Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 187 (2017) 582 – 587

Procedia Engineering

www.elsevier.com/locate/procedia

10th International Scientific Conference Transbaltica 2017:
Transportation Science and Technology

Traffic Flow Prediction using Kalman Filtering Technique

Selvaraj Vasantha Kumar*

School of Civil and Chemical Engineering, VIT University, Vellore, Tamil Nadu, India

Abstract

Traffic flow prediction is an important research problem in many of the Intelligent Transportation Systems (ITS) applications. The use of Autoregressive Integrated Moving average (ARIMA) or seasonal ARIMA (SARIMA) for traffic flow prediction requires huge flow data for model development and hence it may not be possible to use ARIMA in cases where sufficient data are unavailable. To overcome this problem, a prediction scheme based on Kalman filtering technique (KFT) was proposed and evaluated which requires only limited input data. Only previous two days flow observations has been used in the prediction scheme developed using KFT for predicting the next day flow values with a desired accuracy. Traffic flow prediction using both historic (previous two days flow data) and real time data on the day of interest was also attempted. Promising results were obtained with mean absolute percentage error (MAPE) of 10 between observed and predicted flows and this indicates the suitability of the proposed prediction scheme for traffic flow forecasting in ITS applications.

Peer-review under responsibility of the organizing committee of the 10th International Scientific Conference Transbaltica 2017

Keywords: intelligent transportation systems, traffic flow prediction, Kalman filtering, limited data

1. Introduction

Urban population increase due to economic growth in developing countries like India resulted in traffic bottlenecks in most of the metropolitan cities of our country. Availability of cars at affordable prices encourages personal vehicle usage for office and recreational trips thus resulting in lesser use of public transport which leads to

* Corresponding author.
E-mail address: svasanthakumar1982@gmail.com

1877-7058
Peer-review under responsibility of the organizing committee of the 10th International Scientific Conference Transbaltica 2017
doi:10.1016/j.proeng.2017.04.417

congestion on city roads. The solution options for reducing congestion are infrastructure expansion, Transportation System Management (TSM) measures, congestion pricing and technology applications coming under Intelligent Transportation Systems (ITS). Among various options, sustainable solutions like ITS are more preferred than capital-intensive construction strategies. Prediction of traffic volume is an important research problem in ITS addressed by many researchers in the past decades. The various mathematical techniques reported for prediction of traffic volume are regression, neural network, historical average algorithms and time series analysis [1–6]. The use of Autoregressive Integrated Moving average (ARIMA) or seasonal ARIMA (SARIMA) for traffic flow prediction requires huge flow data for model development and hence it may not be possible to use ARIMA in cases where sufficient data are unavailable. Even if huge database is available, requirement of specialized software for time series modelling and time consuming model development process may restrict the use of ARIMA or SARIMA models in problems dealing with real time data such as traffic flow. These constraints with ARIMA model motivate the identification of an alternate technique that could potentially be used for the problem of traffic flow prediction and overcome the existing limitations with ARIMA models. Hence, in the present study, a prediction scheme based on Kalman filtering technique (KFT) was proposed and evaluated. With only few lines of code in MATLAB, the prediction scheme proposed in this study can easily be implemented thus eliminating the requirement for specialized time series analysis software. As contrast to ARIMA models which requires huge database for model development, the proposed scheme requires only limited input data. Previous two days of flow observations converted to passenger car units (PCUs) was used in the prediction scheme to predict the next day flow values with a desired accuracy. Traffic flow prediction using both historic (previous two days flow data) and real time data on the day of interest was also attempted.

2. Study stretch, data collection and extraction

A two lane arterial road in Vellore in Tamilnadu, India was considered as the study stretch. Vellore being close to Chennai and having well connected road and railway network, it experiences a fast and tremendous urban growth with increasing number of vehicles in recent years after it has been declared as a city corporation in 2008. Fig. 1a) shows the Google map view of the study stretch along with prominent landmarks in Vellore such as old and new bus stand, railway station, Vellore Institute of Technology (VIT) University and Christian Medical College (CMC) hospital. As seen in Fig. 1a), both state (SH-9 connecting Cuddalore and Chittor) and national highways (NH-234 connecting Mangalore and Villupuram) passes through the study stretch and as there is no ring road in Vellore, all the bypassable traffic such as cars, trucks and multi-axle trailers from NH-234 to SH-9 or vice versa, uses the study stretch. In addition to the bypassable traffic, the regular traffic proceeding from Vellore town and adjoining residential areas such as Sathuvachari towards VIT and Katpadi railway station or vice versa also uses the same road. The proposal to construct a bridge across Palar river connecting Sathuvachari with Kangeyanallur and VIT has not yet been materialized and hence the study stretch experiences heavy traffic during most times of the day.

Traffic flow data was collected at the midblock of the selected study stretch using video survey from 7 am to 11 am consecutively for three days on 24th, 25th and 26th March 2014. Only one side of the traffic proceeding from Katpadi towards Vellore was considered for analysis. In order to take into account the vehicle heterogeneity as existing in India, five vehicle classes were considered, namely, two-wheeler, three-wheeler (autos), passenger car, light commercial vehicle (LCV), and bus/trucks. Each 5 minute traffic volume from 7 am to 11 am for all vehicle classes was manually extracted using the collected video data. The observed class-wise traffic volume was converted to equivalent passenger car units (PCU) using the PCU factors suggested in IRC-106 [7]. The PCU factors used were 0.75, 2, 1, 1.4 and 2.2 for two-wheeler, three-wheeler (autos), passenger car, light commercial vehicle (LCV), bus/trucks respectively. In IRC-106, two sets of PCU values were given according to the percentage composition of different vehicle types in the traffic stream. For the present case, since the proportion of two-wheelers and share autos were more than 10%, PCU values of 0.75 and 2 were used for two-wheelers and autos respectively instead of 0.5 and 1.2. Fig. 1b shows the two days PCU converted flow data which consists of a total of 96 flow values (48 observations of each 5 min. flow from 7 am to 11 am × 2 days). It can be seen from Fig. 1b that the flow gradually increases from morning 7 am and reaches a peak at around 10 am with maximum flow of 356 PCU's and then gradually decreases till 11 am. Similar phenomenon can be observed on day 2 also with flow gradually increases and reaches a maximum flow of 290 PCU's around 10 am and then it gradually decreases. Using the two days PCU converted flow data as input, the model using KFT was developed to predict the flow on the next consecutive day (26^{th} March 2014), the procedure of which is explained in the following section.

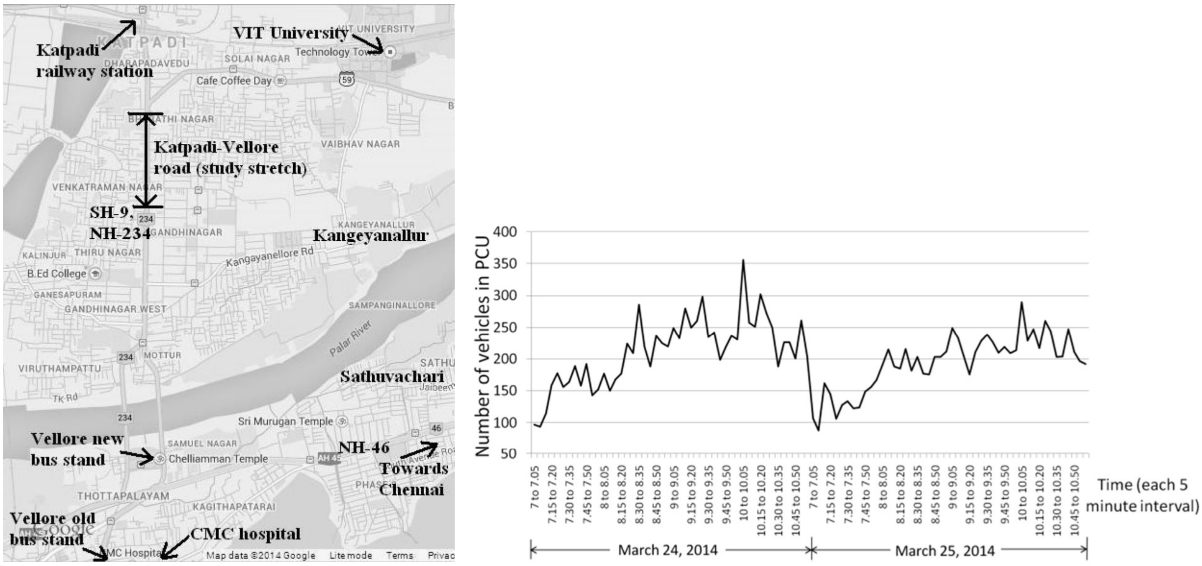

Fig. 1. Google map view of the study stretch along with prominent landmarks in Vellore (a); Input flow data used for model development (b).

3. Development of traffic flow prediction scheme using KFT

The Kalman filter [8] allows a unified approach for prediction of all processes that can be given a state space representation. According to [9], state space representations and the associated Kalman filter have a profound impact on many application areas. A state space model is generally represented in two equations: first equation is called the state equation which finds the state X_{t+1} in time $(t + 1)$ using the previous state X_t and a noise term as shown below.

$$X_{t+1} = F_t X_t + W_t \,,\; t = 1, 2, \ldots, \tag{1}$$

where F_t is a sequence of $v \times v$ matrices and W_t is the process disturbance ~ N (0, $\{Q_t\}$). The second equation is the observation equation which expresses the w dimensional observation Z_t as a function of a v dimensional state variable X_t and noise. Thus:

$$Z_t = \mathbf{G}_t X_t + V_t \,,\;\; r = 1, 2, \ldots \tag{2}$$

where V_t is the measurement noise ~ N (0, $\{R_t\}$) and $\{G_t\}$is a sequence of $w \times v$matrices and $\{W_t\}$ and $\{V_t\}$ are uncorrelated. In the present study, previous two days of flow observations (24^{th} and 25^{th} March 2014) were used as input for predicting the next day flow values (26^{th} March 2014) using the prediction scheme developed using KFT. The steps involved in the KFT algorithm based on (1) and (2) are explained below.

1. The time from 7 am to 11 am was divided into each 5 min. intervals and F_t in (1) was calculated using the flow data collected on 24^{th} March 2014 by dividing the flow X_{t+1} in time $(t + 1)$ by flow X_t in time (t).
2. The apriori estimate of the traffic flow on 26^{th} March 2014 was calculated using:

$$\hat{x}^{-}_{(t+1)} = F_t \hat{x}^{+}_{(t)} \,. \tag{3}$$

The symbol hat '^' indicates the estimate and the superscript '–' denotes the apriori estimate & the superscript '+'denotes the aposteriori estimate. Thus, the variable $\hat{x}^{-}_{(t+1)}$ is the apriori estimate of the predicted flow at time $(t+1)$. The variable $\hat{x}^{+}_{(t)}$ is the aposteriori estimate of the predicted flow at time t. The F_t calculated in step 1 was used in (3) to calculate the apriori estimate of the predicted flow at time $(t+1)$. Since the predicted traffic flow in the first time interval (7 am to 7.05 am) on 26th March 2014 is unknown, the actual or observed flow is taken in the place of $\hat{x}^{+}_{(t)}$ for the first 5 min. interval.

3. The apriori error variance (denoted by P^{-}) was calculated using:

$$P^{-}_{(t+1)} = F_t P^{+}_{(t)} F_t + Q_k \,. \tag{4}$$

4. The Kalman gain (denoted by K) was calculated using:

$$K_{(t+1)} = P^{-}_{(t+1)} \left[P^{-}_{(t+1)} + R_{(t+1)} \right]^{-1} . \tag{5}$$

The equations shown above from steps from 2 to 4 are the "time update equations". The time update equations projects forward in time the current state and error covariance estimates to obtain the apriori estimates for the next time step. The steps 5 and 6 as discussed below are called as the "measurement update" equations. The measurement update equations incorporate a new measurement into the apriori estimate to obtain an improved aposteriori estimate. The time update equations and measurement update equations are sometimes called as "predictor" and "corrector" equations respectively.

Steps 2 to 6 are executed recursively to obtain the predicted flow on 26th March 2014. A flowchart showing the proposed methodology is shown in Fig. 2.

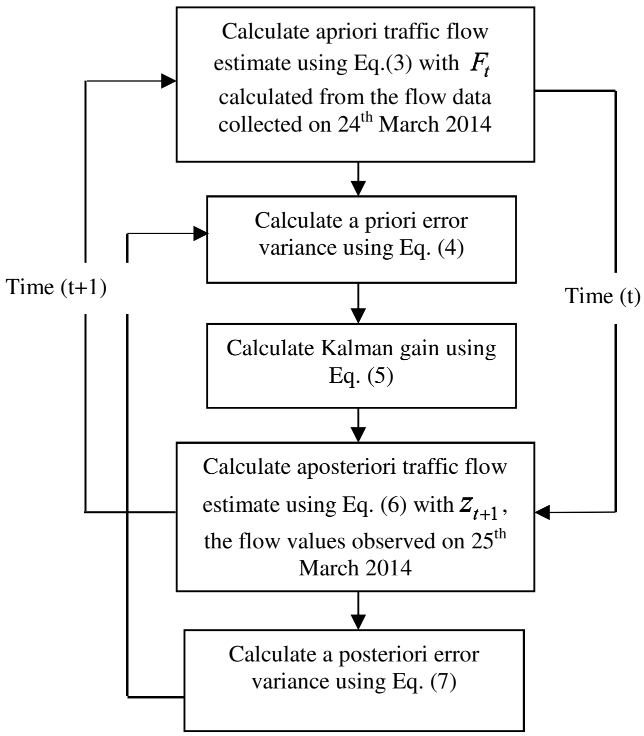

Fig. 2. Flowchart showing the proposed methodology.

5. The aposteriori estimate of the traffic flow on 26th March 2014 was calculated using,

$$\hat{x}^{+}_{(t+1)} = \hat{x}^{-}_{(t+1)} + K_{t+1}\left[z_{t+1} - \hat{x}^{-}_{(t+1)}\right]. \tag{6}$$

The variable $\hat{x}^{+}_{(t+1)}$ is the aposteriori estimate of the predicted traffic flow at time $(t + 1)$ on 26th March 2014, K_{t+1} is the Kalman gain calculated in step 4, $\hat{x}^{-}_{(t+1)}$ is the apriori estimate of the predicted flow at time $(t + 1)$ calculated in step 2. The flow values observed on 25th March 2014 was used in place of z_{t+1} in (6) for correcting the apriori estimate calculated in step 2.

6. The aposteriori error variance was calculated using

$$P^{+}_{(t+1)} = \left[1 - K_{t+1}\right]P^{-}_{(t+1)}. \tag{7}$$

4. Corroboration of the prediction scheme

The validation step involved the forecasting for 26th March 2014 using the previous two days of flow data as input, i.e., 24th & 25th March 2014 and comparing the predicted flows with that of observed values. The results are shown in Fig. 3. It can be seen that the predicted and observed flows follow closely and this shows the good performance of the developed model. The Mean Absolute Percentage Error (MAPE) is used as a measure of estimation accuracy and is calculated using

$$\boldsymbol{MAPE} = \frac{1}{\boldsymbol{n}}\sum_{i=1}^{n}\left|\frac{\text{Predicted flow} - \text{Observed flow}}{\text{Observed flow}}\right| \times 100, \tag{8}$$

where n is 48, the number of predicted flows. A MAPE of 10.56 was obtained between observed and predicted flows shown in Fig. 3. In general, any forecast with a MAPE of less than 10% can be considered highly accurate, 11–20% is good, 21–50% is reasonable and 51% or more is inaccurate. Based on this, it can be said that the model performs highly accurate with MAPE around 10%.

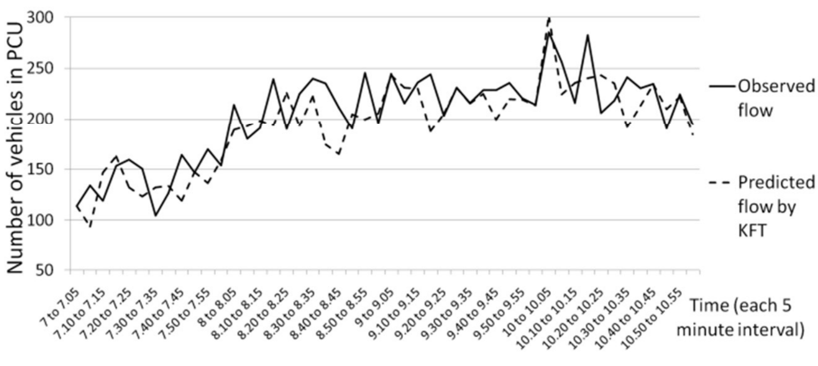

Fig. 3. Observed and predicted flow on 26th March 2014.

Traffic flow prediction taking into account the real time data of 26th March 2014 was also attempted using the prediction scheme developed based on KFT. That is, the real time data observed at time (t) on 26th March was used to predict the traffic flow in the next time interval $(t + 1)$. The results of predicted flows against the actual flow values are shown in Fig. 4. A MAPE of 10.23 was obtained and since the MAPE is around 10%, prediction result can be considered as highly accurate.

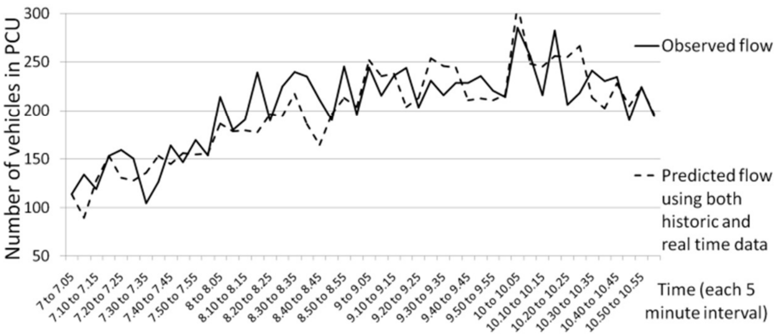

Fig. 4. Observed and predicted flow using both historic and real time data on 26th March 2014.

5. Concluding remarks

Predicting future traffic conditions is an important element in ITS applications. Most of the reported studies on traffic flow prediction use time series techniques such as ARIMA, which has some limitations such as software dependence, requirement of huge database, etc. To overcome the limitations with ARIMA models, in the present study, a prediction scheme based on KFT was proposed and evaluated. The proposed scheme requires only limited input data. Only previous two days flow observations converted to passenger car units (PCUs) and aggregated to the required time interval has been used in the prediction scheme developed using KFT for predicting the next day flow values with a desired accuracy. Traffic flow prediction using both historic (previous two days flow data) and real time data on the day of interest was also attempted. The results were promising with MAPE of around 10 was found between observed and predicted flows and this shows the suitability of the proposed prediction scheme in traffic flow forecasting for ITS applications.

References

[1] Y. Rajabzadeh, A. H. Rezaie, H. Amindavar, Short-term traffic flow prediction using time-varying Vasicek model, Transportation Research Part C 74 (2017) 168–181.
[2] H. Yang and X. Hu, Wavelet neural network with improved genetic algorithm for traffic flow time series prediction, Optik 127 (2016) 8103–8110.
[3] A. Cheng, X. Jiang, Y. Li, C. Zhang, H. Zhu, Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method, Physica A 466 (2017) 422–434.
[4] C. Wang and Z. Ye, Traffic flow forecasting based on a hybrid model, Journal of Intelligent Transportation Systems Technology, Planning, and Operations 20(5) (2016) 428–437.
[5] K. Kumar, M. Parida and V. K. Katiyar, Short term traffic flow prediction in heterogeneous condition using artificial neural network, Transport Journal 30(4) (2015) 397–405.
[6] C. Xu, Z. Li, W. Wang, Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming, Transport Journal 31(3) (2016) 343–358.
[7] IRC: 106, Guidelines for capacity of urban roads in plain areas, Indian Roads Congress (IRC), New Delhi. 1990.
[8] R. E. Kalman, A new approach to linear filtering and prediction problems, Trans ASME J Basic Eng. 82(1) (1960) 35–45.
[9] P. J. Brockwell, R. A. Davis, Introduction to time series and forecasting, Springer-Verlag, New York. 2002.