

RESEARCH ARTICLE

Use of artificial neural networks and support vector machines to predict lacking traffic survey data

Mohammed Saiful Alam Siddiquee¹ and Kalum Priyanath Udagepola^{2*}

¹ Department of Civil Engineering, Faculty of Engineering, King Abdul Aziz University, Jeddah, Saudi Arabia.

² Department of Information and Computing Sciences, Scientific Research Development Institute of Technology Australia, Brisbane, Australia.

Revised: 21 December 2016; Accepted: 19 January 2017

Abstract: The aim of this paper was to predict lacking data from a traffic survey along a principal highway in Bangladesh using artificial neural network (ANN) combined with the support vector machine (SVM). Traffic data were obtained at an hourly rate using a methodical inquiry over a four-year period at the Jamuna toll collection point, which is located along the North Bengal corridor of Bangladesh. Two evolutionary computational statistical procedures were used along with its corresponding numerical model. The neural network and SVM were fed with data from 13 recurring weeks over a four-year period. The missing data were predicted with significant accuracy using both methods. Accuracy of the methods was compared, which showed that the SVM method is much more accurate than the ANN technique. Combination of both the ANN and SVM models can be used to obtain trends in traffic data more accurately.

Keywords: Artificial neural network, estimation, support vector machine, traffic data prediction.

INTRODUCTION

In the absence of long duration traffic count data, there are two approaches for the estimation of annual average daily traffic (AADT) from short counts, either by using a regression model or by applying previously established expansion factors of similar facilities (Garber & Hoel, 1990). Several studies (Kaub, 1988; Sharma, 1989) carried out in developed countries have shown that if the continuous flow data is available for a particular road segment of a highway, a model could be built considering the recursive pattern of traffic flow in order to predict the AADT from short count (i.e. less than 24 hours) data. In India (Kumara *et al.*, 2013) an attempt

has been made to establish a trend in traffic volumes pertaining the daily, weekly, monthly and yearly flow patterns, and to use these particulars for forecasting traffic volumes. Short-term traffic flow is a major feature in areas such as traffic management and control (Abdi & Moshiri, 2015). There are chaotic and frantic properties displayed in short-term traffic. Despite some studies that prioritise long-term traffic flow forecasting, we have considered short-term forecasting (Sun *et al.*, 2008). In long-term traffic flow forecasting, events such as school timings, holidays and other periodic events are utilised for judging the future traffic flows (Cetiner *et al.*, 2010). In this study, by applying the short-term forecaster, the minimum resolution of forecasting is one minute, which is obtained from a 2-lane eastbound stretch of the I-494 freeway in Minnesota, USA. But in the case of long-term traffic flow forecasting, five minutes or one-hour traffic data such as historical data at each major junction in the city are considered (Wang *et al.*, 2006; Cetiner *et al.*, 2010; Chiou *et al.*, 2010). Although these studies were carried out in developed, and developing countries such as India many years ago, so far no such effort in this subject has been made in Bangladesh. One of the main reasons as stated earlier is the lack of long duration round-the-clock flow data. Moreover, to make a time series model in the context of Bangladesh is somewhat complex and intricate in nature as compared to that of developed countries for the following reasons:

- The shifting nature of the month of Ramadan and Eid festival greatly affects the traffic flow pattern and makes it difficult to predict monthly expansion factors. In the western world, main festivals like Christmas, New Year or even summer vacation do not shift with

* Corresponding author (kalumu@srdita.com.au)

- the calendar year.
- During the monsoon season, a substantial portion of the bulk freight is transported through waterways. Since the duration of monsoon is not constant, it is difficult to develop a similar yearly flow pattern.
 - Unusual events such as hartals (political strikes), transport strikes, etc., which greatly affect normal traffic flow are more frequent in Bangladesh than in developed countries.
 - Natural calamities like heavy rainfall and long flood duration is also an external parameter, which needs to be considered at the same time to evaluate the pattern of traffic flow in Bangladesh (Roads and Highways Department, 1994; 2005).

The design considerations and performance evaluation of highways is impossible due to incorrect or assumed traffic data (Matson *et al.*, 1955; Barewood, 1965; Pignataro, 1973). This may render the road networks to be inadequately facilitated or over-designed. This problem arises due to the unavailability of long duration continuous traffic data upon which a reliable pattern analysis can be done and expansion factors or equations can be determined. However, the collected and preserved traffic flow data on the Jamuna Bridge by the toll operators is so far providing a good opportunity for the transportation planners to utilise the flow data productively. The present study uses a fair amount of quality traffic data for the analyses of traffic flow characteristics and expansion of short term counts in the North Bengal corridor. The data for the short hourly counts of traffic data are collected by a team of consultants from the Bangladesh University of Engineering and Technology (BUET) over a period of four years, using four weeks per annum.

Due to the sparse form of short count data from the survey of BUET consultants, artificial neural network (ANN) and support vector machine (SVM) were employed to interpolate and extrapolate the short count data whenever necessary. The predictions and missing data interpretations by these two methods are compared in this study without the effects of road characteristics (Yeh, 2004).

For traffic parameter estimation, a method preferably using artificial intelligence (AI) is used (Yin *et al.*, 2002). In the midst of other existing methods in AI, neural network (NN) approach is mostly deployed in traffic prediction issues (Dougherty *et al.*, 1994; Ledoux, 1997). On the other hand, an approach using ANN will not require defining the relationship between casual input vector and output vectors; this makes it beneficial. Furthermore, ANN techniques take time to process on a computer because of its error minimisation strategy,

which simulates any level of non-linear complexities of a framework in a more direct way. Additionally, NN techniques have been utilised for short-term estimation and traffic parameter forecast (Dougherty *et al.*, 1994; Ledoux, 1997). Back-propagation of NN for forecasting of traffic has been used by researchers (Dochy *et al.*, 1996; Dougherty & Cobbett, 1997). In Utrecht, Rotterdam and the Hague region of the Netherlands, back-propagation of NN was used in predicting short-term traffic flow, rate and occupancy (Dougherty & Cobbett, 1997; Chiou *et al.*, 2014). A more direct application of NN is in the prediction of traffic flow along a motorway link (Kirby *et al.*, 1997). Every one of these tasks builds the use of the NN approach in the measurement of the rate of traffic stream in a mixed traffic environment. For these various reasons, ANN was used in this research (Federal Highway Administration, 2001; Basu, 2006).

SVM is relatively new in the field of traffic flow estimation. Unlike ANN, SVM has a unique solution to any regression problem, and it can be solved in a fixed number of steps. Their success in practice is a result of the solid theoretical foundations based on Vapnik-Chervonenk theory because its abstraction performance relies not on the magnitude of the input space. Expert users perform many SVM regression application studies due to the good understanding of its methodology. However, the fundamental issue for experts trying to apply SVM regression is on setting these parameter values for a particular set of data, as the quality of SVM models depends on a proper setting of SVM meta-parameters. A serial clustering method based on the extended entropy information bottleneck theory has been proposed by Yang *et al.* (2006). This theory is used to split a day into different time sections according to the record of traffic flow data. Thus, SVM is used to predict the amount of traffic flow. In another study, Bayesian inference has been used to fix the SVM kernel parameters to improve the regression precision. In order to gain an excellent SVM forecasting model and to understand the accurate prediction of urban traffic flow, a genetic algorithm SVM is used to predict urban traffic flow. In order to improve the parameters of the SVM, a genetic algorithm is introduced to optimise the criterion of SVM in this model, which can provide the optimised prediction model (Xiao-bin, 2010). The genetic algorithm support vector model (GA-SVM) proposed by Xiao-bin (2010) shows that GA-SVM returns a more accurate result than ANN. This is deduced from the experimental results. Kumara *et al.* (2013) used ANN for forecasting of traffic flow using old traffic data for short term prediction. The model incorporated traffic volume, speed, density, time and day of a week as input variables (Xu *et al.*, 2007).

Results showed that the ANN has consistent performance even if the time interval for traffic flow prediction was increased from 5 minutes to 15 minutes and produced good results even when the speeds of each category of vehicles were considered separately as input variables.

METHODOLOGY

It is already mentioned that daily traffic data collected by the toll collecting authority [Jamuna Operation and Maintenance Company (JOMAC)] is continuous over a period of 4 years (from 1998 to 2002). However, the hourly data is not continuous over this period. The total 13-weeks-survey of hourly data was collected over a time of 4 years. So it is rather difficult to directly link this hourly data to the daily average data from the toll collecting authority. With this end in view, a novel idea is introduced, and an ANN was trained with limited data from the BUET consultants' survey. First, the hourly data was trained with respect to its hour, day, month and year. After a successful training of hourly data with the ANN, it was possible to predict the missing hourly data on other hours of the missing days. Although the data used for training was not continuous over time, it could predict the missing data as well as forecasted data. In order to train the ANN, an in-house developed ANN software was used. The interface of the software is shown in Figure 1.

In order to model the traffic flow data by SVM, a non-commercial SVM software (SVMlight) was used. This software is an implementation tool of the Vapnik's support vector machine (Vapnik, 1995) to address the problem of pattern recognition, regression problems and the rank learning function. SVMlight has used optimisation algorithms, which are described by Joachims (1998a; 1999a; 2002a). The algorithm is

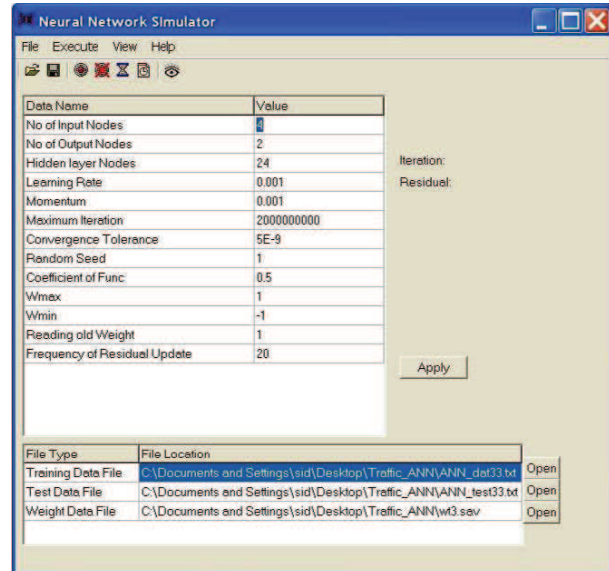


Figure 1: Main interface of the ANN programme used for training and prediction of traffic data

efficient and is capable of handling thousands of support vectors as well as having a scalable memory. The assessment of generalisation performance proves to be efficient using the SVMlight software, which includes two efficient estimation methods in both detection of errors and precision. On the other hand, XiAlpha's estimation (Joachims, 1998b; 1999b; 2000; 2002a) is conservatively biased, but it can be estimated with negligible costs. Approximately unbiased computations provide leave-one-out testing. However, more than 99 % of the results of most leave-one-outs are often exploited by SVMlight, which is usually fixed and does not require computation (Joachims, 2002a; Robichaud & Gordon, 2002).

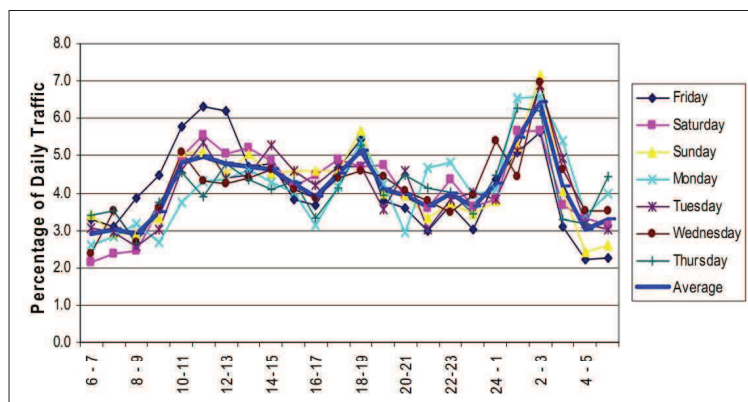


Figure 2: Hourly flow variation on Jamuna Multipurpose Bridge
(Source: 13 surveys conducted by BUET)

There are four major highway corridors in Bangladesh: 1. Dhaka-Chittagong highway; 2. Dhaka-Sylhet highway; 3. Dhaka-North Bengal corridor and 4. Dhaka-Aricha road to South Bengal. Different corridors exhibit distinct characteristics because of the variations in economic activities. All these considerations are to play a role in flow pattern analysis and the development of the framework model. However, in either case, for an efficient analysis of the features of flow data on rural highways, a large quantity of reliable data is required.

The Jamuna Multipurpose Bridge was opened to the public in June 1998. Since then, its operators have collected and preserved daily traffic data in an orderly fashion. Considering that it has taken 6 months for the corridor to be stabilised, 7 years of daily traffic data is now available. Moreover, the Bangladesh University of Engineering and Technology has conducted 13 weekly surveys on the Jamuna Bridge, where hourly traffic was also counted (Rahman, 2002).

The selected corridor should not only possess long duration traffic data, but also the quality of the data has to be high. Although traffic generation and flow are random events, research has shown that the traffic flow in a particular corridor maintains a definite flow pattern; and also the same is repetitive in nature. Figure 2 shows a typical hourly traffic flow pattern on the Jamuna Multipurpose Bridge.

From Figure 2, it can be seen that the 13 wks covered during the surveys repeat a similar nature in the curve. This undoubtedly proves the quality of the data and only this type of quality data is worth using for flow characteristics' analyses and the prediction of ADT (abstract data type) at any future date. As such, this set of high quality data has been used in this research to predict the variation of traffic flow with different time frames. The data from the 4 corridors are analysed in a combined way as all the traffic passes through the Jamuna Multipurpose Bridge.

The ANN programme used in this study was developed by the first author and is available at <http://teacher.buet.ac.bd/sid>. The programme uses a general form of a back propagation algorithm to the network for input and output data. In this paper, one ANN has been used and the data structure is described in Table 1. The hourly traffic data is available in limited form over a time of 13 wks spread over 4 years. The data is arranged in the following format (Table 1) to fit into the training module of the ANN programme.

Table 1: Data structure used for the training of ANN

Year	Input			Output	
	Month	Day	Hours	Hourly traffic	Daily total traffic
1998	10	24	6	57	2185
1998	10	24	7	67	2185
....

Using the above structured data, the training of back-propagation algorithm started. It needs quite a long time to properly train the data using the interface (Figure 1). The learning rate and momentum of the training module of the programme can be dynamically adjusted during the training session. Here, 24 hidden nodes are used heuristically. The training was stopped when the error norm reached a stable value, which did not change any more even after tweaking the parameters. Once the training is over, the programme can be used to predict the values using existing data patterns.

In order to model the traffic flow data by SVM, a non-commercial SVM software (SVMlight) is used. The SVMlight software has its own data format, which can be converted by another free software: LightDataAgentv21.exe. First, the data is formatted as given in Table 2. The hourly and daily total traffic data are separated as SVM can handle only one output. Therefore two tables were separately trained and missing data are predicted.

Table 2: Data structure used for the training of SVM

Year	Input			Output
	Month	Day	Hours	Hourly/daily traffic
1998	10	24	6	57
1998	10	24	7	67
....

RESULTS AND DISCUSSION

Once the training is over, a few predictions of traffic volume with hours, days and years are carried out to understand the predicting capability of ANN and SVM. Figures 3 and 4 show the hourly traffic volume against hours of a day. Two segments of data from the years 1999 and 2002 were selected for prediction. The ANN model

of the traffic data is accurate enough to capture the trend of data over the hours of traffic volume measurement. Figures 5 and 6 show the variation of total traffic volume for that day over the hours of the day. This should be a constant value over the hours as it is a single value (sum of all the hourly traffic volume). This has been checked to find out how the ANN captures the data pattern. It is found that it is not exactly constant in both cases, because of the influence of all the other traffic data and the errors in data due to the factors discussed earlier. Figures 7 and 8 show the variation of daily traffic volume with the day of the week. It is found to be well predicted. In Figure 8, the sudden spike in the data is automatically eliminated.

In order to determine the accuracy of both methods, paired t-tests were performed on, (i) predicted data by ANN and measured data, and (ii) predicted data by SVM and measured data. In the first case, the null hypothesis was that there is no difference between predicted data by ANN and measured data. It was found that $T_{obs} = 1.380129$ and $T_{crit} = -1.71387$. Since $T_{obs} > T_{crit}$, the null hypothesis was rejected and it was concluded with 95 % confidence that the significant difference in simulation by ANN and actual data exists. On the other hand, in the second case, the null hypothesis was that there is no difference between predicted data by SVM and measured data. It was found that $T_{obs} = -1.0$ and $T_{crit} = -1.71387$. Since $T_{obs} < T_{crit}$, the

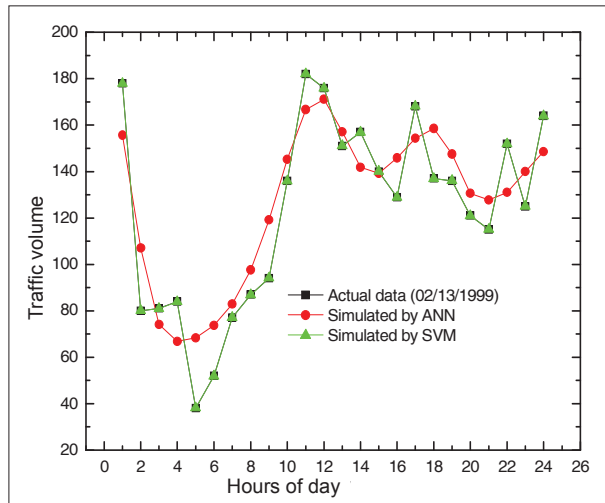


Figure 3: Plot of hourly traffic volume vs hours of a day for the date 13th February, 1999

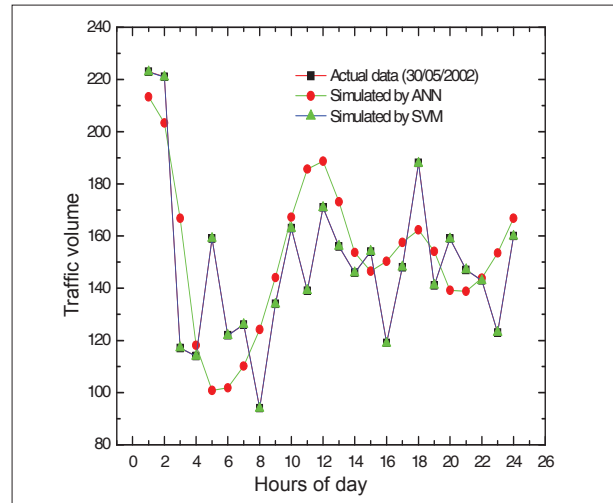


Figure 4: Plot of hourly traffic volume vs hours of a day for the date 30th May, 2002

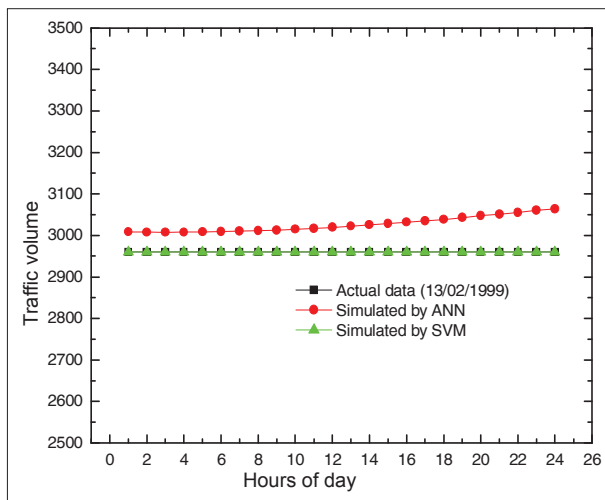


Figure 5: Plot of daily traffic volume vs hours of a day for the date 13th February, 1999

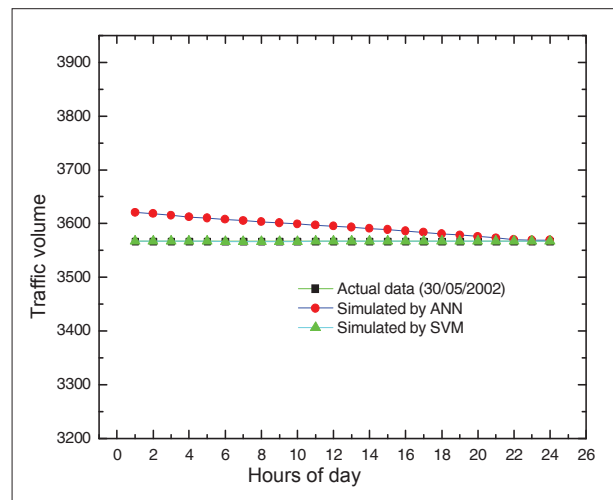


Figure 6: Plot of daily traffic volume vs hours of a day for the date 30th May, 2002

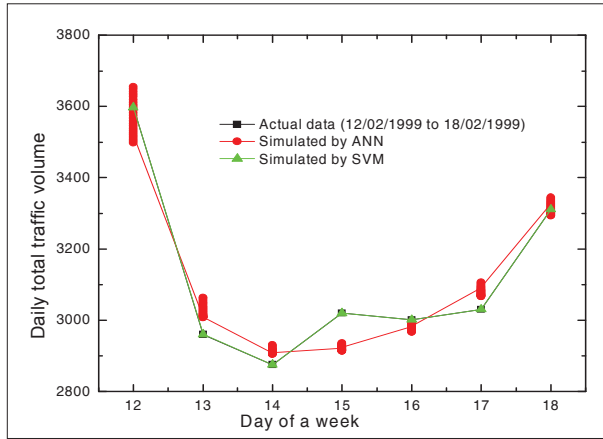


Figure 7: Plot of daily total traffic volume vs day of a week (12 to 18 of February, 1999)

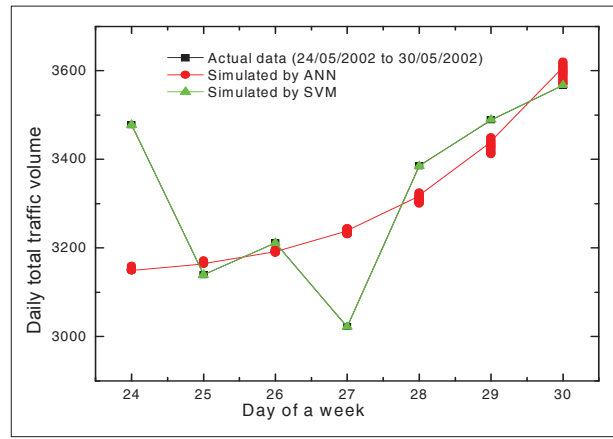


Figure 8: Plot of daily total traffic volume vs day of a week (24 to 30 of May, 2002)

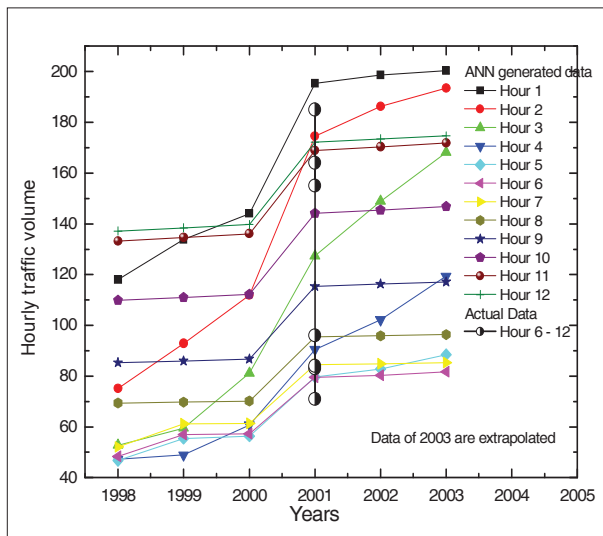


Figure 9: ANN generated plot of hourly traffic volume vs years (1999 to 2003)

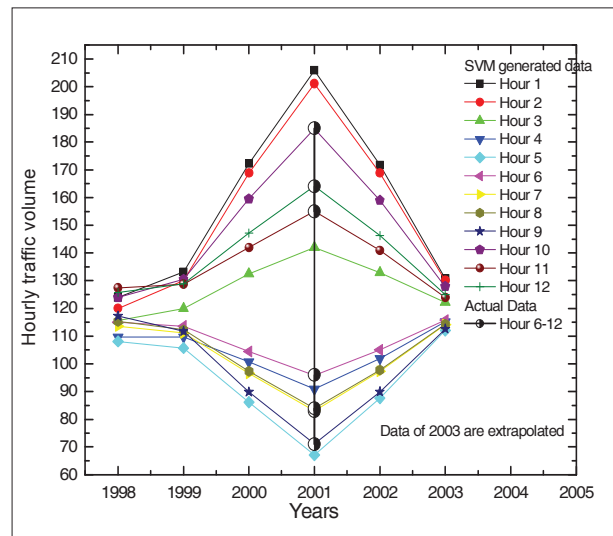


Figure 10: SVM generated plot of hourly traffic volume vs years (1999 to 2003)

null hypothesis was rejected and it was concluded with 95 % confidence that there is no significant difference in simulation by SVM and actual data. That means the SVM predicted the observed data exactly.

All the predictions of data as presented in Figures 3 through 8 by SVM is almost 100 % accurate. In these figures, actual data points are overlapped by SVM simulated points. While predictions by ANN models are smooth, it could not capture the sudden jumps in actual traffic data, but the SVM could capture all the details of the traffic flow data. It could capture both the hourly

traffic volume and daily traffic volume. Therefore, SVM technique of traffic volume forecasting appears to be more accurate.

The potential of using ANN and SVM models to capture all the possible trends inside traffic data is shown. Figures 9 and 10 show the variation of hourly traffic data with its years of origin. In this way, the individual growth factor of each hourly traffic data can be seen. It can be observed that the hourly variation also changes with the year. For the clarity of the figure, only up to 12 hours of data are plotted.

CONCLUSION

The following conclusions are deduced from this study using ANN and SVM to assimilate sporadic traffic data and predict missing data from the models:

- 1) The generated ANN and SVM models can capture the hourly traffic data patterns and predict the pattern quite reasonably.
- 2) The trained ANN and SVM models can also capture the hourly variation of daily traffic volume.
- 3) The ANN and SVM models predict the variation of daily total traffic volume over the days of a week quite successfully.
- 4) Above all, both methods can predict the hourly growth rate of the traffic volume.
- 5) SVM results are more accurate than ANN results in all respects.

REFERENCES

1. Abdi J. & Moshiri B. (2015). Application of temporal difference learning rules in short-term traffic flow prediction. *Expert Systems* **32**(1): 49 – 64. DOI: <https://doi.org/10.1111/exsy.12055>
2. Barewood J.E. (1965). *Traffic Engineering Handbook*, 3rd edition. Institute of Traffic Engineers, Washington DC., USA.
3. Basu D. (2006). Modeling stream speed in heterogeneous traffic environment using ANN-lessons learnt. *Transport* **21**: 269 – 273.
4. Cetiner B.G., Sari M. & Borat O. (2010). A neural network based traffic-flow prediction model. *Mathematical and Computational Applications* **15**: 269 – 278. DOI: <https://doi.org/10.3390/mca15020269>
5. Chiou Y.C., Lan L.W. & Tseng C.M. (2010). Estimation of dynamic freeway origin-destination matrices with cell-based arrival distribution modeling. *Journal of Eastern Asia Society for Transportation Studies* **7**: 1 – 16.
6. Chiou Y.C., Lan L.W. & Tseng C.M. (2014). A novel method to predict traffic features based on rolling self-structured traffic patterns. *Journal of Intelligent Transportation Systems* **18**: 352 – 366.
7. Dochy T., Danech-Pajouh M. & Lechevallier Y. (1996). Short term road traffic forecasting using neural networks. *RTS English Issue* **11**: 73 – 82.
8. Dougherty M. & Cobbett M. (1997). Short-term inter-urban traffic forecasts using neural networks. *International Journal of Forecasting* **13**(1): 21 – 31.
9. Dougherty M.S., Kirby H.R. & Boyle R.D. (1994). Using neural networks to recognise, predict and model traffic. *Artificial Intelligence Applications to Traffic Engineering* (eds. M. Bielli, G. Ambrosino & M. Boero), pp. 233 – 250. VSP, Utrecht, The Netherlands.
10. Federal Highway Administration (2001). *Traffic Monitoring Guidelines*. US Department of Transportation, Federal Highway Administration, New Jersey, Washington DC, USA.
11. Garber N.J. & Hoel L.A. (1990). *Traffic and Highway Engineering*, revised 2nd edition, PWS Publishing, USA.
12. Joachims T. (1998a). Making large-scale SVM learning practical, *Technical Report, SFB 47*. Universität Dortmund, Germany.
13. Joachims T. (1998b). Text categorization with support vector machines: learning with many relevant features. *Proceedings of the 10th European Conference on Machine Learning*, 21 – 23 April, Chemnitz, Germany, pp. 137 – 142. DOI: <https://doi.org/10.1007/BF0026683>
14. Joachims T. (1999a). Making large-scale support vector machine learning practical. *Advances in Kernel Methods* (eds. B. Schölkopf, C. Burges & A. Smola). MIT Press, Cambridge, USA.
15. Joachims T. (1999b). Transductive inference for text classification using support vector machines. *Proceedings of the Sixteenth International Conference on Machine Learning*, 27 – 30 June, pp. 200 – 209.
16. Joachims T. (2000). Estimating the generalization performance of a svm efficiently. *Proceedings of the Seventeenth International Conference on Machine Learning*, 29 June – 02 July, pp. 431 – 438.
17. Joachims T. (2002a). *Learning to Classify Text Using Support Vector Machines*. Springer, USA. DOI: <https://doi.org/10.1007/978-1-4615-0907-3>
18. Joachims T. (2002b). Optimizing search engines using clickthrough data, *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 23 – 26 July, Alberta, Canada, pp. 133 – 142. DOI: <https://doi.org/10.1145/775047.775067>
19. Kaub A.R. (1998). Traffic safety prediction model, available at <http://www.freepatentsonline.com/5798949.html>, Accessed 15 March 2016.
20. Kirby H., Dougherty M. & Watson S. (1997). Should we use neural networks or statistical models for short term motorway traffic forecasting? *International Journal of Forecasting* **13**: 43 – 50.
21. Kumara K., Paridab M. & Katiyarc V.K. (2013). Short term traffic flow prediction for a non urban highway using artificial neural network. In: *Procedia - social and behavioural sciences. Proceedings of the 2nd Conference of Transportation Research Group of India (2nd CTRG)*, volume 104, 2 December, pp. 755 – 764. DOI: <https://doi.org/10.1016/j.sbspro.2013.11.170>
22. Ledoux C. (1997). An urban traffic flow model integrating neural network. *Transportation Research* **5C**: 287 – 300.
23. Matson T.M., Smith W.S. & Hurd F.W. (1955). *Traffic Engineering*. McGraw-Hill Book Company, New York, USA.
24. Pignataro L.J. (1973). *Traffic Engineering, Theory and Practice*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, USA.
25. Rahman S.K. (2002). *Study of Vehicular Flow Pattern on Jamuna Multipurpose Bridge Access Road*. Department

- of Civil Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh.
26. Roads and Highways Department, Government of Bangladesh (1994). Development of Geometric Design Standards. *Road Material and Standards Study Bangladesh*, volume VIIA. Roads and Highways Department, Government of Bangladesh.
 27. Roads and Highways Department, Government of Bangladesh (2005). *Pavement Design Guide*. Roads and Highways Department, Government of Bangladesh.
 28. Robichaud K. & Gordon M. (2002). An assessment of data collection techniques for highway agencies. *Transportation Research Record: Journal of the Transportation Research Board* **1855**: 129 – 135.
DOI: <https://doi.org/10.3141/1855-16>
 29. Sharma V.D., Sharma R.R., Pandey B.D. & Gupta N. (1989). Nonlinear analysis of a traffic flow. *Journal of Applied Mathematics and Physics* **40**(6): 828 – 837.
DOI: <https://doi.org/10.1007/BF00945805>
 30. Sun Z., Pan J. & Duan Q. (2008). Study on a new traffic flow forecasting method, *Proceedings of the 2008 Fourth International Conference on Natural Computation*, volume 3, 18 – 20 October, Jinan, China, pp. 349 – 353.
DOI: <https://doi.org/10.1109/ICNC.2008.883>
 31. Vapnik V.N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, USA.
DOI: <https://doi.org/10.1007/978-1-4757-2440-0>
 32. Wang X., Jia L., Zhao J. & Chen Y. (2006). Urban traffic flow forecasting model of double RBF neural network based on PSO. *Sixth International Conference on Intelligent Systems Design and Applications*, 16 – 18 October, Jinan, China, pp. 892 – 896.
DOI : <http://doi.ieeecomputersociety.org/10.1109/ISDA.2006.277>
 33. Xiao-bin L. (2010). *Forecasting Urban Traffic Flow Based on Support Vector Machine Optimized by Genetic Algorithm*. Institute of Automation Science and Engineering, South China University of Technology, Guangzhou, China.
 34. Xu J., Lu Y. & Ho D.W.C. (2007). A combined genetic algorithm and orthogonal transformation for designing feedforward neural networks, *Proceedings of the 3rd International Conference on Natural Computation (ICNC 2007)*, 24 – 27 August, Haikou, Hainan, China, pp. 10 – 14.
DOI: <https://doi.org/10.1109/ICNC.2007.13>
 35. Yang Z., Wang Y. & Guan Q. (2006). Short-term traffic flow prediction method based on SVM. *Journal of Jilin University Engineering and Technology Edition* **36**: 881 – 884.
 36. Yeh I.C. (2004). *Applied Neural Network*, 3rd edition. Ru-Lin Co., Taipei, Taiwan.
 37. Yin H., Wong S.C., Xu J. & Wong C.K. (2002). Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research (part C)* **10**: 85 – 98.