

11th Transportation Planning and Implementation Methodologies for Developing Countries,  
TPMDC 2014, 10-12 December 2014, Mumbai, India

## Application of data mining techniques for traffic density estimation and prediction

Jithin Raj<sup>a</sup>, Hareesh Bahuleyan<sup>a</sup>, Lelitha Devi Vanajakshi<sup>a\*</sup>

*Department of Civil Engineering, Indian Institute of Technology Madras, Chennai-600036, India*

---

### Abstract

Advanced Traveller Information Systems (ATIS) is one of the functional areas of Intelligent Transportation Systems (ITS) and it aims at providing real time traffic information to the travellers for making better travel decisions. This information would be most effective if provided to travellers during or before the start of their trip. Therefore, accurate prediction models are required in ATIS for conveying reliable information about the future state of traffic. Different methods used for the prediction of traffic parameters include historic averaging, regression analysis, Kalman filtering, time series analysis, machine learning, etc. The objective of this research is to explore the use of automated sensor data and data driven techniques for traffic state prediction under Indian traffic conditions. Travel time and traffic density (as an indicator of congestion) are used commonly to inform users about the state of a traffic system. However, these two parameters are spatial in nature and direct measurement from field is difficult. Therefore, estimation of these parameters from location based data is a challenge in many of the ITS implementations. The present study addresses the problem of estimation of traffic density with the help of location based sensors which are capable of measuring parameters such as volume and Time Mean Speed (TMS). Machine learning techniques namely, k-Nearest Neighbour (k-NN) and Artificial Neural Network (ANN) are selected as the estimation and prediction tool in this study, based on acceptable performance of the same in earlier studies.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Department of Civil Engineering, Indian Institute of Technology Bombay

**Keywords:** ATIS; Traffic Density; Estimation; Prediction; k-NN; ANN

---

---

\* Corresponding author. Tel.: +91-4422574291; fax: +0-000-000-0000 .  
E-mail address: [lelitha@iitm.ac.in](mailto:lelitha@iitm.ac.in)

## 1. Introduction

Traffic congestion is a major challenge in the area of transportation planning as well as traffic management. Congestion usually relates to an excess of vehicles on a portion of roadway at a particular time resulting in speeds that are slower – sometimes much slower – than normal or free flow speeds (FHWA, 2005). Even though congestion can be solved to some extent, the problem cannot be solved completely. Therefore, informing road users about congestion conditions of different roadways will be an advantage to them while performing their journey for making suitable decisions. This can be done only by quantifying congestion using some quantifiable parameters and making it available for the road users as part of Advanced Traveller Information Systems (ATIS). Travel time and traffic density are the commonly used traffic measures used for quantifying congestion on roadways. Being spatial in nature, measurement of these parameters directly from field is very difficult. Travel time can be defined as the time required for road users to travel from one point of roadway to other. Field measurement of travel time is done using spatial based sensors such as GPS or vehicle identification devices such as Bluetooth sensors. Traffic density is defined as the number of vehicles occupying a given length of roadway. Traffic density is considered as the primary measure for quantifying congestion of roadways other than signalised intersections (HCM, 2000). However, aerial photography is the direct method for field measurement of traffic density, which is very difficult to implement in field. Since density is difficult to measure, indirect methods of estimating density from other parameters such as flow, speed or occupancy is usually adopted. Hence, this problem of estimation of density from location based parameters is of importance and is attempted in the present study under Indian conditions. Also, the travel decisions made by road users are more affected by the traffic conditions during the trip than the conditions at the start of the trip. Therefore, accurate prediction models are also required in ATIS for giving reliable information about the future state of traffic.

In this study, data mining techniques were implemented for the estimation and prediction of the traffic state variables. Data mining processes use computational tools to extract useful knowledge from large datasets (Silwattananusarn&Tuamusk, 2012). Machine learning techniques such as k-Nearest Neighbour (k-NN) and Artificial Neural Network (ANN) were selected in this study as the tools for data mining, based on acceptable performance of the same in earlier studies.

The k-NN algorithm is one of the simplest data mining techniques, widely used for classification and regression. It is a non-parametric and supervised algorithm, which classifies a new record by comparing it to similar records in the training data set. The most common method of defining 'similarity' is based on the Euclidean distance between the records. Since the algorithm would be used to estimate density based on speed and flow values, the training dataset would have speed and flow as inputs and density as target. The algorithm will identify the closest/most significant input values. When a new data is provided, the algorithm looks for the 'k' nearest neighbours in terms of Euclidean distance which is calculated based on the input.

ANN is a popular machine learning tool inspired by biological nervous systems. Neural network can be trained to perform a particular function by adjusting the weights of the connections between units. Each of these processing units are known as neurons. Neural networks are trained to adjust the weights of these neurons so that a particular input leads to a specific target. In each neuron, a scalar input is transmitted through a connection that multiplies it by the scalar weight and then added by a bias value, and then applies a transfer function to form an output. This weights and bias values can be adjusted to get some desired output. The output of this neuron will be the input of some other neurons based on connectivity and will be repeated for every data points for a series of iterations. A network is thus developed with the corresponding inputs and targets which can give similar outputs when new data is provided.

For the prediction, the same tools are used with the input being the time series of estimated density values to find out the density values for the future time intervals. The required data for the present study were collected using a location based sensor namely, TIRTL, which uses infra-red technology for the detection of vehicles. The sensor was installed near Perungudi toll plaza in Rajiv Gandhi Salai, Chennai facing the northbound traffic. This road is a representative Indian road with heterogeneous and lane less traffic conditions, i.e. different class of vehicles like 2 wheelers, 3 wheelers, passenger cars, bus and trucks etc. travelling together with zero lane discipline. Volume, speed

and occupancy at this location were collected for four days of two different weeks each, and density was calculated from the occupancy by using the occupancy-density relation. One week of data was used for training the machine learning algorithms and the other week's data were used for validation.

## 2. Literature Review

The first part of the literature review covers some of the popular studies in the area of estimation and prediction of traffic parameters. Studies on estimation and prediction techniques showed the use of various techniques such as historic average technique, statistical methods, machine learning techniques and model-based techniques as the most commonly used ones.

Estimation of density from flow data obtained by aerial photography was reported in some of the earlier studies (Nahian and Trivedi, 1973; Chang and Ghazis, 1975). Flow and occupancy data collected using loop detectors were used for the estimation of density once the loop detectors became popular (Kurkjian et al., 1979; Coifman, 2003). Wang, Papageorgiou and Messmer (2007) estimated traffic flow variables such as flow, space mean speed and density using online parameter model estimator, which also discussed about the estimator's tracking ability and sensitivity to various initial conditions. Studies in the area of speed estimation also showed the usage of loop detector data for the estimation. Hazelton (2004) estimated traffic speed from volume and occupancy data obtained from loop detectors using Markov chain Monte Carlo methods. Wang and Nihan (2000) in another study estimated freeway traffic speed from single loop detector outputs such as volume and occupancy by introducing a relation between these parameters. Another method has been proposed by Soriguera and Robuste (2011) to obtain space mean speeds directly from loop detector data with a mean relative error of approximately 0.5%. Thus, it can be seen that most of the studies in the area of estimation used loop detector data as the main input.

Studies reporting traffic parameter prediction using data driven approaches are discussed here. Zhang and Rice (2003) used a varying coefficients linear model, which varies as a smooth function of departure time, with past instantaneous travel time to predict future travel time. Principal component analysis and nearest neighbour approach were investigated by Rice and Zwet (2004) by combining the historical data and the instantaneous travel time data. Other methods such as linear regression and Kalman filtering are widely used in the area of prediction. Kwon et al. (2000) estimated travel time on a freeway using flow and occupancy data obtained from loop detectors and predicted to future time steps using linear regression. Yu et al. (2008) proposed a variation based online travel time prediction approach using clustered neural networks. Padiath et al. (2009) compared the performance of a historic technique, an ANN based technique and a model based technique for the prediction of traffic density under Indian traffic conditions. However, the study used limited data which were collected manually. Kalman Filter method was employed by Chen and Chein (2001) to predict travel time using data obtained from probe vehicles. Chein and Kuchipudi (2003) also used Kalman Filter method for the prediction of travel time. Overall it can be seen that there are only very few application of data mining techniques in both estimation and prediction problems. Fabrizi and Ragona (2014) developed a pattern matching method of prediction, which tries to identify patterns in the past data that describes the actual traffic load, and produces forecasts supposing that the trend will repeat. Valenti et al. (2010) used and compared five predictive models including ANN and k-NN for the prediction of incident duration. Various studies on estimation and prediction of traffic parameters using ANN shows that the method is suitable if large samples of data are available. Some of the related literature are discussed below. Cetiner et al. (2010) reported prediction of traffic flow from the historical data using ANN and showed a correlation coefficient in the range of 0.85 to 0.95. Ozkurt and Camci (2009) presented an automatic traffic density estimation and vehicle classification method using ANN and 94% accuracy was reported. Vanajakshi and Rilett (2004) predicted traffic speed ranging from 2 minutes ahead up to an hour in to future using artificial neural networks and support vector machines and reported that performance of ANN largely depended on the amount of data. Missing values in traffic data collected by radar detection systems and loops in three types of rural roads including arterial, highway, and freeway roads in Iran have been estimated using ANN by Mahmoudabadi and Fakharian, (2010). Results of another study from Iran showed that ANN can be used to estimate the average speed of vehicles (Mahmoudabadi, 2010). Jiang and Wah (2003) proposed a new approach of constructing and training neural networks for pattern classification. Drakopoulos and Abdulkader (2005) studied the neural network training of heterogeneous data and proposed data pruning (removal of noisy data) and ordered training (partitioning of data) as the effective methods to deal with

heterogeneous data. Zheng and Zuylen (2013) used a three layer artificial neural network for the estimation of urban link travel time from vehicle positions, link ids, timestamps and speed using the probe vehicle data. Thus, it can be seen that most of the studies in the area of estimation used loop detector data as the main input.

k-NN has also been used extensively for similar applications. Embreda and Dalagan (2012) explored the use of k-NN Algorithm in determining the traffic weight of a given snapshot and they reported that accuracy was highly dependent on the training data. Dubey and Pathak (2013) studied the application of data mining technique to automatically extract computer users' normal behaviour from training network traffic data. Cruz et al. (2012) studied the problem of processing k-NN queries in road networks considering traffic conditions and the query return the k points of interest that can be reached in the minimum amount of time. Zhang et al. (2013) presented a method for short term urban expressway flow prediction system with accuracy over 90% using k-NN. Lin et al. (2013) applied k-NN method to form the training dataset for local linear wavelet neural network instead of taking the whole historical dataset for training for the short term prediction of five minutes volume. Xiaoyu et al. (2013) proposed a two-tier k-nearest neighbour algorithm combined with the actual traffic flow to improve the calculation speed and the accuracy of the algorithm. Zheng and Su (2014) developed a kNN-LSPC (k Nearest Neighbour - Linearly Sewing Principle Component) for the prediction of traffic volume, which outperformed eight other algorithms.

The above review shows the use of various model based and data driven approaches for the estimation and prediction of traffic parameters. However, under Indian conditions, automated sensors were not available and hence the prediction problems using data driven approaches such as ANN or k-NN were not attempted exhaustively. Based on the literature, and based on the availability of a good automated database, two data driven approaches namely Artificial Neural Networks (ANN) and K Nearest Neighbour (k-NN) approach were selected for the estimation of traffic density in the present study.

### 3. Test bed and data collection

The required data for the present study were collected using a location based sensor namely, TIRTL, which uses infra-red technology for the detection of vehicles. The sensor was installed near Perungudi toll plaza in Rajiv Gandhi Salai, Chennai facing the northbound traffic. This road is a representative Indian road with heterogeneous and lane less traffic conditions, i.e. different class of vehicles like 2 wheelers, 3 wheelers, passenger cars, bus and trucks etc. Fig. 1 below shows the location of the sensor.



Fig. 1. Location of data collection point (Source: a. Google maps, b. Photograph).

Traffic density is the most commonly used parameter to indicate the level of congestion on a roadway. However, the measurement of density directly from field is very difficult as it is a spatial parameter. Therefore, there is a need to estimate traffic density from available location specific parameters such as volume and speed. The present study uses kNN and ANN for the estimation of density from speed and volume obtained from automated sensors. For validating the results, occupancy given by the infrared sensor was used. Occupancy is defined as the proportion of time a roadway section is occupied by vehicle. It is considered as a surrogate measure of density and is related to

density as shown in equation 1(May, 1990).

$$k = \frac{\%occ}{L_v + L_d} \quad (1)$$

where,  $\%occ$  is the percentage of time detector being occupied by vehicles,  $L_v$  is the vehicle length and  $L_d$  is the detector length. This parameter is provided by the infrared sensor accurately and hence is used to calculate density and is used for validation.

## 4. Methodology

### 4.1. K-Nearest Neighbour Algorithm (k-NN)

The k-Nearest Neighbour (k-NN) is one of the simplest machine learning algorithms, most widely used for classification. It is a non-parametric and supervised algorithm, which classifies a new unclassified record by comparing it to similar records in the training data set. The most common method of defining 'similarity' is based on the Euclidean distance between the records in the feature space. In this paper, the k-NN regression method is utilized, where the output is not a predefined class, but a continuous value. This is explained with the help of the Fig. 2 below:

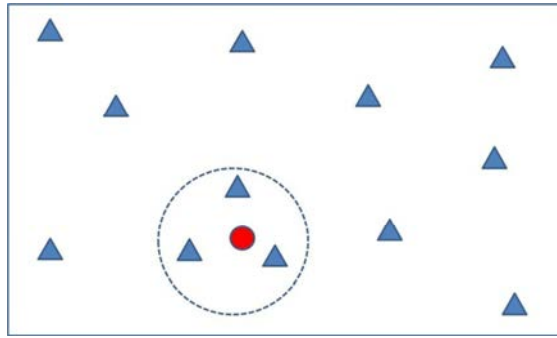


Fig. 2. Demonstration of k-NN algorithm

In the density estimation process carried out in this study, the flow and speed are inputs to the algorithm to get the corresponding density as the output. The blue triangles are the trained data points positioned in the feature space based on their input values. When a new input is provided (red circle), based on its speed and flow values, it occupies a position in the feature space. Next, the algorithm looks for the 'k' nearest neighbours, in terms of Euclidean distance, which is calculated based on the input variables. For example, if  $k=3$ , the output density would be decided by the densities of the three blue triangles shown within the dashed circle as these are the three closest neighbours. Let  $s_1$ ,  $s_2$  and  $s_3$  be the speeds;  $v_1$ ,  $v_2$  and  $v_3$  the flows; and  $d_1$ ,  $d_2$  and  $d_3$  the density of the data points indicated by these triangles. For the test data point (red circle), let the speed be  $s$  and flow be  $v$  which are provided as input and we are required to estimate the corresponding density  $d$ . Then the Euclidean distances are given by equations (2),

$$\begin{aligned} P1 &= \sqrt{(s_1 - s)^2 + (v_1 - v)^2}, \\ P2 &= \sqrt{(s_2 - s)^2 + (v_2 - v)^2}, \\ P3 &= \sqrt{(s_3 - s)^2 + (v_3 - v)^2}. \end{aligned} \quad (2)$$

Since triangles 1, 2 and 3 are the closest neighbours, any distance  $P_i$ ,  $i \neq 1, 2$  and  $3$ ; will only be greater than  $P_1$ ,  $P_2$  and  $P_3$ . And the output density of the new record would be calculated as a simple arithmetic average as given in equation (3).

$$d = \frac{d_1 + d_2 + d_3}{3} \quad (3)$$

In the second application of k-NN algorithm discussed in this paper, the density of the current time step is to be predicted based on the densities of the previous 5 time steps which are provided as inputs to the algorithm.

#### 4.2. Artificial Neural Network(ANN)

ANN is a popular machine learning tool inspired by biological nervous systems and is composed of units operating in parallel. Neural network can be trained to perform a particular function by adjusting the weights of the connections between units. Each of these processing units is known as neurons. Neural networks are trained to adjust the weights of these neurons so that a particular input leads to a specific target. In each neuron, scalar input  $p$  is transmitted through a connection that multiplies it by the scalar weight  $w$  and then added by a bias value  $b$ , to form the result  $wp+b$ . This sum is the input for the transfer function  $f$  which gives an output  $f(wp+b)$ . The variables  $w$  and  $b$  can be adjusted to get some desired output is the basic idea. The output of this neuron will be the input of some other neurons based on connectivity. The transfer function  $f$  can be hardlim, sigmoid, purelin etc. based on the requirement. After passing through all the connected neurons, it should be able to produce the desired target. In this study, back-propagation algorithm was used for training, which automatically recalculates and makes the error minimum. The training continues till one of the stopping criteria is reached, which may be number of iterations, maximum number of validation checks, or performance in terms of mean squared error.

#### 4.3. Combining k-NN with ANN

Along with testing ANN and k-NN individually, a model which is a fusion of the two techniques was also developed. The fusion methodology adopted is as follows: The volume and speed are provided as inputs to the kNN algorithm. The kNN is required to determine the first  $k$  (taken as 100 in this study) nearest neighbours of the new input record, which in turn would form the training dataset for the ANN. Once the ANN is trained, it can now predict a value of density for the original record. The description of the fusion model is described in the Fig. 3.

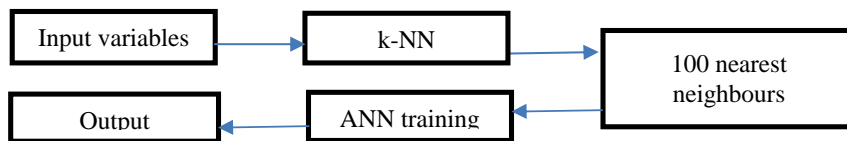


Fig. 3. Model for combining ANN with KNN

### 5. Implementation and results

Traffic density was estimated and then predicted to future using the models already discussed. The comparison of actual and estimated or predicted density are explained separately in detail. The error in estimation and prediction models were quantified using the statistical measure, Mean Absolute Percent Error (MAPE), given by the equation (4).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{y_i} \right| \times 100 \quad (4)$$

where  $y_i$  is the actual density and  $x_i$  is the estimated or predicted density.

#### 5.1. Estimation of traffic density

The available data set were divided in to two subsets- training data set and testing data set. Four days data were

taken as the training data set, and testing data set was selected as the same days of another week. The training data is used to identify the pattern of data and the test data set is used for checking the performance. The training set had speed and flow data at every 5 minute interval as input and the corresponding actual density, obtained from the occupancy values, as the target variable (Fig. 4). Fig. 5 is a graph that shows the comparison between the actual density and the density estimated by each of the techniques described in section 4.

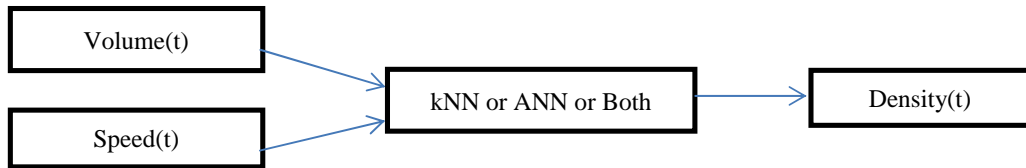


Fig. 4. Model for estimation of density

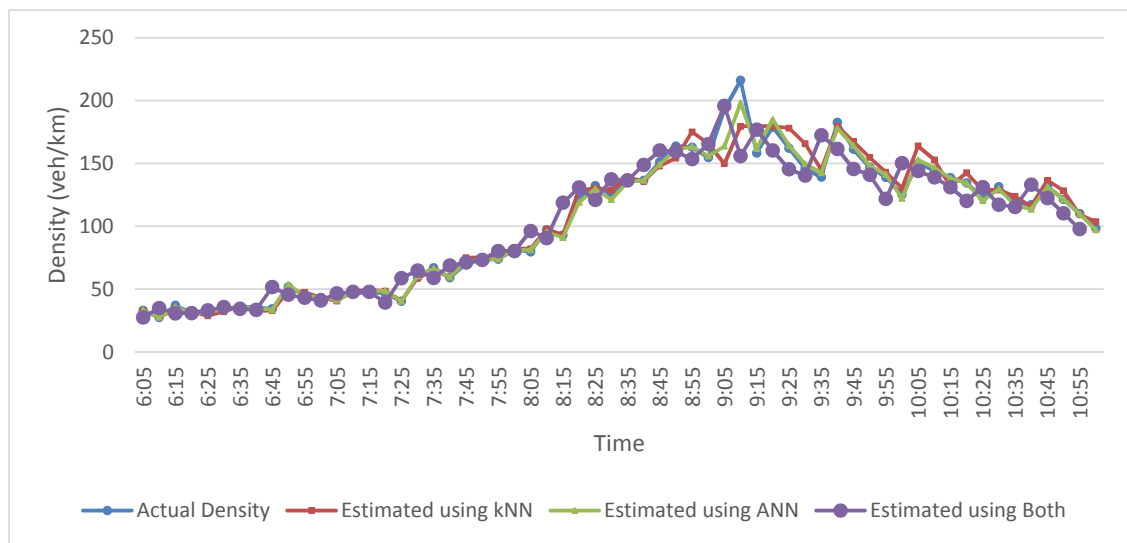


Fig. 5. Sample comparison of actual and estimated density

From Fig. 5, it can be observed that the estimated value of density is in line with the actual density. The errors were quantified using Mean Absolute Percentage Error (MAPE) and the values obtained are tabulated in Table 1. It can be seen that the performance of the ANN algorithm is comparatively better. The results of the combined ANN-kNN model do not show any improvement in performance. This may be indicating that giving enough training data set is more important for ANNs performance than giving reduced set of best inputs. ANN process is able to generate the best network, if enough training sets are provided.

Table 1. MAPE (%) obtained for different days for density estimation.

Day Tested	kNN	ANN	kNN-ANN (100 neighbours)
Day1	3.46	1.94	2.49
Day2	4.41	1.78	2.61
Day3	3.74	1.54	1.96
Day4	3.20	1.50	2.01



### 5.2. Prediction of traffic density

Road users, in general, will be more interested in finding information about future traffic conditions, in order to know about what they can expect while making the trip, than the present scenario. Therefore, it is always of interest to predict the estimated values to the future time steps and provide that information to the users. Hence, in this study, the estimated density values are predicted to future time intervals. This can be carried out by identifying the evolution pattern of parameters and projecting it to future. Thus, the previous several density values can be used as input to predict to the next time interval (s). In this study, previous 5 time intervals density are taken as input (which is equivalent to previous 25 minutes) to predict the density in future time intervals. Figure 6 shows the inputs and outputs used for density prediction. The scheme was implemented using MATLAB and a sample result obtained is shown in Fig. 7.

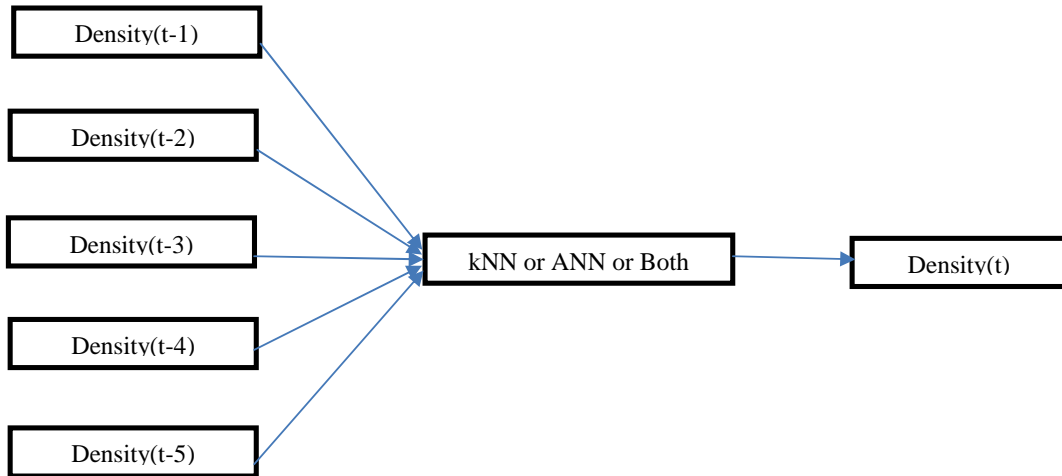


Fig. 6. Model for prediction of density

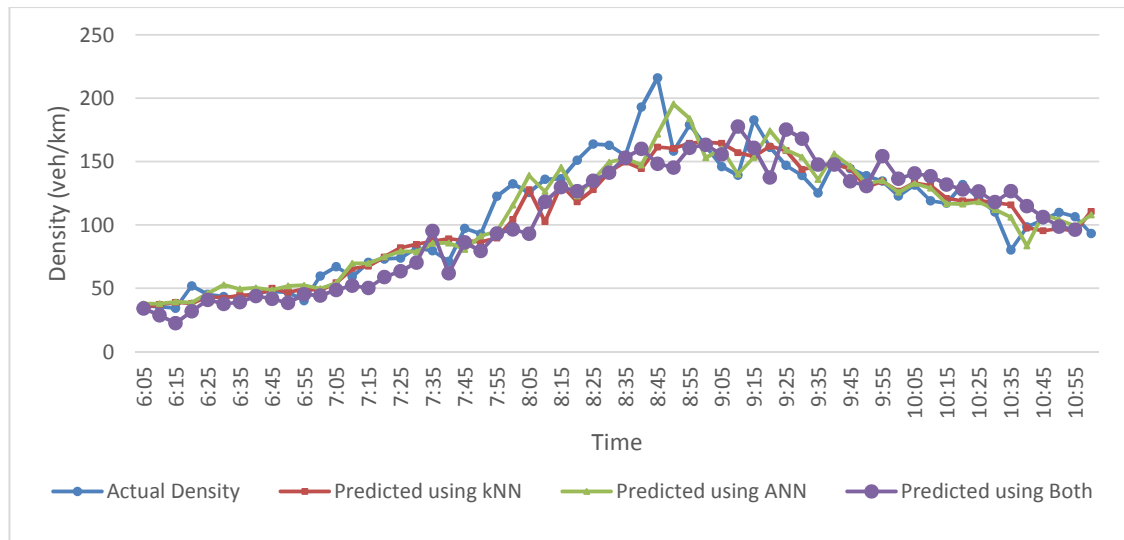


Fig. 7. Sample comparison of actual and predicted density



It can be seen that all the methods performed equally. The performance was quantified using MAPE and the values obtained are shown in Table 2. It can be seen that the individual methods are performing better than the combined approach. Here also, the reduction in training set in the combined approach may be the reason for the increase in prediction error.

Table 2. MAPE (%) obtained for different days for density prediction

Day Tested	k-NN	ANN	kNN-ANN (100 neighbours)
Day 1	11.69	10.01	14.86
Day 2	12.39	11.66	15.93
Day 3	10.88	11.41	12.42
Day 4	10.02	11.40	14.37

## 6. Summary and conclusions

Traffic density is the primary measure for quantifying congestion in uninterrupted roadway sections and is defined as the number of vehicles occupying a given length of roadway. For measuring density directly from the field, number of vehicles occupying the roadway section has to be measured at a particular instant of time. Aerial photography is a technique for density measurement but it is very difficult to implement. Hence, it is usually estimated from other location based parameters such as speed, flow or occupancy and is the topic of the present study. The data available from automated sensors were utilized for testing the models that were developed. The data generated by most of the sensors are mainly classified volume and speed. Both these parameters are location based and hence may not be of much interest to the users. In this study, these data are used for the estimation and prediction of density, which in turn can be used to inform users about congestion.

The techniques used for the estimation and prediction are based on two machine learning techniques: artificial neural networks (ANN) and k-nearest neighbour (kNN). A model which uses the output from kNN as a training set for ANN was also tested. In density estimation, where speed and volume were the inputs to estimate the target variable density produced MAPE in the range of about 2-5% using 5 minute interval data. In the density prediction, where the densities of the previous time steps were used to predict the current density, the MAPE was in the range of 10-12%. Thus, the use of data driven techniques like ANN and kNN along with automated sensor data are promising for traffic state estimation problems for ITS applications in Indian traffic conditions. However, combining these two techniques did not show any significant improvement in performance and hence is not recommended. This may be mainly due to reduction in training set in the combined approach, indicating that for better performance of ANN more training data is important than providing significant input pairs alone. However, the performance is comparable and in the case of large scale problems, the training time may be saved by using the reduced training set.

## Acknowledgements

The data collection effort in this project was made possible through sub-project CIE/10-11/169/IITM/LELI under the Centre of Excellence in Urban Transport project funded by the Ministry of Urban Development, Government of India, through letter No. N-11025/30/2008-UCD.

## References

- Cetiner, B.G., Sari, M., Borat, O., 2010. A Neural Network Based Traffic-Flow Prediction Model. *Mathematical and Computational Applications*, 15(2), 269-278.
- Chang, M., Ghazis, D.C., 1975. Traffic Density Estimation with Consideration of Lane-Changing. *Transportation Science*, 9, 308-320.
- Chen, M.,Chien, S., 2001. Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link based vs Path based. *Transportation Research Board 80th Annual meeting*, Washington, D.C.

- Chien, S.I., Kuchipudi, C.M., 2003. Dynamic Travel Time Prediction with Real time and Historic data. Approved by Journal of Transportation Engineering.
- Coifman, B., 2003. Estimating Density and Lane Inflow on a Freeway Segment. *Transportation Research: Part A*, 38(8), 689-701.
- Cruz, L.A., Nascimento, M.A., Macedo, J.A.F., 2012. k-NearestNeighbors Queries in Time-Dependent Road Networks. *Journal of Information and Data Management*, 3(3), 211–226.
- Darkopoulos, J.A., Abdulkader, A., 2005. Training Neural Networks with Heterogeneous Data. *Neural Network*, 18, 595-601.
- Dubey, R., Pathak, P.N., 2013. KNN based Classifier Systems for Intrusion Detection. *International Journal of Advanced Computer Techniques (IJACT)*, 2(4), 44-48.
- Emberda, G., Dalagan, N., 2012. Recognition of Traffic Weight Using Sobel Edge Detection Method and K-Nearest Neighbor Algorithm, *UIC Research Journal* 18(1), 21-30.
- Fabrizi, V., Ragona, R., 2014. A pattern matching approach to speed forecasting of traffic networks. *European Transport Research Review*. DOI 10.1007/s12544-014-0133-5
- Federal Highway Administration., 2005. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation. Prepared by Camebridge Systematics Inc. with Texas Transportation Institute
- Hazelton, M.L., 2004. Estimating Vehicle Speed from Traffic Count and Occupancy Data. *Journal of Data Science*, 2, 231-244.
- Highway Capacity Manual., 2000. Transportation Research Board. National Research Council, Washington, D.C.
- Jiang, X., Wah, A.H.K.S., 2003. Constructing and Training Feed-forward Neural Networks for pattern classification, *Pattern Recognition*, 36, 853-867.
- Kurkjian, A., Houpt, P.K., Gershwin, S.B., Willsky, A.S., 1979. Estimation of Roadway Traffic Density on Freeways using Presence Detector Data. *Transportation Science*, 14, 232-261.
- Kwon, J., Coifman, B., Bickel, P., 2000. Day -to-day travel time trends and travel time prediction from loop detector data. *Transportation Research Record*, 1717, 120-129.
- Lin, L., Li, Y., Sadek, A., 2013. A k Nearest Neighbor based Local Linear Wavelet Neural Network Model for On-line Short-term Traffic Volume Prediction. *Procedia Social and Behavioural Sciences*, 96, 2066-2077.
- Mahamoudabadi, A., 2010. Using Artificial Neural Network to Estimate Average Speed of Vehicles in Rural Roads. *International Conference on Intelligent Network and Computing*.
- Mahamoudabadi, A., Fakharian, S., 2010. Estimating of Missing Traffic Data Using Artificial Neural Network. *International Conference on Intelligent Network and Computing*.
- May, A.D., 1990. Fundamentals of Traffic Flow. Prentice - Hall, Inc. Englewood Cliff New Jersey 07632, second edition.
- Nahi, N., Trivedi, A., 1973. Recursive Estimation of Traffic Variables: Section Density and Average Speed. *Transportation Science*, 7, 269-286.
- Ozkurt, C., Camci, F., 2009. Automatic Traffic Density Estimation and Vehicle Classification for Traffic Surveillance Systems using Neural Networks. *Mathematical and Computational Applications* 14(3), 187-196.
- Padiath, A., Vanajakshi, L., Subramanian, S., Manda, H. 2009. Prediction of Traffic Density for Congestion Analysis under Indian Traffic Conditions. 12th International IEEE Conference on Intelligent Transportation System,
- Rice, J., Zwet.E., 2004. A Simple and Effective Method for Predicting Travel Times on Freeways. *IEEE Transactions on Intelligent Transport Systems*, 5(3).
- Silwattanasarn, T., Tuamusk, K., 2012. Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, 2.
- Soriguera, F., Robuste, F., 2011. Estimation of Traffic Stream Space Mean Speed from Time Aggregations of Double Loop Detector Data. *Transportation Research: Part C*, 19, 15-129.
- Valenti, G., Lelli M., Cucina, D., 2010. A Comparative Study of Models for the Incident Duration Prediction. *European Transport Research Review*, 2, 103-111
- Vanajakshi, L., Rilett, L.R., 2004. A comparison of the performance of artificial neural networks and support vector machines for the prediction of vehicle speed. *IEEE Intelligent Vehicles Symposium*, Parma, Italy, 194-199
- Wang, Y., Nihan, N.L., 2000. Freeway Traffic Speed Estimation using Single Loop Outputs. *International Transportation Research Record: Journal of Transportation Research*, 1727, 120-126.
- Wang, Y., Papageorgiou, M., Messmer, A., 2007. Real-Time freeway Traffic State Estimation Based on Extended Kalman Filter: A Case Study. *Transportation Science*, 41(2), 167-181.
- Xiaoyu, H., Yisheng, W., Siyu, H., 2013. Short-Term Traffic Flow Forecasting Based On Two-Tier K-Nearest Neighbor Algorithm. *Procedia Social and Behavioural Sciences*, 96, 2529-2536.
- Yu, J., Chang, G., Ho, H.W., Liu, Y., 2008. Variation Based Online Travel Time Prediction Using Clustered Neural Networks.
- Zhang, L., Liu, Q., Yang, W., Wei, N., Dong, D., 2013. An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. *Procedia Social and Behavioural Sciences*, 96, 653-662.
- Zhang, X., Rice, J.A., 2003. Short term Travel Time Prediction. *Transportation Research Part C: Emerging Technologies*, 11(3-4), 187-210.
- Zheng, F., Zuylen, H.V., 2013. Urban Link Travel Time Estimation based on Sparse Probe Vehicle Data. *Transportation Research: Part C*, 31, 145-157.
- Zheng, Z., Su, D., 2014. Short Term Traffic Volume Forecasting: A k- Nearest Neighbor Approach Enhanced by Constrained Linearly Sewing Principle Component Algorithm. *Transportation Research: Part C*. Accepted 13 February 2014.