Short-term traffic volume prediction using neural networks

Nassim Sohaee* Hiranya Garbha Kumar† Aswin Sreeram‡

Abstract

There are many modeling techniques that can predict the behavior of complex systems, such as traffic volumes in regional transportation systems, with high accuracy. However, predictive power suffers significantly when non-recurring events, such as adverse weather, occur in these systems. Therefore, introducing novel ways to identify and quantify disruptions can improve projection accuracy and performance. Proactive traffic management requires the ability to predict traffic conditions. A relatively new mathematical model, the neural network, offers an attractive approach to modeling undefined, complex, and nonlinear situations. This algorithm is trained by using both historical data and non-recurring phenomena such as weather. In this study, we test our algorithm on traffic data collected on four highways and high-resolution weather data within the Dallas area. The test indicates the model's high accuracy and efficiency in predicting short-term traffic.

1 Introduction

Predicting short-term traffic conditions is a vital component of advanced traffic management and information systems which aim to influence travel behavior, reduce traffic congestion, improve mobility, and enhance air quality. Previous research on short-term traffic flow prediction primarily focuses on the normal, or non-conditions, environment. However, traffic data is highly nonlinear and varies by time of day and other influential factors. Weather is one such factor that has a significant impact on traffic dynamics. Adverse weather has been shown to lower free-flow speeds, shift critical density, decrease flow capacity, and make roads more prone to congestion. FHWA has presented empirical studies and statistics about the impact of inclement weather on roadways [1].

1.1 Traffic flow prediction The drastic increase in the number of vehicles on the road –also continually increasing trip frequencies and lengths– has resulted in heavy traffic congestion in almost all of the major cities around the world. In the initial periods, reducing traffic congestion was attempted through infrastructural modifications. The main drawback of these approaches is that they are capital intensive. The recent trend is to apply intelligent strategies for increasing the transportation systems sustainability, otherwise known as Intelligent Transportation Systems (ITS). In an ITS, it is essential to predict traffic flow on a short-term basis, using existing information of present traffic conditions and historical traffic observations. Short-term traffic flow forecasting involves predicting traffic volumes in each next time interval, which is generally in the range of five minutes to one hour, decided by traffic authorities considering the corresponding requirements and situations.

Traffic flow is given by the point per unit time period random process, consisting of a set of isolated points collected over time [2]. While modeling such systems, statistical techniques are traditionally used to identify the stochasticity in the observed data [3, 4, 5]. However, in general, either non-parametric [5, 6] or parametric approaches [7, 8, 9, 10, 11, 12, 13] can be used for traffic flow prediction. Various nonparametric techniques used for this purpose are linear and nonlinear regression, historical average algorithms [8], smoothing techniques [7, 8, 9, 10, 11], and autoregressive linear processes [5, 9, 13]. Time series modeling techniques, such as the autoregressive integrated moving average (ARIMA), have been confirmed as one of the most accurate methods for predicting traffic flow [14]. These techniques attempt to identify patterns in historical data through the decomposition of the trends in long-term seasonal patterns, followed by extrapolating the obtained pattern into the next time interval. Since the traffic flow pattern exhibits strong seasonality due to peak and off-peak traffic conditions -which repeat at roughly the same time every day- seasonal ARIMA (SARIMA) models have proven very useful in predicting traffic flow behavior [14]. Numerous investigations have indicated that SARIMA models perform better that than those based on random walk, linear regression, and support vector regression (SVR), historical average, and simple ARIMA [15]. The investigations using SARIMA models to predict traffic flow reveal that this

*Department of Information Systems, University of Texas at Dallas

†Department of Computer Science, University of Texas at Dallas

‡Department of Information Systems, University of Texas at Dallas

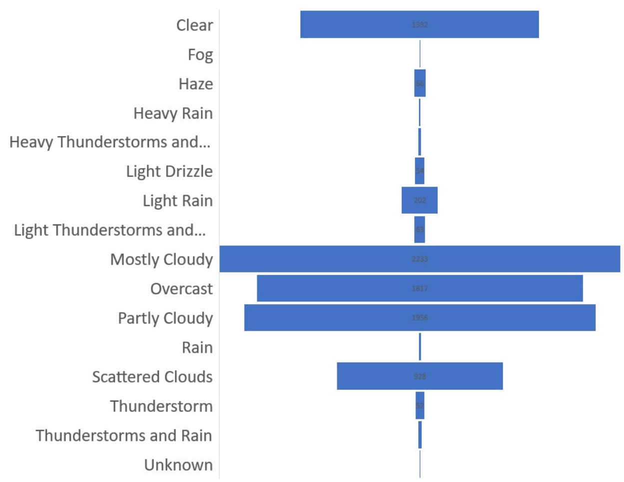

Figure 1: Weather condition in DFW, February to December 2016

technique, though accurate, has a drawback in that it needs a large amount of historical data for proper model development. For instance, Smith et al. [16] have used 15-minute traffic flow observations over a period of 45 days to forecast the next day's traffic flow. Williams and Hoel [17] used more than two months of traffic volume observations in their research. Stathopoulos and Karlaftis used 60, 000 traffic flow observations, which were aggregated for every three-minute interval over a period of 106 [18]. Using such large databases to build the models can be a significant issue in regions where data availability is lacking. Furthermore, the storage and maintenance of such historical databases is also a crucial problem. In addition to these shortcomings, the major issue in developing AARIMA models for short-term traffic flow prediction is computational overhead. Because of the computational intensity, such a process may not be feasible for real-time applications.

2 Methodology

We assume that weather conditions impact the free-flow speed, capacity, and critical density of a road. There are four categories of weather indexes used to express most weather conditions in the Dallas-Fort Worth (DFW) area. For this paper, we used high-resolution weather data, hourly weather data, and other weather indicators such as visibility. Figure 1 shows the distribution of weather conditions in 2016. Current weather data along with short term and long term historical traffic counts were used to predict traffic counts in the next hour.

3 Deep Learning

Deep learning is the application of artificial neural networks to learning tasks that contain more than one hidden layer. Deep learning is part of a broader family of machine learning methods based on learning data representations. Deep learning is a useful tool for identifying patterns. In general, deep learning methods produced favorable results in applications where the target function was complex, and the datasets were large.

A deep learning predictor, denoted by $\hat{y}(x)$, takes an input vector $x = (x_1, , x_n)$ and outputs y via different layers of abstraction, which employ hierarchical predictors by composing non-linear and semi-affine transformations. Let x^t_{t+h} be the forecast of traffic counts at time $t+h$, given measurements up to time t. Our deep learning traffic architecture looks as follows:

$$y(x) := x^t_{t+h}$$

To model traffic flow data $x^t = (x_{t-k}, \ldots, x_t)$ we use predictors x given by:

$$x^t = (x_{t-h}, x_{t-h+1}, \ldots, x_t)$$

Here x_i is the traffic count at time i. In a basic feed-forward neural network, raw input data are presented to processing elements in the input layer. The input values are then assigned weight and passed to the hidden layer. Elements in the hidden layer are then processed and passed to the output layer. The output layer processes the elements and produces the network output.

To develop a neural network model to perform traffic prediction, the network needs to be trained with historical examples of input-output data. As part of the model development process, decisions must be made about the architecture of the neural network. In neural networks, we usually train the network using stochastic or mini-batch gradient descents rather than the entire dataset. Stochastic and mini-batch gradient descent use a batch size number of training examples at each iteration, so at some point, you will have used all the data for training and can start over from the beginning of the dataset. One epoch is one complete pass through the entire training set, meaning multiple iterations of gradient descent updates until you show all the data to the neural network and then start again. For this paper, we conducted some experiments to determine the best neural network architecture.

We implemented a machine learning model for this dataset: Keras with Theano backend to build a sequential neural network for regression. Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or

Table 1: Four locations analyzed in the DFW area

Location ID	Physical location	AADT	Missing data
S126	IH 35 E	215,784	25.21%
S148	IH 35 E	201,267	6.03%
S220	IH 45	82,434	15.34%
S519	IH 30	19,535	3.29%

Theano. Keras was developed with a focus on enabling fast experimentation. The core data structure of Keras is a model, a way to organize layers. The simplest type of model is the Sequential model, which is a linear stack of layers.

3.1 Overfitting In statistics and machine learning, one of the most common tasks is to fit a model to a set of training data with the goal of making reliable predictions on unseen test data. In overfitting, a statistical model describes random error or noise instead of the underlying relationship. A model that has been overfitted has poor predictive performance, as it overreacts to minor fluctuations in the training data. Since both our models have more test accuracy than training accuracy, there doesn't seem to be any overfitting. The small number of epochs (10 iterations) also decreases the likelihood of overfitting. As we increase the number of epochs beyond a certain threshold, overfitting can become an issue.

3.2 Data Description To validate the efficiency of the proposed congestion prediction approach, the model was applied to data collected from four sites in the DFW area. The source of the data is Texas Department of Transportation's (TxDOT) Traffic Count Database System (TCDS). Table 1 shows basic information about the four locations analyzed in this paper.

The collected data are aggregated in 1-hour intervals. In this paper, the traffic count data were collected on weekdays and weekends for most of the weeks in 2016. The data was collected over the course of 11 months, from February to December of 2016. However, the data is not consistent within and across months; i.e., there are missing days in a month and also some missing months in the dataset. We accumulated the data into one file with 14 fields, which include location, date and time, volume count for the past 3 hours, volume count for the past 3 weeks on the same time and day, weather condition, rain, precipitation, and visibility. The total data points add up to about 30,000 points. We used 46% of the observations as training, and the remainder was used for testing the deep learning model.

Table 2: Accuracy of neural network prediction models

			Epoch size		
		Batch size	10	40	100
Training data	Model 1	10	83.35%	97.38%	96.79%
		25	76.92%	96.32%	97.46%
		50	74.65%	86.88%	97.31%
	Model 2	10	90.94%	94.33%	97.17%
		25	70.40%	94.11%	97.01%
		50	33.29%	92.96%	95.94%
	Model 3	10	90.94%	**97.54%**	97.17%
		25	70.40%	94.11%	97.01%
		50	33.29%	92.96%	95.94%
Testing data	Model 1	10	96.57%	97.49%	97.72%
		25	96.07%	97.30%	97.50%
		50	95.70%	96.89%	97.46%
	Model 2	10	96.84%	97.54%	**97.78%**
		25	96.03%	97.39%	97.66%
		50	94.61%	97.25%	97.57%
	Model 3	10	96.84%	94.33%	97.78%
		25	96.03%	97.39%	97.66%
		50	94.61%	97.25%	97.57%

4 Results

We developed a short-term traffic forecast for four locations in the DFW area. All four locations are located at three major interstate highways passing through Dallas, carrying a high volume of local and interstate traffic. To develop a predictive model, a traffic and weather database was developed. The data are stored in 1-hour intervals since January of 2010. We combined the data from all locations into one dataset, which is divided into a training set and the test set.

In general, the neural network model proved to outperform the traditional methods, demonstrating its ability to model complex characteristics. We tested our data set in three different neural network models. Model 1 has only one hidden layer of 8 nodes. Model 2 has two hidden layers of sizes 8 and 4, and Model 3 has three hidden layers of sizes 8, 4, and 4. We also changed the epoch size to see if we could improve accuracy. The outputs of these three models are presented in Table 2.

Our experiment shows all three models are not performing well on training set with small epoch size. However, all three models are producing significant accuracy in testing data. The best accuracy of training data has been achieved at batch size 10, epoch size 40 in the third model. And the second model at batch size 10 and epoch size 100 give the highest accuracy for training set. Note that low epoch size is computationally fast but sacrifices accuracy. Usually, as the batch size increases the computation time will increase too. The trend

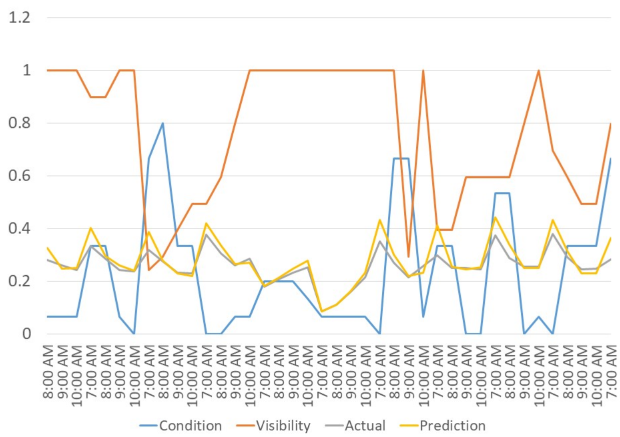

Figure 2: Morning peak

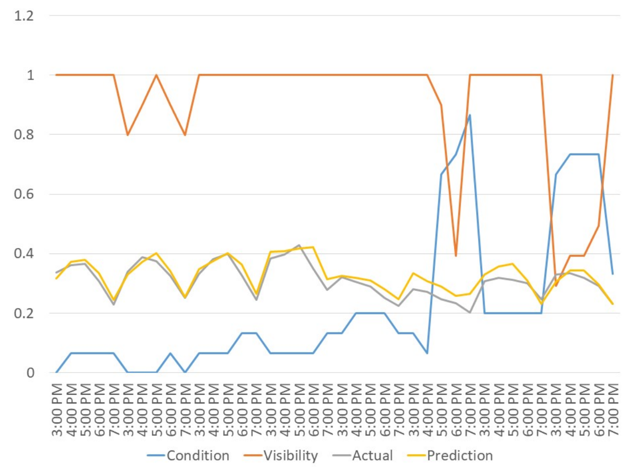

Figure 3: Afternoon peak

shows that increasing epoch size further might provide even better accuracy, but, as mentioned earlier, there is a concern of overfitting. Also, the computation time increases as epoch time increases, so it may not be a good candidate for real-time traffic prediction. If we want to settle for a computationally fast and reliable method, we can invest on Model 1 which has only one hidden layer, and small batch size 10 and epoch size 40. This model also has a consistent performance on both training and testing data set.

Figures 2 and 3 illustrate the performance of our deep learning model in peak hours in the morning and afternoon. We also included weather condition and visibility in these figures to show the effect of weather condition and visibility on traffic pattern. Deep learning model tends to overpredict morning peak volumes. This suggests that we need to include other factors that influence traffic volume. Afternoon peak volume forecast is closely fitted actual volume.

5 Conclusion and future work

Various kinds of traffic flow models are used to describe traffic flow characteristics; however, very few of them describe the explicit negative impact of adverse weather on travel speed, flow capacity, critical density, and many other aspects, such as driving safety. As research has demonstrated that weather conditions indeed impact the driving environment and driver behavior, it is necessary to build a weather-specific prediction model. We also provided a robust deep learning model to predict traffic volumes based on short-term and long-term historical data, as well as high-resolution local weather data. Future studies may also look into a combination of weather-specific data and other on-the-ground events, like maintenance and real-time accident data, and consider their implementation into real-time traffic control and prediction.

References

[1] H. Rakha, M. Farzaneh, M. Arafeh, R. Hranac, E. Sterzin, and D. Krechmer. *Empirical studies on traffic flow in inclement weather. s.l. : Final Report-Phase I* p 107, 2007.

[2] G. Y. Zhang, Q. Zhao, Z. W. Luo, H. Wei, *Short-term traffic flow prediction with ACD and particle filter* Proceedings of the 9th International Conference on Chinese Transportation Professionals, Harbin, China, 2009.

[3] S. I. Chien, L. Xiaobo, K. Ozbay. *Predicting travel times for the South Jersey real-time motorist information system* Transportation Research Record: Journal of the Transportation Research Board, pp. 1855:3240, 2003.

[4] S Vasantha Kumar, Lelitha Vanajakshi, *Short-term traffic flow prediction using seasonal ARIMA model with limited input data* European Transportation Research Review, pp. 7 - 21, 2005.

[5] B. Ghosh, B. Basu, M. O. Mahony, *Bayesian time-series model for short-term traffic flow forecasting,* Journal of Transportation Engineering, pp. 180-189, 2007.

[6] G. A. Davis, N. L. Nihan, *Nonparametric regression and short term freeway traffic forecasting,* Journal of Transportation Engineering, 117(2), pp. 178-188, 1991.

[7] P. C. Vythoulkas, *Alternative approaches to short-term traffic forecasting for use in driver information systems.* 12th International Symposium on traffic flow theory and transportation, Berkeley, California, 1993.

[8] B. L. Smith, M. J. Demetsky, *Short-term traffic flow prediction: neural network approach* Transportation

Research Record: Journal of the Transportation Research Board , 1453, pp. 98104, 1994.

[9] H. R. Kirby, S. M. Watson, M. S. Dougherty, *Should we use neural network or statistical models for short-term motorway traffic forecasting* International Journal of Forecast, 13(1), pp. 43-50, 1997.

[10] H. B. Yin, S. C. Wong, J. M. Xu, C. K. Wong, *Urban traffic flow prediction using a fuzzy-neural approach.* Transportation Research, Part-C, 10(2), pp. 85-98, 2002.

[11] D. H. Lee, W. Z. Zheng, Q. X. Shi, *Short-term freeway traffic flow prediction using a combined neural network model*, Proceedings of the 84th Annual Transportation Research Board Meeting, Washington D.C. 2004.

[12] E. I. Vlahogianni, M. G. Karlaftis, J. C. Golias, *Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach* Transportation Research, Part-C, 13(2), pp. 211-234, 2005.

[13] M. C. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, P. Zhang, *An aggregation approach to short-term traffic flow prediction*, IEEE Transactions of Intelligent Transportation Systems, 10(1), pp. 60-69, 2009.

[14] T. Mai, B. Ghosh, S. Wilson, *Multivariate short-term traffic flow forecasting using Bayesian vector autoregressive moving average model,* Proceedings of the 91st Transportation Research Board Annual Meeting, Washington, D.C. 2012.

[15] E. Chung, N. Rosalion, *Short term traffic flow prediction,* Proceedings of the 24th Australian Transportation Research Forum, Hobart, Tasmania 2001.

[16] B. L. Smith, B. M. Williams, R. K. Oswald, *Comparison of parametric and nonparametric models for traffic flow forecasting* Transportation Research: Part-C, 10, pp. 303321, 2002.

[17] B. M. Williams, L. A. Hoel, *Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results,* Journal of Transportation Engineering, 129(6), pp. 664672, 2003.

[18] A. Stathopoulos, M. G. Karlaftis, *A multivariate state space approach for urban traffic flow modeling and prediction* Transportation Research: Part-C, 11, pp. 121135, 2003.