Evaluation of Recent Computational Approaches in Short-Term Traffic Forecasting

Haofan Yang, Tharam S. Dillon, and Yi-Ping Phoebe Chen(✉)

Department of Computer Science and Information Technology,
La Trobe University, Melbourne, VIC 3086, Australia
Phoebe.Chen@latrobe.edu.au

Abstract. Computational technologies under the domain of intelligent systems are expected to help the rapidly increasing traffic congestion problem in recent traffic management. Traffic management requires efficient and accurate forecasting models to assist real time traffic control systems. Researchers have proposed various computational approaches, especially in short-term traffic flow forecasting, in order to establish reliable traffic patterns models and generate timely prediction results. Forecasting models should have high accuracy and low computational time to be applied in intelligent traffic management. Therefore, this paper aims to evaluate recent computational modeling approaches utilized in short-term traffic flow forecasting. These approaches are evaluated by real-world data collected on the British freeway (M6) from 1st to 30th November in 2014. The results indicate that neural network model outperforms generalized additive model and autoregressive integrated moving average model on the accuracy of freeway traffic forecasting.

Keywords: Computational approach · Short-term traffic forecasting

1 Introduction

Traffic during peak hours in developed and developing countries is usually congested and such problem is apace rising since the past three decades. Computational technologies particularly intelligent systems are expected to help the rapidly increasing traffic congestion problem in recent traffic management. The use of intelligent systems is an important trend in traffic management, and its objective is to provide innovative services to transportation development and enable various users to better use of transport networks. Advances in computing and communications technologies promote using intelligent systems to manage many problems in transportation, especially the traffic congestion [1-3]. Congestion can be reduced by redistributing traffic spatially and temporally. To achieve traffic flow redistribution, it is necessary to get future traffic conditions [4, 5]. The authors in [6] also emphasized that it is necessary to continuously forecast the traffic conditions for short time ahead to enable dynamic traffic control. These cause establishing efficient and accurate traffic flow forecasting models to become an important issue in traffic management.

© IFIP International Federation for Information Processing 2015
T. Dillon (Ed.): IFIP AI 2015, IFIP AICT 465, pp. 108–116, 2015.
DOI: 10.1007/978-3-319-25261-2_10

Lieu, et al. [7, 8] stated that the way to establish a reliable and efficient intelligent traffic system relies on providing continuous information of the traffic conditions over time. Such traffic information need to be updated in a timely fashion and should generate projections on the expected traffic networks [9]. In the new era of intelligent traffic systems, research has focus on establishing forecasting models to manage the traffic networks [10-12]. Many computational approaches have been commonly applied to build forecasting models, such as neural network (*NN*), generalized additive model (*GAM*), and autoregressive integrated moving average (*ARIMA*) [6, 13-16]. In this paper, above approaches are applied to the traffic data collected on the British freeway (M6) from 1st to 30th November in 2014 for evaluation. We evaluate three different approaches and report on their performance. The rest of this paper is organized as follows. Section 2 describes the computational modeling approaches used in this study. Section 3 explains the experimental design. Section 4 discusses the experiment results. Finally, the conclusion is given in Section 5.

2 Description of Used Computational Modeling Approaches

2.1 Neural Network *(NN)*

NN is a kind of information processing technique, and it can be trained to learn relationships in a dataset. The NN model has been applied in short-term traffic forecasting for many years, and it has been proven to be effective in solving problems that existing complex relationships, such as traffic flow forecasting [13, 17-19]. Since the traffic data with lumpiness may reduce the generalization capability on the short-term traffic flow forecasting on unseen data [20], we applied the exponential smoothing method [21, 22] to remove lumpiness in the collected traffic data. After removing the lumpiness, a three-layer feed-forward *NN* with levenberg-Marquardt algorithm was trained to establish a short-term traffic forecasting model. The traffic data is collected from n sensing stations ($S_1, S_2, S_3,\ldots,S_n$), which are located on the freeway (M6). S_i captures two traffic condition measures, i.e., the average vehicle speed $\hat{V}_i(t)$ and the average headway $\hat{h}_i(t)$ between time t and time $t+T_S$, where T_S is the sampling time. Future traffic conditions can be forecasted by the *NN* model, according to the current and past traffic conditions. The current traffic condition is denoted by the current average speed $\hat{V}_i(t)$ and current average headway $\hat{h}_i(t)$. The past traffic condition is indicated by the past average speed $\hat{V}_i(t-kT_s)$ and past average headway $\hat{h}_i(t-kT_s)$, which was collected by Si at time $(t-kT_s)$ with $i=1,2,3,\ldots,n$ and $k=1,2,3,\ldots,p$, whereas the past traffic data within p sampling time period are collected. The future traffic condition can then be generated by the *NN* model, which is indicated by $\hat{h}_L(t+mT_s)$ passing through the L^{th} sensing station S_L at time $(t+mT_s)$, where future traffic condition with m sampling time ahead is forecasted. The ***NN*** model is formulated as:

$$\hat{h}_L(t+mT_s) = \sum_{i=1}^{n}\sum_{j=1}^{M}\left[\beta_{j,i}^{V}\left(\gamma_{0,j,i}^{V} + \sum_{k=1}^{p}\gamma_{k,j,i}^{V}V_i(t-kT_s)\right)\right.$$
$$\left. + \beta_{j,i}^{h}\left(\gamma_{0,j,i}^{h} + \sum_{k=1}^{p}\gamma_{k,j,i}^{h}h_i(t-kT_s)\right)\right] + \alpha_0 \quad (1)$$

where *M* is the number of nodes in the hidden layer; () is the activation function of the hidden set (sigmoid function is used in this study); $\beta_{j,i}^{V}$, $\beta_{j,i}^{h}$, $\gamma_{0,j,i}^{V}$, $\gamma_{0,j,i}^{h}$, $\gamma_{k,j,i}^{V}$, $\gamma_{k,j,i}^{h}$, and α_0 are the parameters of the *NN* model.

2.2 Generalized Additive Model (*GAM*)

The *GAM* has the ability to allow non-parametric fits with relaxed assumptions between predicted and actual values. With this ability, the *GAM* can provide better fits to data than purely parametric models. The *GAM* is based on the additive model (*AM*), which is a nonparametric regression method that assumes the mean of a response variable depends on predictors by a nonlinear function. The *AM* relates a univariate response variable to a set of other response variables. The *GAM* extends the generalized *AM* which assumes linear dependence, by allowing the dependence of the response variable to be nonlinear [23]. The *GAM* is generalizing the *AM* to allow the response variable to follow an exponential family distribution [24, 25] along with a link function. Given a data set $\{y_i, x_{i1}, x_{i2}, \dots, x_{ip}\}_{i=1}^{n}$ of data size *n*, where y_i and x_i are the i^{th} observations of the traffic flow, the *AM* is formulated as:

$$y_i = \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i \quad (2)$$

where $f_j(x_{ij})$ is a smooth function of the i^{th} observation of covariate *j*, and ε_i is the residual. With a link function *G*, the *GAM* can be structured as:

$$\mathrm{G}\left(E\left(y_i | x_{i1}, x_{i2}, \dots, x_{ip}\right)\right) = f_0 + \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i \quad (3)$$

where f_0 is the intercept that is equal to the overall mean of the y_i [23, 25]. In this study, a local scoring algorithm which maximizes a likelihood function is used to estimate the used smooth functions in *GAM*. Also, the generalized cross-sectional validation approach is used to avoid over-fitting in the short-term traffic flow forecasting.

2.3 Auto-Regressive Integrated Moving Average (*ARIMA*)

The *ARIMA* model, one of the most popular time series approach applied in traffic flow forecasting, is generally referred as *ARIMA (p, d, q)*, where *p, d,* and *q* are non-negative integers [26]. *p, d,* and *q* refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. To empirically create a proper model, we followed the Box-Jenkins *ARIMA* modeling procedure which consists of the following steps: identification, estimation, diagnostic checking, and forecasting [27]. Through continuous modification, the most proper forecasting model can be generated.

In the identification step, the collected traffic data is assessed to determine whether it is stationary. Autocorrelation function (*ACF*) and Partial Autocorrelation Function (*PACF*) are used to detect it [28]. If the collected traffic data is not stationary, the differencing approach is applied. The value *d* is the lowest order of differencing applied to achieve stationary. The second step is to decide whether *AR(p)* or *MA(q)* should be used in the model. According to the *ACF* and *PACF* plots of the differenced series, it is possible to tentatively identify the value p and/or q. The error residuals can be calculated when the candidate model *ARIMA (p, d, q)* is estimated. The candidate model is further tested by Akaike Information Criterion (*AIC*) and Schwartz's Bayesian Criterion (*SBC*) in diagnostic checking step [29]. The lower value of *AIC* and *SBC*, the more suitable *ARIMA (p, d, q)* should be. Q-test is also used to determine whether the estimated model is suitable. If the candidate model cannot pass the test, the process should go back to the identification step to develop a better model. The candidate model which passed diagnostic checking is set as the used *ARIMA* model in short-term traffic flow forecasting. In the *ARIMA* model, we predict $\hat{h}_L(t+mT_s)$, the average traffic flow at the L^{th} sensing station S_L at time $(t+mT_s)$, where future traffic condition with *m* sampling time ahead is forecasted based on the past traffic flow condition. The past traffic flow condition P_{tf} is defined as

$$P_{tf} = (1-L)^d \hat{h}_L(t+mT_s) \quad (4)$$

L is the lag operation and the *ARIMA* model is structured as:

$$\left(1-\sum_{i=1}^{p} \Phi_i L_i\right) P_{tf} = \left(1+\sum_{i=1}^{q} \theta_i L_i\right) \varepsilon_T \quad (5)$$

where p is the order of autoregressive and q is the order of moving average. Φ_k and θ_k are the parameters of the autoregressive and moving average parts, respectively. ε_T are error terms for the stationary distribution.

3 Experimental Design

The traffic flow data was collected from the British traffic warehouse, and it was captured by ANPR cameras, in-vehicle GPS and inductive loops built into the road surface located between junctions, J40 and J41, of the freeway M6 in the United Kingdom from 1st to 30th November in 2014. The main aim of traffic flow forecasting is to help for reducing traffic congestion, therefore the peak traffic period on the business days should be the main target. The traffic time periods are set to 15-minute intervals in the day which also refers to 0 to 95 where 0 indicates 00:00 to 00:15 and 95 refers to 23:45 to 24:00. The collected traffic data was analyzed and it was found all business days had similar traffic headway distribution as shown in Fig. 1. The figure illustrates how the data have daily seasonal patterns, which are periodically repeated. The peak traffic flow condition was occurred in mornings between 7:00 and 9:30 (time period 28 to 37) and evenings between 17:00 and 19:30 (time period 68 to 77) on the business days.

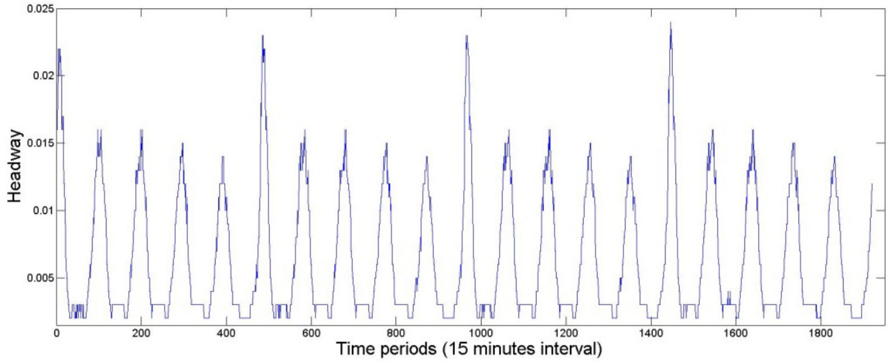

Fig. 1. Traffic headway distribution on the business day between 1st to 30th November in 2014. The time period index (0 to 1920) starts on Monday 00:15 at 3rd November and ends on Friday 24:00 at 28th November in 2014 with 96 intervals (15-minutes intervals) per day.

To train the models (*NN*, *GAM*, and *ARIMA*) of the selected road section, we excluded weekend traffic data since it has different daily patterns. The total number of days used for evaluation is 20 (business days between 3rd to 28th November). The significance level is set to 5% and the confidence interval is 95% in all analyses used in this study. The remained traffic data was divided into two subsets. The data of first subset, namely the training set, was used for establishing the *NN*, *GAM*, and *ARIMA* models. The data of second subset, namely the test set, was applied to evaluate the generalization capability of the trained models. The business days of the first three weeks in November 2014 were set in the training set and the business days of the last week in November 2014 were concluded in the test set. For the time period with 15-minute intervals, training data set and testing data set contain 1440 (3*5*96) data and 480 (1*5*96) data, respectively. Since the generalization capability on traffic flow forecasting would be improved by removing the lumpiness from the raw traffic data, we applied the exponential smoothing method to remove lumpiness in the collected data before applying the data to develop proposed models.

In order to evaluate the accuracy and efficiency of the *NN*, *GAM*, and *ARIMA* models, the performance of these models was evaluated by Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE and RMSE indicate the mean and the sample standard deviation of the differences between observed values and predicted values, respectively.

4 Experiment Results and Discussion

Since the main target in this study is the peak traffic period on the business days, we evaluate the forecast an hour ahead (in the peak traffic period 34 to 37 or 74 to 77) using the models developed in the training phase. The traffic flow conditions on the road section M6-J40 to M6-J41 from the previous one and half hours (in the peak traffic period 28 to 33 or 68 to 73) of the same peak traffic period (morning or evening peak) are used, i.e., the previous 6 records of a peak traffic period are used to predict 4 records ahead during the same peak traffic period. The forecasting models (NN, GAM, and $ARIMA$) are evaluated in different business days using the testing data set. The forecasted results are compared with real observations in order to know the performance of each model which is evaluated by the MAPE and RMSE. The $ARIMA$ model considers the number of lags to forecast a future value, therefore it is used to forecast the time $(t + 2 * mT_s)$ value.

Table 1. Comparison of forecasting models based on MAPE in peak traffic periods. Morning peak 34 to 37 indicates 08:30 to 09:30 and evening peak 74 to 77 denotes 18:30 to 19:30.

Traffic Peak period	Model	MAPE%				
		Mon	Tue	Wed	Thu	Fri
Morning peak 34 to 37	*NN*	2.19	1.85	1.99	2.11	2.61
	GAM	2.76	4.02	4.28	3.36	3.63
	ARIMA	13.32	15.17	16.05	14.54	14.91
Evening peak 74 to 77	*NN*	1.73	2.08	2.24	2.83	1.90
	GAM	3.04	3.88	3.61	2.98	4.11
	ARIMA	17.43	17.58	15.68	16.38	17.09

The forecasting performance of *NN*, *GAM*, and *ARIMA* models applied to peak traffic periods of the testing data set are presented in Table 1 and Table 2. For the comparison results displayed in Table 1 and Table 2, we notice that the MAPEs and RMSEs of the *NN* model are smaller than those of the GAM and ARIMA models during the morning peak and evening peak. Fig.2. shows the comparison between the observed traffic headway values and the predicted headway values which are generated by *NN*, *GAM* or *ARIMA* models on Monday 24th November 2014. The comparison between the observed traffic headway values and the predicted headway values on 25th, 26th, 27th, and 28th November (Tuesday to Friday) presents similar results as Monday 24th.

Table 2. Comparison of forecasting models based on RMSE in peak traffic periods. Morning peak 34 to 37 indicates 08:30 to 09:30 and evening peak 74 to 77 denotes 18:30 to 19:30.

Traffic Peak period	Model	RMSE				
		Mon	Tue	Wed	Thu	Fri
Morning peak 34 to 37	*NN*	1.20	1.01	1.36	1.05	1.14
	GAM	2.38	1.54	3.05	1.56	1.68
	ARIMA	5.60	6.18	4.92	7.60	6.57
Evening peak 74 to 77	*NN*	1.15	1.20	1.72	1.41	1.53
	GAM	2.65	1.69	2.85	1.88	2.03
	ARIMA	7.43	7.26	4.91	8.66	7.59

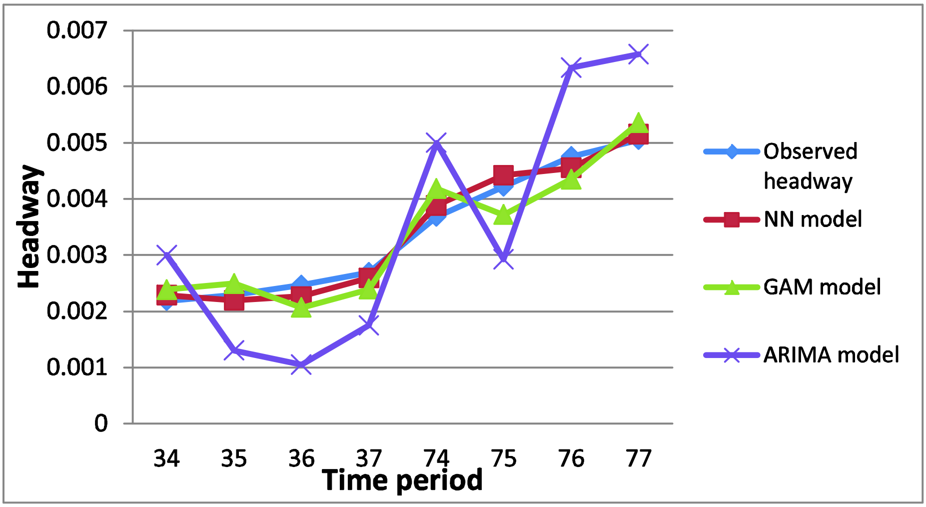

Fig. 2. Headway comparison between observed and predicted values on Monday 24th November 2014. 34 to 37 and 74 to 77 refer to 08:30 to 09:30 and 18:30 to 19:30, respectively.

Both Tables and Fig. 2 show that the *NN* model performs much better than the other models as its predicted headway values are close to the observed headway values. Moreover, its MAPE and RMSE values are less than *GAM* and *ARIMA* models.

5 Conclusion

Due to the fact that traffic congestion problem is rapidly increasing, it is urgent to establish intelligent traffic control systems that can redistribute the traffic flow spatially and temporally. In order to reach it, an efficient and accurate forecasting model is necessary to assist real time traffic control systems. Research has applied many computational approaches on short-term traffic forecasting to build forecasting models. While the traffic flow data of a road link can be collected much easier than before, selecting appropriate computational approaches to deal with the demand of forecasting accuracy and efficiency is getting significant. In this paper, we proposed to evaluate some computational approaches in short-term traffic flow forecasting. From the experiment results, we found that on forecasting accuracy, *NN* model outperforms *GAM* and *ARIMA* models on forecasting accuracy of the short-term freeway traffic forecasting.

References

1. Gu, H., Wang, W., Hu, Y., Qiao, L., Zhan, F.: Study on the intelligent transport system and social economic development. In: CICTP: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient, 2012, pp. 845–855 (2012)
2. Ben-Akiva, M.E., Gao, S., Wei, Z., Wen, Y.: A dynamic traffic assignment model for highly congested urban networks. Transportation Research Part C: Emerging Technologies **24**, 62–82 (2012)
3. Hamilton, A., Waterson, B., Cherrett, T., Robinson, A., Snell, I.: The evolution of urban traffic control: changing policy and technology. Transportation planning and technology **36**, 24–43 (2013)
4. Daganzo, C.F.: A behavioral theory of multi-lane traffic flow. Part II: Merges and the onset of congestion. Transportation Research Part B: Methodological **36**, 159–169 (2002)
5. Kamal, S., Imura, J.-I., Hayakawa, T., Ohata, A., Aihara, K.: Smart Driving of a Vehicle Using Model Predictive Control for Improving Traffic Flow. IEEE Transactions on Intelligent Transportation Systems **15**, 878–888 (2014)
6. Daraghmi, Y.A., Chih-Wei, Y., Tsun-Chieh, C.: Negative Binomial Additive Models for Short-Term Traffic Flow Forecasting in Urban Areas. IEEE Transactions on Intelligent Transportation Systems **15**, 784–793 (2014)
7. Lieu, H., Gartner, N., Messer, C., Rathi, A.: Traffic flow theory. Public Roads **62**, 45–47 (1999)
8. Lieu, H.C.: Traffic estimation and prediction system. 0738-6826 (2000)
9. Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G.: Short-term traffic forecasting: Overview of objectives and methods. Transport reviews **24**, 533–557 (2004)
10. Zhang, Y., Zhang, Y., Haghani, A.: A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. Transportation Research Part C: Emerging Technologies **43**, 65–78 (2014)
11. Wu, S., Yang, Z., Zhu, X., Yu, B.: Improved k-nn for Short-Term Traffic Forecasting Using Temporal and Spatial Information. Journal of Transportation Engineering **140** (2014)
12. Zhang, X., Onieva, E., Perallos, A., Osaba, E., Lee, V.C.: Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Transportation Research Part C: Emerging Technologies **43**, 127–142 (2014)
13. Chan, K.Y., Dillon, T., Chang, E., Singh, J.: Prediction of short-term traffic variables using intelligent swarm-based neural networks. IEEE Transactions on Control Systems Technology **21**, 263–274 (2013)
14. Chen, C., Hu, J., Meng, Q., Zhang, Y.: Short-time traffic flow prediction with ARIMA-GARCH model. In: Intelligent Vehicles Symposium (IV), pp. 607–612. IEEE (2011)
15. Tjondronegoro, D.W., Chen, Y.-P.: Knowledge-discounted event detection in sports video. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans **40**, 1009–1024 (2010)

16. Nahar, J., Chen, Y.-P.P., Ali, S.: Kernel-based naive bayes classifier for breast cancer prediction. Journal of biological systems **15**, 17–25 (2007)
17. Chan, K.Y., Khadem, S., Dillon, T.S., Palade, V., Singh, J., Chang, E.: Selection of significant on-road sensor data for short-term traffic flow forecasting using the Taguchi method. IEEE Transactions on Industrial Informatics **8**, 255–266 (2012)
18. Vlahogianni, E.I., Karlaftis, M.G.: Testing and comparing neural network and statistical approaches for predicting transportation time series. Transportation Research Record: Journal of the Transportation Research Board **2399**, 9–22 (2013)
19. Vanajakshi, L., Rilett, L.R.: A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In: Intelligent Vehicles Symposium, pp. 194–199. IEEE (2004)
20. Chan, K.Y., Dillon, T.S., Singh, J., Chang, E.: Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm. IEEE Transactions on Intelligent Transportation Systems **13**, 644–654 (2012)
21. DeLurgio, S.A.: Forecasting principles and applications (1998)
22. Wanas, N., Auda, G., Kamel, M.S., Karray, F.: On the optimal number of hidden nodes in a neural network. In: IEEE Canadian Conference on Electrical and Computer Engineering, 1998, pp. 918–921 (1998)
23. Hastie, T.J., Tibshirani, R.J.: Generalized additive models, vol. 43. CRC Press (1990)
24. Wood, S.: Generalized additive models: an introduction with R. CRC press (2006)
25. Wood, S.N., Goude, Y., Shaw, S.: Generalized additive models for large data sets. Journal of the Royal Statistical Society: Series C (Applied Statistics) **64**, 139–155 (2015)
26. Williams, B.M., Hoel, L.A.: Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of transportation engineering **129**, 664–672 (2003)
27. Pankratz, A.: Forecasting with univariate Box-Jenkins models: Concepts and cases, vol. 224. John Wiley & Sons (2009)
28. Liu, L.-M.: Identification of seasonal ARIMA models using a filtering method. Communications in Statistics-Theory and Methods **18**, 2279–2288 (1989)
29. Claeskens, G., Hjort, N.L.: Model selection and model averaging, vol. 330. Cambridge University Press, Cambridge (2008)