Chapter 15
Online Traffic Prediction Using Time Series: A Case study

M. Karimpour, A. Karimpour, K. Kompany, and Ali Karimpour

15.1 Introduction

Crashes and traffic congestion are among the most challenging issues in traffic engineering. Road capacities and road accidents have great impacts on traffic congestion. An accurate prediction of traffic flow is one of the significant steps in Intelligent Transportation Systems (ITS) [LiWa13]. ITS have enabled the engineers to get access to real time data [SmEtA02]. Real time data not only can be helpful for the road users to decide their routes for traveling, but also can give the chance to the engineer to manage the routes more effectively [Us12].

Different programs have long been employed to anticipate the point traffic, such as fuzzy neural approach [OgEtA11]. Also, using linear multi-regression dynamic approach helps anticipate data online [QuAl08]. Another approach to predict the traffic is artificial neural network [KuEtA13]. Other methods that have been utilized are Bayesian networks and neural network [SuEtA06].

M. Karimpour
Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: mostafa.karimpour1@gmail.com

A. Karimpour
Department of Highway and Transportation Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: Abolfazl.karimpour@gmail.com

K. Kompany
Department of Highway and Transportation Engineering, Virginia Tech, Blacksburg, VA, USA
e-mail: Kian.kompany@gmail.com

Ali Karimpour (✉)
Ferdowsi University of Mashhad, Mashhad, Iran
e-mail: karimpor@um.ac.ir

Time series are the data and information variables collected in the past and used to predict the data in the future. These data are collected in regular time intervals. One of the applications of time series is anticipating the stock exchange of big cities [CaEtA97]. Some other applications of time series include predicting water demand of metropolitan area using structural models, time series, and neural networks [JaKu07]. Forecasting the Forex financial markets is another application of time series [YaTa00].

Time series are also used to predict traffic intensity in various studies. In a recent study conducted in [CoHu14], time series utilized to predict the trucks short-term traffic by analyzing the recent traffic data of the trucks. Traffic prediction is one of the important factors in intelligent transportation systems, utilizing time series using previous data enable us to predict the forthcoming traffic and be to enhance the traffic conditions [RaEtA14].

This paper is organized as follows. Section 15.2 explains a methodology to examine the traffic in an arbitrary intersection. Section 15.3 provides an in-flow rate prediction method of the traffic data of the target intersection, and then illustrated the proposed scheme. The results are then presented in Section 15.4. Finally, Section 15.5 highlights the main results and draws concluding remakes.

15.2   Traffic Modeling by Mixed Logic Dynamic

Mixed logic dynamic (MLD) is a common framework that combines both continues states with logical states as the following general form [BeMo99]:

$$x(t+1) = A_t x(t) + B_{1t} u(t) + B_{2t} \delta(t) + B_{3t} z(t)$$

$$y(t) = C_t x(t) + D_{1t} u(t) + D_{2t} \delta(t) + D_{3t} z(t)$$

$$E_{2t} \delta(t) + E_{3t} z(t) \leq E_{4t} x(t) + E_{1t} u(t) + E_{5t}$$

where $x(t)$, $u(t)$, and $y(t)$ are the states, inputs, and outputs, respectively, with continues and logical elements as:

$$x = \begin{bmatrix} x_c \\ x_l \end{bmatrix}, x_c \in R^{n_c}, x_l \in \left(0, 1\right)^{n1}, n = n_c + n_l$$

$$u = \begin{bmatrix} u_c \\ u_l \end{bmatrix}, u_c \in R^{m_c}, u_l \in \left(0, 1\right)^{m_l}, m = m_c + m_l$$

$$y = \begin{bmatrix} y_c \\ y_l \end{bmatrix}, y_c \in R^{p_c}, y_l \in \left(0, 1\right)^{p_l}, p = p_c + p_l$$

other variables are logic (binary) variables.

Fig. 15.1 A four connection intersection [Pa15].
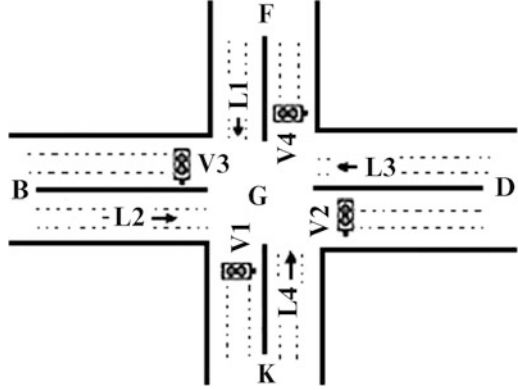
Traffic modeling could be a good candidate for the MLD structure. That is, the queue length can be considered as the continuous state and the signaling of intersection can be considered as the logic values. Figure 15.1 illustrates an arbitrary intersection, the queue length for every approach of the intersection can be derived as:

$$L_i(k+1) = L_i(k) + \tfrac{T_s}{M_{fi}}(q_{in,i}(k) - v_i(k)q_{out,i}(k)) \quad v_i(k) \in (0,1)$$

where L_i is the queue length of ith approach, T_s is sampling period, M_{fi} is the number of lane in the ith approach, $q_{in,i}$ is the input flow rate of the ith approach, $q_{out,i}$ is the output flow rate of the ith approach, and v_i is a binary variable that shows the ith approach signal is green ($v_i = 1$) or red ($v_i = 0$). The output flow rate of the ith approach is:

$$q_{out,i}(k)v_i(k) = min\left(M_{fi}m_iv_i(k), \tfrac{M_{fi}L_i(k)}{T_s}v_i(k)\right)$$

where m_i is the maximum output rate of the ith approach. Since out flow rate has a nonlinear formula, it can be rewritten as [BeMo99]:

$$q_{out,i}(k)v_i(k) = z_{if}(k)\delta_i(k) + z_{3f}(k) \quad (\delta_i(k) = 1 \Leftrightarrow z_{if}(k) \leq 0)$$

where

$$z_{if} = M_{fi}m_fv_i(k) - \tfrac{M_{fi}L_i(k)}{T_s}v_i(k), \quad z_{3f} = \tfrac{M_{fi}L_i(k)}{T_s}v_i(k)$$

and δ_i is a binary variable.

In a four phase intersection at each time one light is green. The order of lights are:

$$v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow v_4 \rightarrow v_1$$

To consider the order in a four phase system, the following equations must be satisfied:

$$v_1(k-1)(1-v_1(k)) - v_2(k) \leq 0$$
$$v_2(k-1)(1-v_2(k)) - v_3(k) \leq 0$$
$$v_3(k-1)(1-v_3(k)) - v_4(k) \leq 0$$
$$v_4(k-1)(1-v_4(k)) - v_1(k) \leq 0$$

Also, to avoid interfering movements the following set of equations must be satisfied:

$$J_j v(k) \leq 1$$

where J_j is jth row of matrix J and J is:

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

By using the recent formula for output flow rate all equations can be mixed in the MLD framework. By modeling an intersection in the MLD framework, and solving the derived formulation by model predictive control [Pa15], Figure 15.2 have been achieved. Figure 15.2 demonstrates the number of cars in the approach #3 (Figure 15.1) for an arbitrary sequence (the signal order is not fixed) and regular sequence. Figure 15.3 shows the line signal of the approach #3 for an arbitrary sequence and regular sequence.
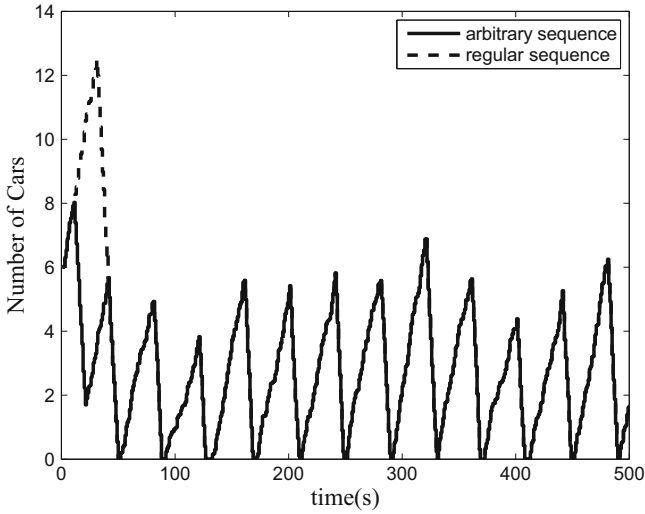
Fig. 15.2 Number of cars in approach #3 in Figure 15.1 for both arbitrary and regular sequence
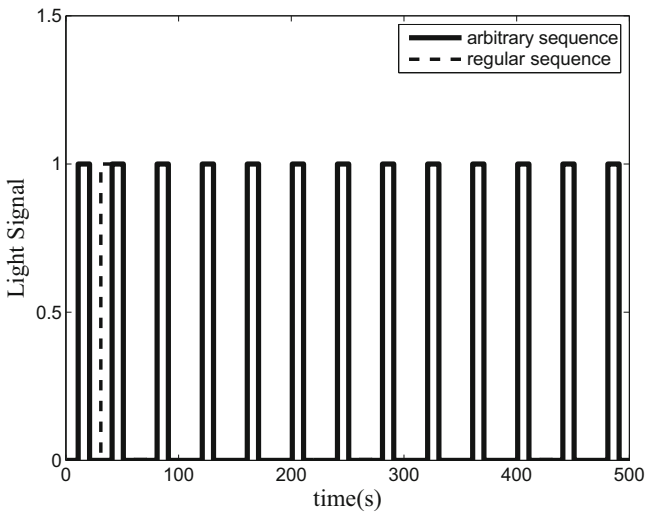

Fig. 15.3 Light signal in approach #3 in Figure 15.1 for both arbitrary and regular sequence

Mixed logic dynamic method is used to model the intersection sequences. For modeling part, the flow rate of each approach is used as the input to our model. Finally, the model is solved using model predictive control formula. Next section will provide an approach to predict input flow rate (in-flow rate) of each approach in an arbitrary intersection.

15.3 In-Flow Rate Prediction

In order to predict the in-flow traffic in an intersection, a time series model is incorporated. All the steps of building a time series for traffic prediction are explained in the following sections.

A. Time Series Models

In the physical applications of time series, when studying on a variable, there are some situations where the relevant input variable is not specified. In other words, there are some important variables which may leave effect on outputs, but one cannot manipulate them.

Statistician and economists use time series terminology to explain these systems. Some examples of time series are the world oil price, daily price of power markets, and traffic flow in an intersection in this study. For instance, traffic flow in an intersection is affected by too many parameters, i.e., ambient temperature, time of day, seasonal situation, etc. But none of these variables can be manipulated.

In time series, there is no manipulated input but there are lots of inputs that affect the output of a system. AR model is one kind of time series representation. In AR model, output of system is derived in an autoregressive (AR) manner based on the previous value of outputs:

$$y(t) + \alpha_1 y(t-1) + \alpha_2 y(t-2) + \ldots + \alpha_n y(t-n) = e(t)$$

$y(t)$ is the output in time t and $e(t)$ is white noise input with variance λ. The parameters of AR models are $\theta = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_n]^T$ and can be derived by least square method (LSM):

$$\theta = [\Phi^T \Phi]^{-1} \Phi^T Y$$

Φ and Y are defined in [Lj99]. In some situation white noise assumption may be restrictive so one can use ARMA model. In ARMA model output of system is derived in an autoregressive manner and the input is in moving average (MA) such that:

$$y(t) + \alpha_1 y(t-1) + \ldots + \alpha_n y(t-n) = e(t) + c_1 e(t-1) + \ldots + c_m e(t-m)$$

$y(t)$ is the value of output in time t and $e(t)$ is white noise input with variance λ. The parameters of ARMA models are

$$\theta = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_n \quad c_1 \quad \ldots A_{\mathbf{.}} \quad c_m]^T$$

and can be derived by several different methods [Lj99].

A more convenient way to define ARMA model is

$$A(q)y(t) = C(q)e(t)$$

where $A(q) = 1 + \alpha_1 q^{-1} + \ldots + \alpha_n q^{-n}$ and $C(q) = 1 + c_1 q^{-1} + \ldots + c_m q^{-m}$ and q is a backward shift operator.

B. Proposed Method

This paper uses ARMA model to predict traffic flow in the specific intersection. Here $y(t)$ is considered as the traffic flow in time t and $e(t)$ is the combined effect of variables which may affect the traffic flow $y(t)$. The data used in this paper are obtained from Mashhad Traffic Organization, for Moallem Blvd. intersection. The traffic flow for each approach of the intersection, for every 15 minute intervals are collected for several days using loop detectors.

15.4   Experimental Results

I. Experiment Number 1

The system is learned with data from the first day with 15 minute intervals. Then, the prediction was made 15 minutes ahead for the following day. Moreover, the ARMA model applied in this system that uses the data of the several 15 minutes before the traffic flow. In other words, 12 samples of traffic flow (3 hours) are applied to predict the traffic flow for the 15 minutes ahead. Coefficients used to predict this model are as follows:

$$A(q) = 1 - 0.8478q - 1 - 0.3553q^{-2} + 0.2323q^{-4} + 0.03152q^{-5} + \ldots$$

$$C(q) = 1 - 0.3099q^{-1}$$

As it is shown in Figure 15.4, prediction is made after the 12th sample and all previous 12 samples are available data. In this condition, the model accuracy is 88.74 % (Table 15.1). To examine the prediction system more thoroughly, the anticipation is also made for one hour ahead. The result is shown in Figure 15.5.

As it is demonstrated in Figure 15.5, accuracy of the system is decreased a little. Considering Figure 15.5, it can be concluded that because the prediction interval has increased, the accuracy of the system has reduced to 81.96 % (Table 15.1).

II. Experiment Number 2

The same system in this step is learned with data from the first day with every 30 minutes intervals. Then the anticipation was made 30 minutes ahead for the following day. The result of the prediction for 30 minute ahead in the following day and 60 minute ahead is illustrated in Table 15.2. From Table 15.2 it can be inferred that the accuracy of the system decreased significantly.
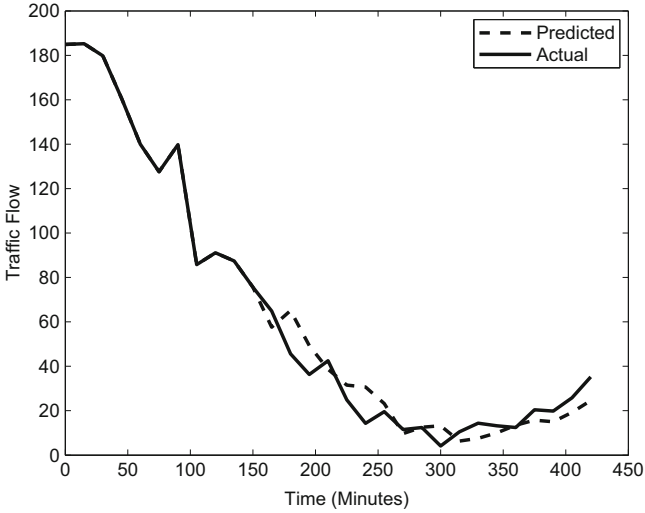
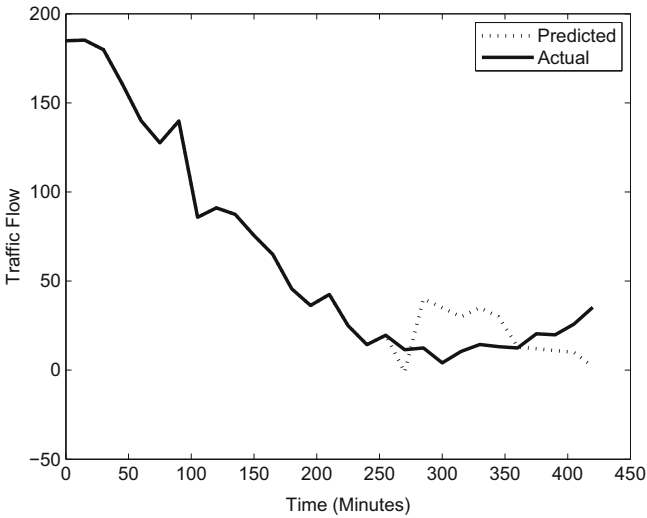**Fig. 15.4** Results for the system predicted output for 15 min later

Fig. 15.5 Results for the system predicted output for 1 hour later

By comparing the two experiments which are shown in Table 15.2, it can be noted that the less the time intervals are, the more the accuracy will be achieved. Thus, it is significant to record data in shorter intervals.

Table 15.1 System accuracy (least square error) for 15 minutes interval learning data

Accuracy of the system	Forecasting ahead
88.74 %	30 min
81.96 %	60 min

Table 15.2 Different system accuracy (least square error) for two experiments

Experiment 1		Experiment 2	
Accuracy of the system	Forecasting ahead	Accuracy of the system	Forecasting ahead
88.74 %	30 min	73.76 %	30 min
81.96 %	60 min	66.01 %	60 min

15.5 Conclusion and Discussion

The burgeoning rise of car production and the following traffic increase have made it inevitable to propose a model for the online traffic control. Implementing a model for predicting traffic flow in an online method for a definite intersection would help traffic engineers to plan for the upcoming traffic of that intersection. Since the ITS systems depend on real time information, it is mandatory that to be able to have access to online models. It is important for the traffic engineers to have meticulous information about the traffic flow, before the planning and construction phase. Moreover, the ability to predict the traffic flow in a built intersection for 15 minutes or one hour later has advantages such as:

1. If the flow is predicted to be more than the intersection capacity, the approaching ways to the intersection can be limited.
2. With precise flow prediction, smaller queue length can be derived.
3. Sufficient time would be available to send police units or in case of other emergencies.
4. The intersection traffic light can be changed easier manually if the intersection is not actualized.

In this study, time series are used to create the model. Two different experiments conducted on the same data. In the first experiment, the proposed model is instructed by the data of the first day having 15 minutes intervals, and the model is tested for 15 and 60 minutes later time intervals in the following day. The average error obtained from this test is less than 18 percent for all the predictions. In the second experiment, the mode is instructed by the data of the first day having 30 minutes intervals, and the model tested for the same period as the first one. But the average error gained from the second experiment increased significantly. So using data by less interval measurements is much better than data with large interval time.

Traffic intersections may also be linked as in a network. That is, the neighbor intersections should follow a green wave to provide the most efficiency. In this

paper, we only analyzed the impact of all the four approaches in an isolated arbitrary intersection. Future research can be conducted on the impact of an intersection in a network.

References

[BeMo99]   Bemporad, A., Morari, M.: Control of systems integrating logic, dynamics, and constraints. Automatica **35**, 407–427 (1999)

[CaEtA97]   Caldarelli, G., Marsili, M., Zhang, Y.C.: A prototype model of stock exchange. Europhys. Lett. **40**(5), 479 (1997)

[CoHu14]   Cordova, F., Huynh, N.: Using economic indicators to perform short-term truck traffic forecasting a time series and truck traffic analysis framework. TRB Annual Meeting, Current Research in Freight Transportation and Logistics Planning and Operations (2014)

[JaKu07]   Jain, A., Kumar, A.M.: Hybrid neural network models for hydrologic time series forecasting. Appl. Soft Comput. **7**(2), 585–592 (2007)

[KuEtA13]   Kumar, K., Parida, M., Katiyar, V.K.: Short term traffic flow prediction for a non-urban highway using artificial neural network. Proc. Soc. Behav. Sci. **104**(2), 755–764 (2013)

[LiWa13]   Liang, Z., Wakahara, Y.: City traffic prediction based on real-time traffic information for intelligent transport systems. In: IEEE, 13th International Conference on ITS Telecommunications (ITST), pp. 378–383 (2013)

[Lj99]   Ljung, L.: System Identification Theory for the User. Prentice Hall Information and System Science Series. Wiley (1999)

[OgEtA11]   Ogunwolu, L., Adedokun, O., Orimoloye, O., Oke, S.A.: A Neuro-Fuzzy approach to vehicular traffic flow prediction for a metropolis in a developing country. J. Ind. Eng. Int. **7**(13), 52–66 (2011)

[Pa15]   Pahnabi, A.H.: Intersection traffic control with uncertainties based on mixed logical dynamical model and robust model predictive control. M.Sc. Dissertation (2015)

[QuAl08]   Queen, C.M., Albers, C.J.: Forecasting traffic flows in road networks: a graphical dynamic model approach. In: Proceedings of the 28th International Symposium of Forecasting, International Institute of Forecasters (2008)

[RaEtA14]   Raeesi, M., Mesgari, M.S., Mahmoudi, P.: Traffic time series forecasting by feedforward neural network: a case study based on traffic data of Monroe. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **40**(2), 219–223 (2014)

[Us12]   United State Department of Transportation (USDOT-2012)

[SmEtA02]   Smith, B.L., Williams B.M., Oswald K.R.: Comparison of parametric and nonparametric models for traffic flow forecasting. Trans. Res. Part C **10**, 303–321 (2002)

[SuEtA06]   Sun, S., Zhang, C., Yu, G.: A Bayesian network approach to traffic flow forecasting. IEEE Trans. Intell. Trans. Syst. **7**(1), 124–132 (2006)

[YaTa00]   Yao, J., Tan, C.L.: Forecasting the forex financial markets. Neurocomputing **34**, 79–98 (2000)