Performance Analysis and Evaluation of Short-Term Travel Forecast Schemes Based on Cellular Mobile Services

Jamal Raiyn[1], Tomer Toledo[2]

Abstract

Various forecast schemes have been proposed to manage the travel data in transportation engineering. Many studies showed that the moving average schemes are offering meaningful results compared to other different forecast schemes. This paper deals with the moving average schemes, namely, simple moving average, weighted moving average, and exponential moving average. Furthermore, the performance analysis of the short-term forecast schemes is discussed and analyzed.

Keywords: *forecast scheme, historical information, transportation engineering*

Published in International Review of Civil Engineering 3(2), 2012

I. Introduction

Traffic congestions and road accidents continues to increase in industry countries. For instance, in the United States (Elvik, 2000)[1] traffic congestion is causing 4.2 billion delay hours and in 2007 had cost $78 billions. There are three basic strategies to relieve congestion (SooBeom, 2009)[2]. The first strategy is to increase the transportation infrastructure. However, this strategy is very expensive and can only be accomplished in the long term. The second strategy is to limit the traffic demand or make traveling more expensive that will be strongly opposed by travelers. The third strategy is to focus on efficient and intelligent utilization of the existing transportation infrastructures. This strategy is gaining more and more attention because its well. Currently, the Intelligent Transportation System (ITS) is the most promising approach to implement the third strategy. Various forecast schemes (Andrada-Felix and Fernandez-Rodriguez, 2008)[3], (Andrawis and Atiya, 2009)[4] have been proposed to manage the travel flow information. Meanwhile, the robustness and the accuracy of the exponential smoothing forecast is high and impressive. This paper reports on the performance of three moving average techniques in predicting average travel speeds up to 10 minutes ahead of time. The Results indicate that all three moving average methods have more or less similar performance in forecasting short-term travel times. However, as one would expect that method using optimized weights produces slightly better predictions at a higher computational cost. Moving average methods overestimate travel speeds in slow-downs and underestimate them as long the congestion is clearing up and speeds are increasing. In general, predictions are not reliable during highly congested periods, while they exhibit improvements in quality when traffic moves toward less congested and eventually free flow conditions. A comprehensive discussion about the nature and the performance of the studied moving average techniques is provided in this paper. The paper is organized as follow. Section 2, introduces the information collection based on cellular phone services. Section 3, introduces the short-term forecast scheme based on historical information. Finally, section 4, discusses the performance analysis of the proposed short-term forecast scheme based on exponential moving average.

II. Information Collection

The development of advanced technologies on data collection and data management lead to the collection of high quality accident data with numerous information types as well as the ability to manage that data more efficiently (Tu at al., 2008)[5]. There are two major strategies for the travel data collections; at the present time, the most and widely used technology is the traditional strategy that is based on the magnetic loop detectors, installed under the roadway surface. The lack of this technology expressed in the high cost of the installation and maintenance of the local detectors, they are typically installed only on a relatively small area of the roadway system, thus providing limited coverage of the entire transportation network. Furthermore in an urban environment there are many traffic interruptions. These interruptions cause delays that are not easily depicted by measuring speeds at any point along the road. To avoid equipment costs along the road network, a modern strategy that is based on the cellular phone service introduced (Zysman et. al., 2001)[6]. The Information collection based on cellular systems can be gathered on millisecond compared to the traffic data collection based on detectors. Conceptually, traffic information (Alger, 2004)[7], (Borzacchielo, 2010)[8] may fall into one of the three categories as follows; Historical information, real-time information, and Predictive information. The historical data is a collection of past observations of the system. Historical data describe the traffic states of a transportation system during previous time periods. It is mainly used to classify daily graphs or special events. Real-time information is most up-to-date and can be calculated, e.g., by on-line simulations. Predictive information, like traffic forecasts, can help to change the travel behavior of road users by providing information about the future state of the network. The real-time information achieved to update the historical adaptive information, special in the case that the real-time information does not matching the historical information. The historical information is necessary to carry out a plausible forecast in real-time. To develop a robust forecast model, it is needed to collect accurate travel information; firstly it is necessary to optimize the resource allocation in cellular systems. The optimization of the resource allocation in cellular system considers various issues like repeated handoff, radio spectrum coverage, call blocking probabilities, delay and interference. Our study and analysis has been showed that the mentioned issues are strongly influence the quality of the collected travel flow information. In the last time (second half of 20^{th} century) the phenomenon of traffic congestion has become predominant due to the rapid increase in the number of vehicles. Traffic congestion appears when too many vehicles attempt to use a common transportation infrastructure with limited capacity. To reduce the traffic congestion, several methods have been proposed (Chrobok *et al.* 2004)[9], such as Time Series, Kalman Filtering (Xie et al. 2007)[10], Neural Networks State Space reconstruct, Non-parameters Regression, were presented. Some of them are successful in practical or simulation forecast. But it’s determined by the complexity of traffic flow process and the properties of forecast methods that it's very difficult to get the accurate forecast results only using one method or model. Some model parameters need to be determined before the model is used to forecast the traffic volume, the structure of the model depends on these parameters, which are computed out from the historical and real time traffic data in practice. For a successful forecast of traffic flow, it ought to apperceive the variety of environment and can adjust the parameters automatically. Furthermore it is important that the forecast model takes into consideration the abnormal conditions that occurred in real-time (Zheng and Liu, 2009)[11]. The ability to achieve accurately forecast future link times in the transportation networks is a critical component for many transportation systems applications. Travel time in an urban traffic environment is highly time-dependant due to random fluctuations in travel demands, interruptions caused by traffic control devices, incidents, and weather conditions. It has been increasingly recognized that for many transportation applications, estimates of the mean and variance of travel times significantly affect the accuracy of the forecasting approaches. Figure1 describes the travel data collection based on cellular systems compared to the data collection based on sensors. The current system supports data resolution of 2.5min.

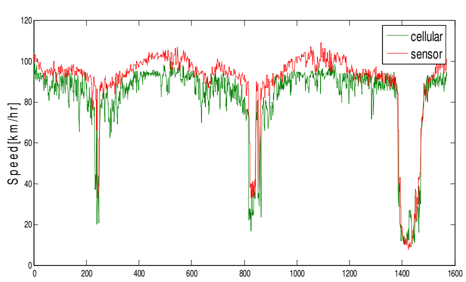

Fig. 1: sensors vs. cellular

III. Methodology

The purpose of this paper is to assess the strengths and limitations of available of the traffic data collection based on the cellular mobile service and their corresponding processing algorithms. The performance of an incident detection system is determined on two levels: data collection technologies and data processing algorithms. Variations in cellular mobile services and algorithm schemes result in a variety of solutions for incident detection. Various short-term traffic forecasting scheme have been proposed. This section introduces the forecast model based on the moving average. There are three types of moving average, that is, simple moving average (SMA), weight moving average (WMA), and exponential moving average (EMA). In this study, an exponential moving average is considered. An exponential moving average uses a weighting or a smoothing factor which decreases exponentially. The weighting for each older data point decreases exponentially, giving much more importance to recent observations while not discarding the older observations entirely. Figure 2 illustrates the proposed forecast model. The forecast model is divided into two phases, namely, detection phase, and forecast phase. The detection phase focused on the collected data analysis. To increase the accuracy of the forecast model, the abnormal events in the collected data should be considered. The forecast scheme is based on the exponential moving average. The robustness and accuracy of the exponential smoothing forecast is high and impressive. The accuracy of the exponential smoothing technique depends on the weight smoothed factor alpha value of the current demand. To determine the optimal alpha factor value, fitting curve has been considered.

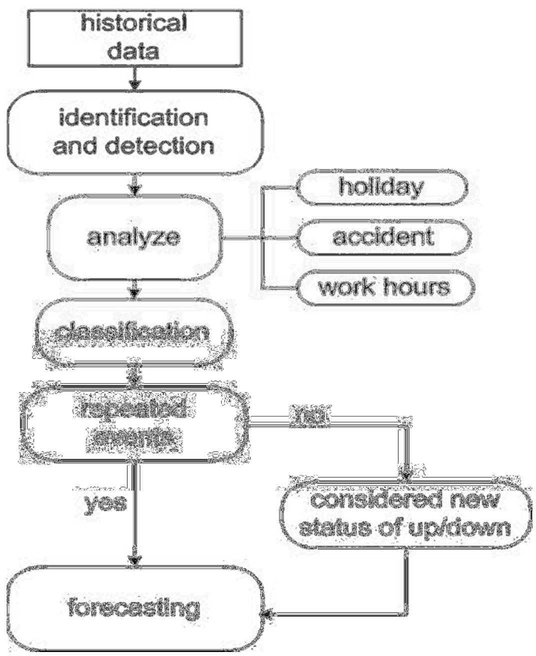

Fig. 2: Algorithm Process

III.1 Forecast based Historic Observations

The historical database is a collection of past travel observations of the system. Exponential smoothing is forecasting method that gives weight to the observed time series unequally. The unequal weight is accomplished by using one or more smoothing parameters, which determine how much weight is given to each observation. The major advantage of exponential smoothing methods is that gives good forecasts in a wide variety of applications. In addition, data storage and computing requirements are minimal, which makes exponential smoothing suitable for real-time application.

$$tt(t+1,k) = \alpha * tt^M(t,k) + (1-\alpha) * tt^H(t,k) \qquad (1)$$

Where $0 < \alpha \leq 1$, $tt^M(t, k)$ the actual travel time in section k at the time t. $tt^H(t, k)$ the historical travel time in section k at time t.

III.2 Smoothed parameter alpha

To achieve short-term traffic flow forecasting with high accuracy, the proposed forecast scheme required to optimize the smoothed parameter alpha. Alpha determines how responsive a forecast is to sudden jumps and drops. It is the percentage weight given to the prior, and the remainder is distributed to the other historical periods. Alpha is used in all exponential smoothing methods. The lower the value of alpha, the less responsive the forecast is to sudden change. The smoothing parameter "alpha" lies between 0 and 1. To determine the optimized smoothing factor, a sum of the square errors between the observed and the forecasted alpha dose rates was analyzed by increasing the smoothing filter factor from 0.1. Sum of the square errors is decreased as the smoothing filter factor is increased as shown in Figure 3.

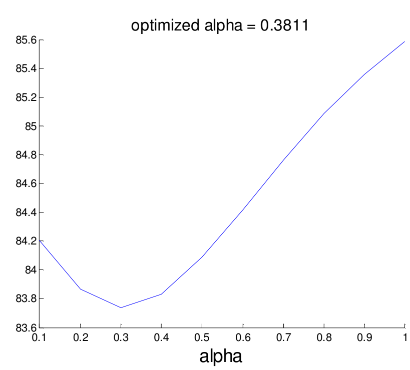

Fig. 3: Smoothed parameter alpha

IV. Performance Analysis

There are various measures of forecasting accuracy techniques proposed in the literature. The aim of this study is to evaluate forecast accuracy travel observations. The forecasting accuracy techniques are used to be able to select the most accurate forecast scheme for many. The forecasting performance of the various models and the measures of the predictive effectiveness was evaluated using various summary statistics. The comparing experiments are carried out under normal traffic condition and abnormal traffic condition to evaluate the performance of four main branches of forecasting models on direct travel time data obtained by license plate matching (LPM).

Statistical Measurements	
MAE	$\dfrac{\sum_{i=1}^{n}\left\lvert x_i - \bar{x}_i\right\rvert}{n}$
RMSE	$\sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2}{n}}$
RAE	$\dfrac{\sum_{i=1}^{n}\left(\lvert P_{ij} - x_i\rvert\right)}{\sum_{i=1}^{n}\left(\lvert x_i - \bar{x}\rvert\right)}$
RRSE	$\sqrt{\dfrac{\sum_{i=1}^{n}\left(P_{ij} - \bar{x}_i\right)^2}{\sum_{i=1}^{n}\left(x_j - \bar{x}_i\right)^2}}$
Theil's Coeff.	$\dfrac{\sqrt{MSE}}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i)^2}{n}} + \sqrt{\dfrac{\sum_{i=1}^{n}(\bar{x}_i)^2}{n}}}$

Table 1: measurements error

The MAE is a measure of overall accuracy that gives an indication of the degree of spread, where all errors are assigned equal weights. The MSE is also a measure of overall accuracy that gives an indication of the degree of spread, but here large errors are given additional weight. It is the most common measure of forecasting accuracy. Often the square root of the MSE, RMSE, is considered, since the seriousness of the forecast error is then denoted in the same dimensions as the actual and forecast values themselves. Mean square percentage error (MSPE) is the relative measure that corresponds to the MSE. The more commonly used measure is the root mean square percentage error (RMSPE). Theil's Coefficient is another statistical measure of forecast accuracy. One specification of theil's compares the accuracy of a forecast model to that of a naïve model. A theil's greater than 1.0 indicates that the forecast model is worse than the naïve model; a values less than 1.0 indicates that it is better. The closer U is to 0, the better the model.

IV.1 Modern vs. traditional traffic data

This section illustrates the simulation results and analysis of the implementation of the measured traffic speeds and travel time. The information of the dual magnet loop detectors will be compared to the information that is provided from cellular phone service. Based on the WEKA platform has been carried out analysis and comparison of different Prediction schemes. WEKA (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms for data mining tasks. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. WEKA is introduced to make comparison between the following schemes:

i. Smoothed Linear Models (LM)
ii. Tree Decision (TD)
iii. Nearest- Neighbor Classifier (NN)

The comparison is focused on various statistical measurements error, mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), and Theil's coefficient. The statistical measurements errors are summarized in Table 1. The results of the quality measurements are summarized in Table 2-4. Table 4 illustrates that the Nearest Neighbor Scheme offers a clear and the best results compared to the linear model and tree decision schemes.

Statistical Measur.	Linear model-LM	
	Cellular	Sensor
MAE	7.973	7.1967
RMSE	11.6976	11.5308
RAE	41.4674%	69.6342%
RRSE	49.1034%	68.9102%

Table 2: Cellular vs. sensor based on LM

Statistical Measur.	Tree decision-TD	
	Cellular	Sensor
MAE	9.5974	7.57
RMSE	14.0847	11.6718
RAE	49.916%	73.2468%
RRSE	59.1237%	69.753%

Table 3: Cellular vs. sensor based on TD

Statistical Measur.	Nearest Neighbour	
	Cellular	Sensor
MAE	6.4734	6.6224
RMSE	10.1594	11.0445
RAE	33.6678%	64.0777%
RRSE	42.6465%	66.0042%

Table 4: Cellular vs. sensor based on NN

statistical Measurements	SMA	WMA	EMA
MAE	6.22	8.11	5.17
RMSE	12.33	14.04	9.57
RAE	11.84	16.54	11.54
Theil's Coefficient	7.21	9.55	5.61

Table 5: SMA vs. WMA vs. EMA

IV.2 Simulation Results

Results indicate that all three moving average methods, SMA, WMA and EMA, have more or less similar performance in forecasting short-term travel times. However, as one would expect the method using optimized weights produced slightly better forecasts at a higher computational cost. Quality of forecasts is diminished as the time for which forecasts are made is farther in the future. Moving average methods overestimate travel speeds in slow-downs and underestimate them when the congestion is clearing up and speeds are increasing. Figure 4 illustrates the exponential moving average related to actual observations. EMA offers high accuracy based historical information. Figure 5 illustrates optimized exponential moving average compared to exponential moving average. Figure 6 describes the comparison between SMA, WMA and EMA based on the RMSE. Figure 7 illustrates the comparison between SMA, WMA and EMA based on the Theil's Coefficient. Figure 8 illustrates the comparison between SMA, WMA and EMA based on the MAE. Evaluation of performance analysis of moving average schemes illustrates that the exponential moving average scheme offers the best results.

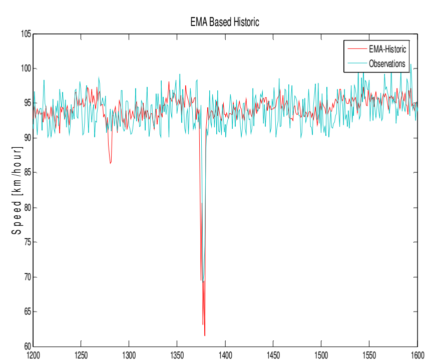

Fig. 4: EMA-H vs. Actual Observations

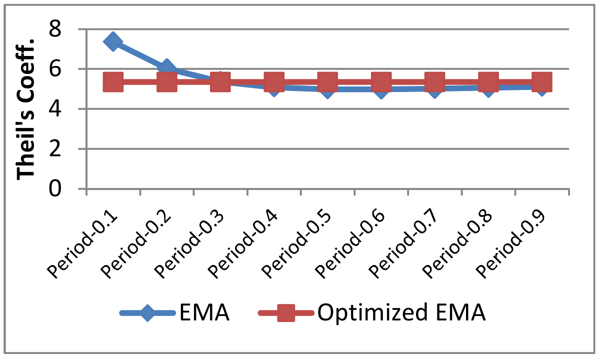

Fig. 5: Comparison between EMA and Opt-EMA

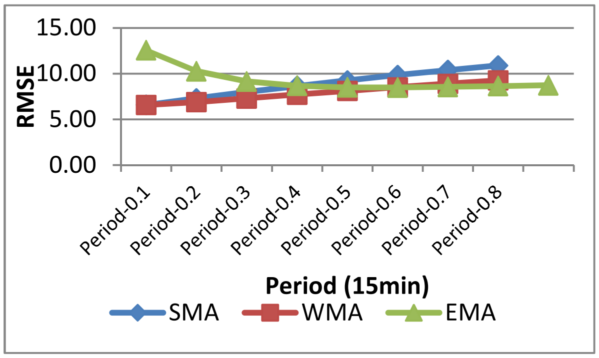

Fig. 6: SMA,WMA, EMA in Comparison

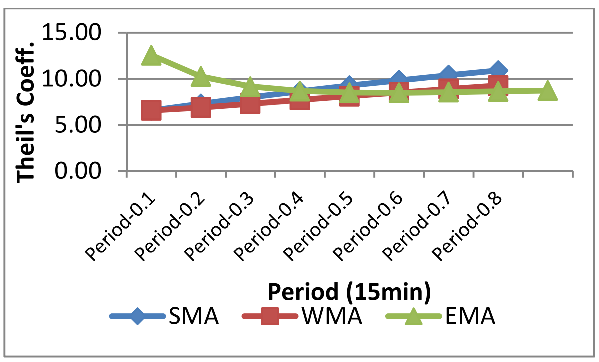

Fig. 7: Theil's Coeff.

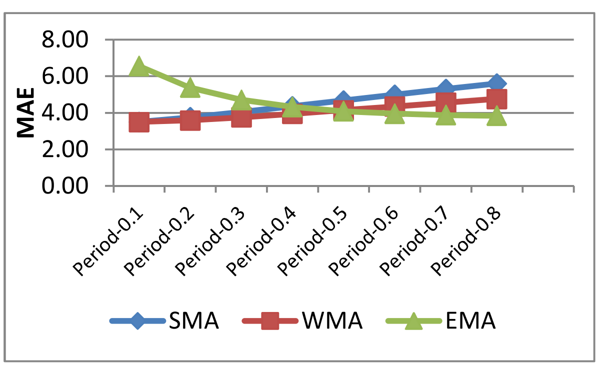

Fig. 8: MAE

V. Conclusion

Various forecast schemes have been proposed to manage the travel flow. This paper introduced various forecast schemes based on the historical data. Furthermore this paper discussed and summarized some forecast methods based on their performance analysis. Analysis illustrates that the exponential moving average is the most accurate method, when few data are available. Moreover the proposed algorithm has been given best solution for traffic travel forecast. However the road accidents increased rapidly. To reduce the incidents should be developed a new detection scheme that considers driver's behaviors.

Reference

[1] R. Elvik, How much do road accidents cost the national economy?. *Accid. Anal. Prev., 32: 2000, pp. 849-851.*

[2] L. SooBeom, Analysis of design elements for urban highway safety. *Proceedings of the 4th International IRTAD Conference on Road Safety, September 16-17, 2009, Seoul, Korea, 2009, pp: 251-262.*

[3] J. Andrada-Felix, and F. Fernandez-Rodriguez, Improving moving average trading rules with boosting and statistical learning methods. *J. Forecast., 27, 2008, pp.433-449.*

[4] R. R. Andrawis, and F.A. Atiya, A New Bayesian Formulation for Holt's Exponential Smoothing, *Journal of Forecasting, 28, 2009, pp.218-234.*

[5] H. Tu, H. V. Lint and H. V. Zuylen, The effects of traffic accidents on travel time reliability. *Proceedings of the IEEE Conference on Intelligent Transportation Systems, October 12-15, 2008, Beijing, China.*

[6] G.L. Zysman, H. Menkes and B. Qi, Wireless mobile communications at the start of the 21st Century. *IEEE Commun. Mag., 39: 2001, pp. 110-116.*

[7] M. Alger, Real-time traffic monitoring using mobile phone data. *Proceedings on 49th European Study Group with Industry. March 29-April 2, 2004, Oxford, England.*

[8] T.M. Borzacchielo, The use of data from mobile phone networks for transportation applications. *The Transportation Research Board (TRB) 89th Annual Meeting, January 10-14, 2010, Washington, USA.*

[9] R. Chrobok, O. Kaumann, J. Wahle and M. Schreckenberg, Different methods of traffic forecast based on real data. *Eur. J. Operational Res.,155, 2004, pp. 558-568.*

[10] Y. Xie, Y. Zhang and Z. Ye, Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition. *Comput. Aided Civ. Infrastruct. Eng., 22:5. 2007,pp. 326-334.*

[11] X. Zheng, and M. Liu, An overview of accident forecasting methodologies. *J. Loss Prev. Process Ind., 22: 2009, pp. 484- 491.*

Authors' Information

1 Jamal Raiyn
Computer Science Department,
Al-Qasemi, Academic College of Education,
P.O.BOX 124, Baka El-Gharbia 30100, Israel.
raiyn@qsm.ac.il

[2] Dr. Tomer Toledo,
Faculty of Civil and Environmental Engineering,
Department of Transportation Engineering,
Technion - Israel Institute of Technology,
Haifa, Israel.
toledo@technion.ac.il