A review of various short-term traffic speed forecasting models

John Boaz Lee
Information Sys. and Computer Science Dept.
Ateneo de Manila University, Philippines
jlee@ateneo.edu

Kardi Teknomo
Information Sys. and Computer Science Dept.
Ateneo de Manila University, Philippines
teknomo@gmail.com

ABSTRACT

The widespread adoption of smartphones in recent years has made it possible for us to collect large amounts of traffic data. Special software installed on the phones of drivers allow us to gather GPS trajectories of their vehicles on the road network. In this paper, we simulate the trajectories of multiple agents on a road network and use various models to forecast the short-term traffic speed of various links. Our results show that traditional techniques like multiple regression and artificial neural networks work well but simpler adaptive models that do not require prior training also perform comparatively well.

1. INTRODUCTION

In the past, intelligent transportation systems (ITS) relied on speed sensors installed in various locations to monitor traffic conditions. Because of the relatively large number of roads in urban areas and the high cost to set-up and maintain these traditional speed sensors, most ITS operators are only able to monitor the traffic conditions of urban areas partially. Now that the use of GPS-enabled devices has become quite pervasive, vehicles with these devices can be treated as a new kind of mobile sensor. In the city of Beijing alone, more than 50,000 licensed taxicabs (many of which are GPS-equipped) generate over 1 million vehicle trajectories per day [21]. This large amount of data, when processed and analysed, can be used in various applications.

Vehicle trajectories have been used to solve various problems in the domain of urban computing, which is an emerging field of study concerned with the application of technology in the urban setting [11]. In [21], for instance, researchers used the trajectories generated by around 30,000 taxicabs to identify defects in urban planning and regions with salient traffic problems. The same type of data was also used to identify unusual traffic patterns in another study [13]. This allowed researchers to automatically identify special events, like celebrations, protests, road closures, or large-scale business promotions, that disrupted the normal flow of traffic in an area.

Wei et al. [18] devised an algorithm that could analyse vehicle trajectories to identify popular routes. Yuan et al. [20], on the other hand, analysed the trajectories of taxi drivers to give driving directions. By leveraging the intelligence of taxi drivers, the developed system [20] was able to suggest routes that were often faster than routes suggested by other approaches.

Now that real-time traffic information has become more accessible due to the relative ease in collecting GPS trajectories of vehicles, the next logical step is to forecast future traffic condition [14]. This is a problem that has many practical applications. For instance, a system that suggests routes to drivers can take advantage of traffic forecasts to recommend routes that are predicted to be less congested in the near future.

The problem of forecasting short-term traffic condition is one that is well-studied in the literature; many different approaches have already been proposed. In some of the earlier studies, univariate time series models like the Box-Jenkins autoregressive integrated moving average (ARIMA) and various exponential smoothing models were used to forecast short-term traffic [1, 4, 15]. More recently, various approaches utilizing neural networks were introduced [2, 5, 10]. An approach based on dynamic wavelet neural networks was proposed in [10] while one that combined fuzzy logic, wavelet transforms, and neural networks was studied in a paper by Xiao et al. [19].

More advanced approaches include the model proposed by Min and Wynter [14] that utilizes a spatial-temporal autoregressive model to predict short-term traffic. Castro-Neto et al. [1], on the other hand, presents a supervised statistical learning technique called the Online Support Vector machine for Regression that is able to predict traffic flow under both typical and atypical conditions. Finally, [7] introduces a model that combines clustering and regression-based techniques to predict traffic speed.

In this paper, we survey various models for forecasting short-term traffic speed. The goal is not to devise a novel approach for solving the problem of traffic speed forecasting. Instead, we test several techniques to discover which works best with GPS trajectory data in practice.
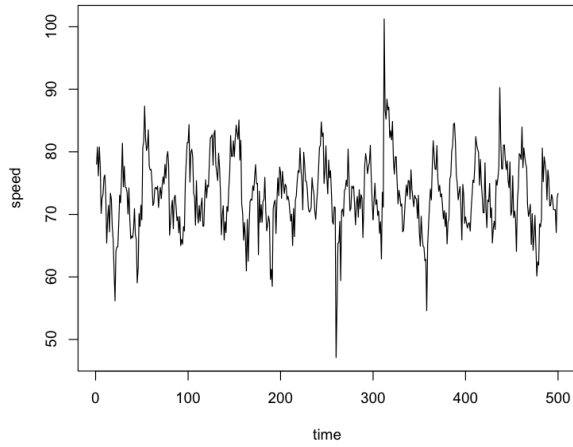
Figure 1: Simulated speed time series generated using a fitted SARIMA model.

The rest of the paper is structured as follows. In the next section we talk about the data and the models to be tested. We then explain the experiments and elaborate on the experimental results in the third section. Finally, we conclude the paper with some suggestions for future work.

2. METHODOLOGY

2.1 Dataset

Although the researchers have already implemented a system to collect GPS trajectories of vehicles in Metro Manila, the volume of data collected at the time of the study was insufficient for forecasting. Because of this, a synthetic vehicle trajectory dataset was constructed.

A graph to represent a road network was first created, time series speed data was then generated for each link or road in the network. Since real-world traffic speed follow cyclical and seasonal patterns, the generated data should also exhibit these features. To ensure this, we fitted a seasonal ARIMA model [9] on real-world traffic speed data taken from the North Luzon Expressway [3]. This model was then used to create the simulated speed data. An example of a simulated speed time series is shown in fig. 1.

We then simulated the trajectories of multiple agents on the graph. Even though the movements of the agents across the network were random, they were still subject to several constraints. First of all, the number of trajectories utilizing a certain link on the network is inversely proportional to the underlying traffic speed of the link at the current time point. Intuitively, the reason for this is because we can expect to capture the trajectories of more vehicles along a heavily congested road than a road that has light traffic and high speed. Secondly, the actual speed logged by a vehicle's trajectory as it traverses a link is drawn from a normal speed distribution with mean equals to the speed of the link at the current time point and standard deviation of 10. We can expect the speed of different trajectories traversing the same link to be slightly different just as different drivers drive their vehicles

	No seasonality	Additive seasonality	Multiplicative Seasonality
No Trend	$\hat{x}_{t+h} = a_t$ $a_t = \alpha x_t + (1-\alpha)a_{t-1}$	$\hat{x}_{t+h} = a_t + c_{t-s+h}$ $a_t = \alpha(x_t - c_{t-s}) + (1-\alpha)a_{t-1}$ $c_t = \gamma(x_t - a_t) + (1-\gamma)c_{t-s}$	$\hat{x}_{t+h} = a_t c_{t-s+h}$ $a_t = \alpha\dfrac{x_t}{c_{t-s}} + (1-\alpha)a_{t-1}$ $c_t = \gamma\dfrac{x_t}{a_t} + (1-\gamma)c_{t-s}$
Add. Trend	$\hat{x}_{t+h} = a_t + b_t h$ $a_t = \alpha x_t + (1-\alpha)(a_{t-1}+b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$	$\hat{x}_{t+h} = a_t + b_t h + c_{t-s+h}$ $a_t = \alpha(x_t - c_{t-s}) + (1-\alpha)(a_{t-1}+b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$ $c_t = \gamma(x_t - a_t) + (1-\gamma)c_{t-s}$	$\hat{x}_{t+h} = (a_t + b_t h)c_{t-s+h}$ $a_t = \alpha\dfrac{x_t}{c_{t-s}} + (1-\alpha)(a_{t-1}+b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$ $c_t = \gamma\dfrac{x_t}{a_t} + (1-\gamma)c_{t-s}$
Mult. Trend	$\hat{x}_{t+h} = a_t b_t^h$ $a_t = \alpha x_t + (1-\alpha)(a_{t-1}b_{t-1})$ $b_t = \beta\left(\dfrac{a_t}{a_{t-1}}\right) + (1-\beta)b_{t-1}$	$\hat{x}_{t+h} = a_t b_t^h + c_{t-s+h}$ $a_t = \alpha(x_t - c_{t-s}) + (1-\alpha)(a_{t-1}b_{t-1})$ $b_t = \beta\left(\dfrac{a_t}{a_{t-1}}\right) + (1-\beta)b_{t-1}$ $c_t = \gamma(x_t - a_t) + (1-\gamma)c_{t-s}$	$\hat{x}_{t+h} = a_t b_t^h c_{t-s+h}$ $a_t = \alpha\dfrac{x_t}{c_{t-s}} + (1-\alpha)(a_{t-1}b_{t-1})$ $b_t = \beta\left(\dfrac{a_t}{a_{t-1}}\right) + (1-\beta)b_{t-1}$ $c_t = \gamma\dfrac{x_t}{a_t} + (1-\gamma)c_{t-s}$

Table 1: Formula for the different exponential smoothing models.

Finally, to capture seasonality in the data, one can use Holt and Winter's formulation [9]

$$\mu_{t+h} = (a_t + b_t h)c_{t-s+h}$$
$$a_t = \alpha\frac{x_t}{c_{t-s}} + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$$
$$c_t = \gamma\frac{x_t}{a_t} + (1-\gamma)c_{t-s}$$

here s is the length of the season. In this model, the level is initialized by taking the average of the first season $a_s = \frac{1}{s}\sum_{i=1}^{s} x_i$, trend is initialized $b_s = \frac{1}{s}\sum_{i=1}^{s}\frac{x_{s+i}-x_i}{s}$, and the seasonal indices c_i are initialized $c_i = \frac{x_i}{a_s}$. Table 1 shows the formulation of the different exponential smoothing methods that we will be testing. Aside from the methods listed in table 1, we also tested the adaptive version of the traditional exponential smoothing model. One can refer to [9] for a thorough explanation of exponential smoothing.

3. EXPERIMENTAL SETUP, RESULTS, AND DISCUSSION

We tried to forecast short-term traffic speed using the methods discussed in the previous section. Feed-forward neural networks* were utilized and trained using the simulated data. Various versions with different decay rates and number of hidden units were tested to identify the best setup. To test the exponential smoothing models, we tested each model with different combinations of values for each parameter. The parameters, α, β, and γ were given values from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

The methods were first tested on a simulation where the average number of trajectories traversing a link was, at any given time, around 30. This average was then increased by 100 and then by 200.

---
*The implementation found in the caret package of R was utilized. Every time the model is retrained, it performs slightly differently.

|	ave = 30	ave = 130	ave = 230
NSNT	5.135	5.080	4.869
NSAT	5.153	5.099	4.887
NSMT	5.172	5.109	4.903
ASNT	**4.752**	**4.761**	**4.577**
ASAT	5.029	5.054	4.805
ASMT	5.962	5.910	5.807
MSNT	4.849	4.857	4.678
MSAT	5.115	5.133	4.900
MSMT	5.051	5.077	4.863
Neural Network	**4.725**	**4.829**	-
Multiple Regression	**4.557**	**4.530**	**4.375**

Table 2: The performance of the different models based on RMSE. The top three models are highlighted.

Table 2 shows the performance of the different methods in forecasting short-term traffic speed. Acronyms are used to describe the various exponential smoothing approaches; for instance, NSNT stands for no seasonality and no trend. We use root-mean-square error (RMSE) to evaluate performance, RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$$

where $\hat{y}_t$ and y_t are the predicted speed and the actual speed at time t, respectively. RMSE is a good measure that tells us how far the predicted values are from the actual observed values.

As expected, the supervised learning approaches performed better in the evaluation. However, it is interesting to see that the performance of the methods based on exponential smoothing did not differ too much from that of the super-
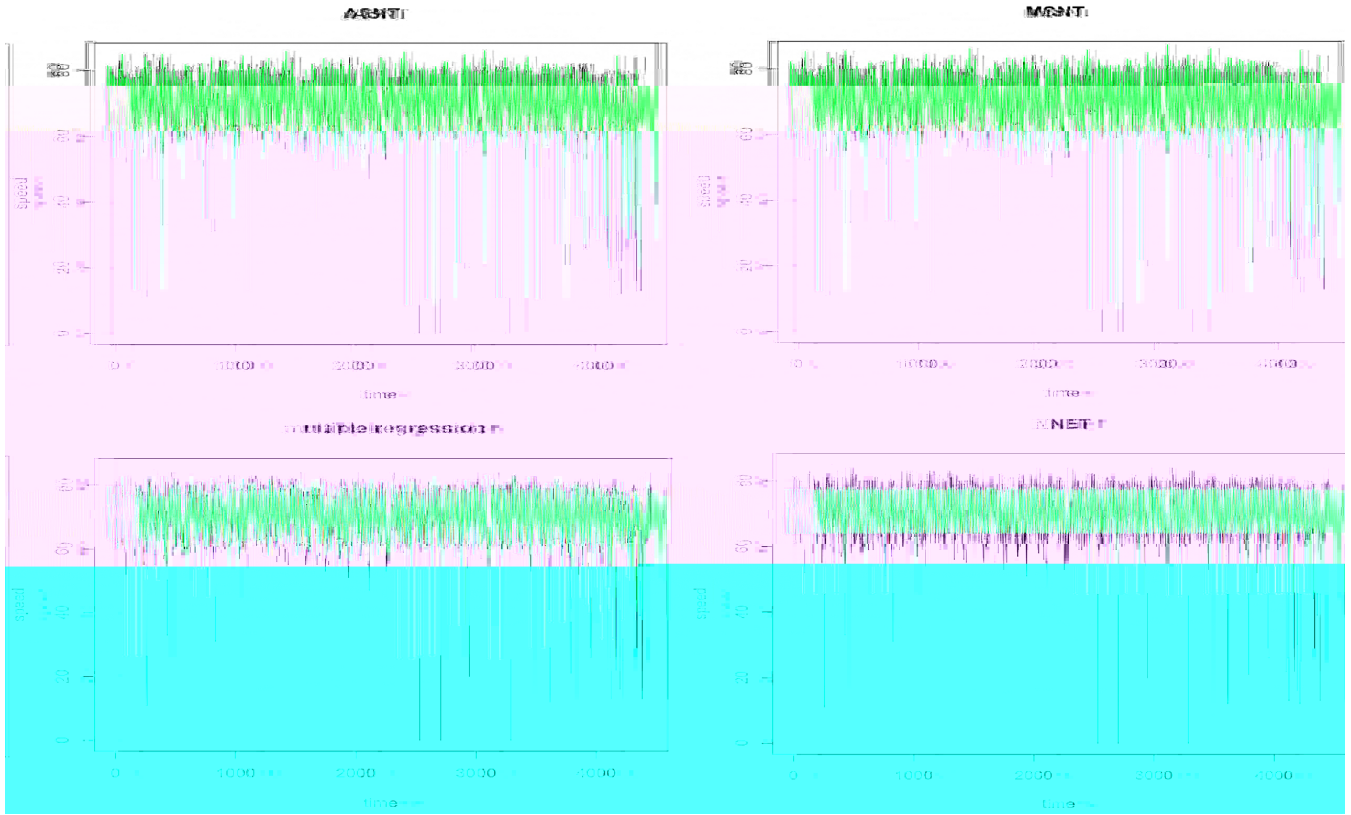
Figure 2: Single link traffic speed forecast using ASNT, MSNT, multiple regression, and neural network.

vised techniques. In particular, the models with seasonality and no trend (additive seasonality and no trend, multiplicative seasonality and no trend) performed comparatively well and are way better than some of the other variants. This may be due to fact that even though traffic data does exhibit long-term trend, at the micro-level the traffic speed fluctuates frequently and does not show prolonged trend in one direction. Thus adding trend in the formula could be counterproductive in this case. In fact, when we simply use exponential smoothing without seasonality, the formulation that works best is the one without trend as well.

Figure 2 shows the forecasted values overlapped with the actual traffic speed values for the four best performing methods. It is interesting to note that the multiple regression seems to be able to predict traffic speed fairly accurately during typical conditions but almost always underestimates the speed when there are sharp drops or spikes in speed (outlier conditions). Adding additional predictors that indicate special events may help the model, however the question is how we can do this automatically. The exponential smoothing-based methods, on the other hand, seem to be more prone to overestimation of speed as evidenced by their graphs. Because of this, however, they are able to predict the speed fairly accurately during atypical conditions. The neural network forecasts, in this instance, seem to be the most conservative and are mostly centered around the mean. This may be due to the fact that neural networks are known to require large amounts of training data to perform well.

	ideal parametric values
NSNT	$\alpha = 0.9$
NSAT	$\alpha = 0.8;\ \beta = 0.1$
NSMT	$\alpha = 0.8;\ \beta = 0.1$
ASNT	$\alpha = 0.5;\ \gamma = 0.1$
ASAT	$\alpha = 0.6;\ \beta = 0.1;\ \gamma = 0.1$
ASMT	$\alpha = 0.3;\ \beta = 0.1;\ \gamma = 0.2$
MSNT	$\alpha = 0.5;\ \gamma = 0.1$
MSAT	$\alpha = 0.6;\ \beta = 0.1;\ \gamma = 0.1$
MSMT	$\alpha = 0.7;\ \beta = 0.8;\ \gamma = 0.1$

Table 3: The optimal values for the parameters α, β, and γ.

Even though supervised techniques work well in forecasting, there is an overhead cost of retraining and discovering the ideal parameters of these models for every link in the road network. Given the potentially large amount of trajectory data that a real-world traffic forecasting system gathers continuously, it would be advantageous to use simpler models like ASNT or MSNT exponential smoothing models for forecasting at the cost of a slight reduction in accuracy.

3.1 Ideal parametric values
Table 3 shows the optimal values of the parameters used in the exponential smoothing models. For the models without seasonality, higher values for α are preferred which means

that recently observed speed is more important in the forecast than long term information. On the other hand, lower values for β are preferred which means that if trend is to be considered, it is best to update the trend gradually rather than update it immediately based on the slope of the last pair of observed speed.

It is interesting to note that in the model with additive seasonality and no trend, both the previous forecasted speed as well as the forecasted speed from the previous season are equally important. However, lower values for γ and β are preferred.

In fact, for all variants of exponential smoothing tested, low values for γ yield the optimal results.

4. CONCLUSION AND SOME SUGGESTIONS FOR FUTURE WORK

In this work, we studied the problem of short-term traffic speed forecasting using vehicle trajectory data. Various models were evaluated and it was shown that learning based techniques outperform those based on exponential smoothing although certain variants of the second group of methods also performed quite well. The optimal values for the parameters of the various models were also identified and analysed.

We now list down some possible directions for future work. Since we were unable to collect a substantial amount of vehicle trajectory data for forecasting, we had to resort to using simulated data. In the future, we would like to test the models on real-world trajectory data. At the moment, we assume that the paths the trajectories take are all defined. In the real-world, however, most collected trajectories are uncertain which means that the actual path connecting two consecutive GPS coordinates is uncertain or unknown [12]. A traffic forecasting system will have to take this into account.

We would also like to consider testing other techniques for forecasting. In particular, we are interested in testing ensemble methods [6] which can yield better performance than stand-alone techniques when applied appropriately. Eventually, we plan to install systems on taxis as well as various public transportation vehicles which will collect the GPS trajectories. This will result in a large stream of data being uploaded continuously to the system's server, especially during peak hours. There is a need to develop a system which can handle the load. Furthermore, the system will ideally employ an online learning strategy [16] so it can adapt dynamically.

Recent research have shown that spatial features are also important when forecasting short-term traffic condition [14]. In future work, we would like to test the effect of incorporating spatially relevant features as input to our learning models.

Finally, it would be interesting to see how we can extend the models based on exponential smoothing to make them better. This can be done by combining the models with other techniques or by redefining or reformulating certain parts.

5. REFERENCES

[1] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han. Online-svr for short-term traffic ow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, 36, 2009.

[2] H. Chen, S. Grant-Muller, L. Mussone, and F. Montgomery. A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Computing and Applications*, 10, 2001.

[3] J. B. Clemente, R. M. U. Tadlas, R. A. B. Juayong, J. A. Malinao, and H. N. Adorna. Visualization of data signatures using non-metric multidimensional scaling for visualization of data signatures using non-metric multidimensional for time series traffic data analysis. In *Proceedings of International Congress on Computer Applications and Computational Science*, 2010.

[4] M. V. Der Voort, M. Dougherty, and S. Watson. Combining kohonen maps with arima time series models to forecast traffic ow. *Transportation Research Part C*, 4(5), 1996.

[5] H. Dia. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*, 131(2), 2001.

[6] T. G. Dietterich. Ensemble methods in machine learning. In *Proceedings of First International Workshop on Multiple Classifier Systems*, 2000.

[7] B.-W. Huang, K.-W. Wang, L.-Y. Wei, and W.-C. Peng. A hybrid prediction algorithm for traffic speed prediction. In *Proceedings of the International Computer Symposium*, 2012.

[8] M. Hustim and M. I. Ramli. The vehicle speed distribution on heterogeneous traffic: Space mean speed analysis of light vehicles and motorcycles in makassar - indonesia. In *Proceedings of the Eastern Asia Society for Transportation Studies*, 2013.

[9] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2013.

[10] X. Jiang and H. Adeli. Dynamic wavelet neural network model for traffic ow forecasting. *Journal of Transportation Engineering*, 131(10), 2005.

[11] T. Kindberg, M. Chalmers, and E. Paulos. Guest editor's introduction: Urban computing. *IEEE Pervasive Computing*, 6(3), 2007.

[12] H. Liu, L.-Y. Wei, Y. Zheng, M. Schneider, and W.-C. Peng. Route discovery from mining uncertain trajectories. In *Proceedings of IEEE 11th International Conference on Data Mining Workshops*, 2011.

[13] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xie. Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of Conference on Knowledge Discovery and Data Mining*, 2011.

[14] W. Min and L. Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C*, 19, 2011.

[15] B. Park, C. J. Messer, and T. Urbanik. Short-term traffic volume forecasting using radial basis function

neural network. *Transportation Research Record*, 1998.

[16] S. Shalev-Shwartz. Online learning: Theory, algorithms, and applications, 2007.

[17] H. Wang, W. Fan, P. S. Yu, and J. Han. Mining concept-drifting data streams using ensemble classi ers. In *Proceedings of Conference on Knowledge Discovery and Data Mining*, 2003.

[18] L.-Y. Wei, Y. Zheng, and W.-C. Peng. Constructing popular routes from uncertain trajectories. In *Proceedings of Conference on Knowledge Discovery and Data Mining*, 2012.

[19] H. Xiao, H. Sun, and B. Ran. The fuzzy-neural network tra c prediction framework with wavelet decomposition the fuzzy-neural network tra c prediction framework with wavelet decomposition. In *Proceedings of 82nd Transportation Research Board Annual Meeting*, 2003.

[20] J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Transactions on Knowledge and Data Engineering*, 25(1), 2013.

[21] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2011.