions and Recommendations

The present study is the first step toward developing a model to incorporate data from separate managed lanes at several measurement locations for the purpose of forecasting traffic on the I-580 Express Lanes in California as a case study. The results indicate promising performance from the proposed model for practical use. For the completion of this work, the following should be addressed:

- Analyzing the effect of the number of training sample timesteps is a matter of interest for reaching an optimal balance between forecasting accuracy and computational burden.

Table 1: Final Loss Function and MSE for Each Case

	Training Loss Function Value	Training MAE	Validation Loss Function Value	Validation MAE
Westbound Volume	0.0101	6.7 %	0.0131	8.5 %
Westbound Speed	0.0060	4.5 %	0.0313	10%
Eastbound Volume	0.0111	7.1%	0.0102	8.3 %
Eastbound Speed	0.0071	5.2 %	0.0323	11%

- Separate the forecasting of the GP lanes and the express lanes in order to develop a model for optimizing the pricing policies of express lanes.
- Incorporate other data sources in the model, such as weather forecasting models.
- Enhance the performance of the current architecture by constructing heatmaps with more dimensions (e.g., heatmaps with volume and speed as two channels).

References

[1] J. Bates, *Papers on stated preference methods in transport research*, Journal of Transport Economics and Policy, 22 (1988), pp. 7–9.

[2] M. E. Ben-Akiva and S. R. Lerman, *Discrete choice analysis: theory and application to travel demand*, vol. 9, MIT press, 1985.

[3] E. Ben-Elia and Y. Shiftan, *Which road do i take? a learning-based model of route-choice behavior with real-time information*, Transportation Research Part A: Policy and Practice, 44 (2010), pp. 249–264.

[4] C. R. Bhat and S. Castelar, *A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the san francisco bay area*, Transportation Research Part B: Methodological, 36 (2002), pp. 593–616.

[5] M. C. Bliemer, J. M. Rose, and S. Hess, *Approximation of bayesian efficiency in experimental choice designs*, Journal of Choice Modelling, 1 (2008), pp. 98–126.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 1970.

[7] R. C. Boyer, W. T. Scherer, and M. C. Smith, *Trends over two decades of transportation research: a machine learning approach*, Transportation research record, 2614 (2017), pp. 1–9.

[8] Cambridge Systematics, Inc., *Eastside corridor independent traffic and revenue study*, September 2012. Final Report.

[9] C&M Associates Inc., *Modeling for managed lane traffic and revenue forecasting*, 2015. Final Report.

[10] ——, *I-73 intermediate traffic and revenue study, south carolina*, 2016. Final Report.

[11] ——, *Ntta special projects system: Traffic and revenue bring down letter*, 2016. Final Report.

[12] A. de Palma and N. Picard, *Route choice decision under travel time uncertainty*, Transportation Research Part A: Policy and Practice, 39 (2005), pp. 295–324.

[13] A. Dharia and H. Adeli, *Neural network model for rapid forecasting of freeway link travel time*, Engineering Applications of Artificial Intelligence, 16 (2003), pp. 607–613.

[14] A. Elfar, A. Talebpour, and H. S. Mahmassani, *Machine learning approach to short-term traffic congestion prediction in a connected environment*, Transportation Research Record, 2672 (2018), pp. 185–195.

[15] D. A. Hensher, J. M. Rose, J. M. Rose, and W. H. Greene, *Applied choice analysis: a primer*, Cambridge university press, 2005.

[16] HNTB Corporation, *Atlanta regional managed lane system plan, technical memorandum 6*, 2010. Final Report.

[17] C. Huang, M. Burris, and W. D. Shaw, *Differences in probability weighting for individual travelers: a managed lane choice application*, Transportation, 44 (2017), pp. 375–393.

[18] ——, *Models of transportation choice with risk: an application to managed lanes*, Transportation letters, 9 (2017), pp. 276–295.

[19] C. A. Inc., *Traffic & revenue study for i-66 express lanes outside the capital beltway, virginia*, September 2015. Final Report.

[20] Larso Transportation Institute, *Public perspective on traffic and revenue forecasts for public-private partnerships*, 2017.

[21] J. Luan, F. Guo, J. Polak, N. Hoose, and R. Krishnan, *Investigating the transferability of machine learning methods in short-term travel time prediction*, in Transportation Research Board 97th Annual Meeting, Washington, DC, 2018.

[22] Y. Lv, X. Zhang, W. Kang, and Y. Duan, *Managing emergency traffic evacuation with a partially random destination allocation strategy: A computational-experiment-based optimization approach*, IEEE Transactions on Intelligent Transportation Systems, 16 (2015), pp. 2182–2191.

[23] MnDOT, *MnPASS system study phase 3*, 2017. Final Report.

[24] R. Phillips, C. Gocmen, and G. van Ryzin, *Revenue maximizing dynamic tolls for managed lanes: A simulation study*, 10 2015.

[25] S. Poddar, K. Ozcan, P. Chakraborty, V. Ahsani, A. Sharma, and S. Sarkar, *Comparison of machine learning algorithms to determine traffic congestion from camera images*, in Transportation Research Board 97th Annual Meeting, Washington, DC, 2018.

[26] B. Sana, J. Castiglione, D. Cooper, and D. Tischler, *Using google's passive data and machine learning for origin-destination demand estimation*, Transportation Research Record, 2672 (2018), pp. 73–82.

[27] J. Schmidhuber and S. Hochreiter, *Long short-term memory*, Neural Comput., 9 (1997), pp. 1735–1780.

[28] WilburSmith Associates, *SH 183 managed lanes intermediate level 2*. Final Report.

Please cite this article using:

Farzad Karami, Shahram Bohluli, Chao Huang, Nassim Sohaee, Deep learning model for express lane traffic forecasting, AUT J. Math. Com., 3(2) (2022) 129-135
DOI: 10.22060/AJMC.2022.21395.1089