International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Index Copernicus Value (2015): 56.67 | Impact Factor (2017): 5.156

A Novelty Markov Design to Model Traffic in a Road Network

Gayani K. Thalagoda[1], Sachith P. Abeysundara[2]

Department of Statistics and Computer Science, University of Peradeniya, 20400, Sri Lanka

Abstract: ***Traffic congestion in the urban areas has become an unremitting problem faced by city planners. Stochastic Methods such as Markov Chains and Monte Carlo Simulations produce fruitful results, hence, are proven to be highly effective in terms of predictions along a segment of highway, freeway etc. However, a major drawback of these methods is the lack of feasibility when applying to the road networks in whole, due to the high complexity of road network. Our work introduces a novelty concept based on Markov Models to overcome the above drawback of stochastic models when applied to traffic congestion in a road network in whole. The research focuses on striking a balance between the route capacity and total surface area of vehicles of each type. A road system with multiple junctions and lanes is considered and represented in matrix form. The counts of each vehicle type that enters a road segment are also represented in matrix form. Thereafter, the total surface area of vehicles occupying each road segment is expressed as a proportion of road capacity. We estimate the traffic state transition for each corresponding segment. Such transition matrices can be used to identify the converging traffic state in future. We initially conduct simulation runs for different initial conditions. The proposed method successfully provides transition matrices for the given road network. Future studies involve extending the proposed method by considering other factors such as the vehicles parked on either sides of the roadway, vehicle speed and weather conditions.***

Keywords: Traffic congestion, Markov chains, Stochastic models

1. Introduction

1.1. Introduction to the problem

The large urban agglomerations are facing the negative effects of traffic congestion, the most unpleasant phenomenon of the contemporary society (Chen and Li 2010). It has caused serious impediments to the quality of life of urban populations. Many adverse situations such as excessive fuel consumption, escalation in vehicle operating costs, unrestrained time waste, escalation in road accidents and fatalities due to frustrated drivers stuck up in traffic could arise due to the above phenomenon. Thus, it emphasizes the need to perceive an optimum solution for the traffic scenario. In the quest for meeting the capacity demand for road traffic, it has now been realized that constructing new roads is not a feasible solution given the high initial cost and lack of space for new roadways in the metropolitan areas. Moreover, the time it takes to finish constructing a new road way and the disruption it can bring about, to the current road network during the process of construction could be considered as major factors which make construction of new road ways infeasible. However, the recent advances in electronics, sensors and computers provide room to develop flexible strategies through effective utilization of existing infrastructure. Use of this advanced technology will significantly improve the quality of data obtained and will help to control traffic scenarios more accurately.

1.2. Background

Traffic phenomenon has been extensively studied under diverse methodologies. Mathematical deterministic models and stochastic models stand out as two major approaches of analysis. However, the highly complex dynamical nature of traffic, tend to repudiate mathematical models due to its simplicity and in- adequate capacity for analysis. Thus, stochastic models, which use simulation mechanism, are more preferred. Simulation models are developed using necessary software to suit the traffic flow of a given network. Then, the developed model is validated using real data and then fine-tuned to suit other different scenarios and control strategies. Stochastic Methods such as Markov Chains, Monte Carlo Simulation, Grey Markov models (Zang et al. 2011) produce fruitful results and are proven to be highly effective in terms of predictions along one segment of a highway, freeway etc. (Antoniou 2007). However, a major limitation in all these methods is the lack of feasibility in applying to the road networks in whole, due to the high complexity of road network. (Qi 2010). Under the study in concern, the primary objective is to build a stochastic Markov model to overcome the above drawback of stochastic models and to determine when and where heavy traffic is most likely to occur in a freeway road network system at a given future time interval. To construct a convenient model based on the vehicle type that could be utilized even by the traffic control units to predict the traffic congestion along particular road segments with in a city. This paper uses total surface area occupied by vehicles as a measure of traffic density, which ultimately acts as an index to vehicular traffic. Association between road capacity and vehicular surface area (Surface area of road covered by the vehicle) is being examined. It should be significantly noted that the term 'surface area' mentioned under the study exclusively refer to the surface area of the road occupied by the vehicle. The term 'surface area' does not imply the surface area of the whole body of the vehicle. Since the surface area of a vehicle varies with the type of the vehicle, this particular study is designed with the 'vehicle type' element incorporated. Investigating the effect of vehicle type on traffic phenomenon is also to be carried out through the study. Analyzing the contribution to traffic by each vehicle type is also a significant aim of the study in concern. This implies investigating how vehicle type can contribute to traffic congestion during peak periods. Another objective is to determine when and where traffic congestion takes place in a system. By developing the model in such a way that the model predicts the traffic condition in a particular road segment during a future time interval, and there by advising

International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Index Copernicus Value (2015): 56.67 | Impact Factor (2017): 5.156

the drivers to determine best possible time to start the journey and best possible route to take to reach their final destination. The model assumes that the transition states are stationary over time. Thus, this model is divided in to peak times and off-peak times for a better interpretation, (Metroul 2015). This model will later be analyzed with real data for validation.

In the very initial stage of the study, it attempts to design and develop an innovative macroscopic stochastic traffic flow model to suit the specific needs of the objectives of the study. This includes, a major concern awarded to the vehicle type. A simple road network system with comparatively fewer number of road segments connected by a fewer number of junctions is considered. A simulation study to check the performance of the designed model is conducted next. Due to the absence of readily available data to fit the purpose, Statistical techniques are used to generate necessary data at this level for the simulation study. In this stage, the simulated results are applied to the model and model performance is examined. A numerical algorithm for future traffic state prediction based on this model is obtained at this level. Since the designed experiment yields promising results, further development of the model is discussed. Thereafter, the proposed model is applied to a real world scenario and the results are discussed. The viability of the model is comprehensively discussed. Finally, a conclusion regarding the overall performance of the model is eloquently expressed. Further improvements necessary for the model too are elucidated.

2. Methodology

Traffic congestion is a condition on transport networks that occurs as use increases. Slower speeds, longer trip times, and increased vehicular queuing characterize it. The most common example is the physical use of roads by vehicles. When traffic demand is great enough to affect the interaction between vehicles, it slows the speed of the traffic stream, resulting in congestion. As demand approaches the capacity of a road (or of the intersections along the road), extreme traffic congestion sets in. Traffic congestion can lead to drivers becoming frustrated and engaging in road rage. Thus, the significance of studying traffic congestion and pursuing solutions are of vital necessity. Traffic congestion scenario can be studied under various methodologies. Both deterministic methods and stochastic methods can be utilized for analyzing purposes. However, stochastic processes are proven to be more viable. In deterministic models, it is assumed that the output of the model is fully determined by the parameter values and the initial conditions, where as in stochastic models, there is a certain inherent randomness involved. The same set of parameter values and initial conditions will lead to an ensemble of different outputs. Since traffic scenario too has a significant random component associated with it, stochastic processes are more viable in studying traffic congestion along road networks.

2.1. Stochastic Processes

A stochastic process is a random process evolving with time (Dodge and Yadolah 2006) According to the probability theory; a stochastic process is a time sequence representing the evolution of some system represented by a variable whose change is subject to a random variation (Lawler 2006). Given a probability space $\{\Omega, F, P\}$ and a measurable space $\{S, \Sigma\}$, an S-valued stochastic process is a collection of S-valued random variables on $\{\Omega\}$, indexed by a totally ordered set T (”time”). That is, a stochastic process X is a collection, $\{X_{t:}\ t \in T\}$ here each $\{X_t\}$ is an S- valued random variable on $\{\Omega\}$. The space S is then called the state space of the process. In a stochastic or random process, there is some indeterminacy element involved. Even if the initial condition or initial parameters of the state is known, there are several directions in which the process may evolve. In many stochastic processes, the movement to the next state or position depends on only on the current state, and is independent from prior states or values the process has taken. Common examples of stochastic processes include stock market and exchange rate fluctuations and also random movements such as Brownian motion or random walks (Thiele et al. 1880). The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire journals, and not as an independent document. Please do not revise any of the current designations.

2.2. Markov Processes

Markov Chains is a major type of a stochastic process. If one can make predictions for the future of the process based solely on its present state just as well as one could know the process’s full history. i.e., conditional on the present state of the system, its future and past are independent. In discrete time, the process is known as a discrete-time Markov chain (DTMC). It undergoes transitions from one state to another on a state space, with the probability distribution of the next state depending only on the current state and not on the sequence of events that preceded it. When time is taken as continuous, the process is known as a Continuous-time Markov chain (Norris 1097). It takes values in some finite state space and the time spent in each state takes non-negative real values and has an Exponential distribution. Future behavior of the model (both remaining time in current state and next state) depends only on the current state of the model and not on historical behavior. A stochastic process has the Markov property if the conditional probability distribution of future states of the process (conditional on both past and present states) depends only upon the present state, not on the sequence of events that preceded it. A process with this property is called a Markov process. A very fine example is a game of snakes and ladders. The next state of the board depends only on the current state, and the next roll of the dice. It doesn’t depend on how things got to their current state. The previous moves of the game don’t have an effect on the next move. Thus, the next state of the game depends only on the current state. Under the study in concern, we hope to apply a Discrete Markov Chain by taking vehicle density as the main variables.

2.3. The Study in concern

The main attempt of the research study was to build from scratch, a simple but effective stochastic model that could be conveniently used even by external parties to prognosticate when and where traffic would occur. Since building a stochastic model is a major aspect of the study, designing an experiment to suit the need is crucial. When considering the road congestion, different vehicle types contribute to traffic congestion in different levels. Contribution to road traffic by containers and other large vehicles is relatively higher than the contribution by motorcycles or three wheelers. Thus, a major focus is attributed to the vehicle type. Further, the designed experiment shall also address to the matter of handling complexity of a road network in its entirety. The selected stochastic approach is a Markov Model. Under the study in concern, we first design an experiment using Markov concept. The designed experiment tends to both the issues mentioned above. It takes in to consideration both vehicle type as well as the road network at its entirety. Thereafter, the designed experiment is tested with simulated data. Since readily available data necessary for the experiment is not found, the model building in initially carried out using simulated data. As the initial step, we focus on rather a simple road network system with comparatively fewer number of road segments connected by a fewer number of junctions. Under the study in concern, we chose a road network system with m junctions. Next we graphically represent the road network system. Then it is converted in to a matrix form. We define a $m \times m$ matrix 'A' called the 'Adjacency matrix' (Reiter 2015), which shows all the connections of the road network. The elements of the matrix M can be defined as follows (equation 1).

$$A_{ij} = \begin{cases} 1 \text{ ; if i and j are directly connected} \\ 0 \text{ ; if i and j are not directly connected} \end{cases} \quad (1)$$

Thereafter, counts of vehicles of each vehicle type in consideration are obtained. The vehicle type v could be fundamentally recognized as cars, vans, Lorries, three wheelers etc. It should be noted that the any number of types of vehicles could be considered under the model, due to the dynamic and convenient nature of the model. The data collected are inserted in to separate matrices in such a way that each matrix represents one vehicle type. Next, the surface area occupied by each vehicle type is sought. The surface area occupied by each vehicle on the surface of the road is hereafter addressed as the "areal density". Then, the total surface area of vehicles occupying each road segment is separately calculated and expressed as a proportion of road capacity. Based on the congestion values obtained, traffic states S are identified as follows.

Definition of traffic congestion states;

$$\text{State} = S_i \text{if} c_{i-1} \leq \; d \leq c_i ; i{=}1, 2, \ldots, k \;\; and \; c_0 = 0$$

Where d is the areal density and the constants c_1 through c_k can be determined according the nature of the traffic network. The task is carried out by analyzing graphically, the proportional values. The traffic congestion variation within each time period is observed graphically and thereby, optimum classification of traffic congestion states is carried out.

Thereafter, each road segment having a direct connection is separately considered and the manner in which how the traffic congestion along that road segment changes within the three-hour period is graphically represented. Next, the transition matrix for each road segment is constructed and Markov Chain concept is implemented to prognosticate the traffic congestion state along a particular road segment in future time intervals. The model results are discussed. Further, the behavior of the transition of traffic states under different initial conditions are analyzed and discussed.

3. Study Design and Simulation

3.1. Summary of the simulated data set

Initially, an experiment to test the designed experiment with the use of simulated data is carried out. The main objective of perpetrating a simulated study is to check the appropriateness of the designed model in attaining our core objectives and to distinguish areas of the study that needs to be developed further.

The main variables of the data set under study, is the vehicle type, count, areal density. Four vehicle types, namely, cars, buses, motorcycles, three wheelers are considered to start with. The count of each vehicle type along a particular road segment during five minutes intervals is obtained. This procedure is carried out for an aggregate time period of three hours.

Due to the unavailability of a readymade data set, data has to be simulated. Since the basic stages of the study deals with the simulated data, the vehicle flow is sought reasoning. Vehicle flow assumed to have a poison distribution in general. Vehicle flow of each vehicle type (cars, vans, motorcycles, buses) along a particular road segment is reckoned to have poison distributions with separate parameter values. It is noteworthy to point out that, the parameter values of each vehicle type, along a particular road segment are also simulated for the convenience and un-biasedness of the study. Once the data have been simulated, the data are converted in to the matrix form.

The matrix form of the data depends solely on the nature of the road network. The number of rows and number of columns of the matrix is equal to the number of junctions in the selected road network. Poisson Distribution is given below.

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (2)$$

Where x is the number of vehicles and λ is the average number of events (vehicles) per unit time.

3.2 Graphical representation of the network

The road network under simulation study is as below. It comprises of 7 junctions, thus the adjacency matrix contains seven rows and seven columns. If, two junctions are directly connected by one road segment, it is represented by "1" in the adjacency matrix. Similarly if two junctions are not directly connected, then it is denoted by a zero.

Once the adjacency matrix is obtained, next step is to obtain the respective vehicle counts of each road segment. A 3D matrix of dimension (7, 7, 36) is obtained. There are 20 connections, given that vehicles can travel either ways along all the road segments. This implies there exist 36 of 7*7 matrices, having 20 non-zero values, one for each time frame of 5 minutes. A 3D matrix for each vehicle type is obtained the same way. Thus, 4 such matrices are taken in to consideration under the simulation study.

Time Frame 1: T_1

$$\begin{bmatrix} 0 & 14 & 0 & 0 & 0 & 0 & 3 \\ 7 & 0 & 17 & 0 & 7 & 0 & 8 \\ 0 & 9 & 0 & 10 & 5 & 0 & 0 \\ 0 & 0 & 15 & 0 & 0 & 8 & 0 \\ 0 & 1 & 6 & 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 5 & 4 & 0 & 13 \\ 7 & 15 & 0 & 0 & 0 & 8 & 0 \end{bmatrix}$$

Time Frame 2: T_2

$$\begin{bmatrix} 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 7 & 0 & 10 & 0 & 13 & 0 & 7 \\ 0 & 10 & 0 & 14 & 7 & 0 & 0 \\ 0 & 0 & 15 & 0 & 0 & 10 & 0 \\ 0 & 10 & 5 & 0 & 0 & 21 & 0 \\ 0 & 0 & 0 & 3 & 5 & 0 & 14 \\ 7 & 15 & 0 & 0 & 0 & 10 & 0 \end{bmatrix}$$

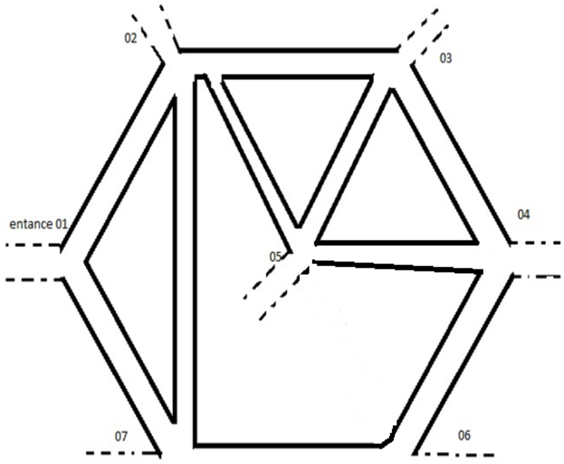

Figure 1: Graphical representation of the road network

3.3. Determination of road occupancy

Once the data matrices, line matrix (L) and road surface area matrix (R) are obtained, the study is preceded to the next level. At this level, sum of surface areas of all the vehicles during a 5-minute interval is calculated. This results in 36 matrices, comprising of the total surface area occupied by vehicles along each segment.

Time Frame 1: T_1

$$\begin{bmatrix} 0.000 & 669.308 & 0.000 & 0.000 & 0.000 & 0.000 & 441.931 \\ 460.656 & 0.000 & 614.276 & 0.000 & 500.105 & 0.000 & 488.226 \\ 0.000 & 376.613 & 0.000 & 386.143 & 386.608 & 0.000 & 0.000 \\ 0.000 & 0.000 & 689.653 & 0.000 & 0.000 & 324.265 & 0.000 \\ 0.000 & 319.023 & 384.799 & 0.000 & 0.000 & 711.732 & 0.000 \\ 0.000 & 0.000 & 0.000 & 247.339 & 306.150 & 0.000 & 508.511 \\ 447.962 & 696.964 & 0.000 & 0.000 & 0.000 & 621.453 & 0.000 \end{bmatrix}$$

Once the set of matrices containing the sum of the surface areas of the vehicles are obtained, it is represented as a ratio of the road capacity area. Here after, the measure is known areal density. The density matrices depict the road occupancy by vehicles during a given time interval. This, step is very crucial to our study, since our Markov Model is built around this variable. It is noteworthy that since it is still at the simulation level, Surface area of the road is arbitrarily taken.

Time Frame 1: T_1

$$\begin{bmatrix} 0.000 & 0.335 & 0.000 & 0.000 & 0.000 & 0.000 & 0.295 \\ 0.307 & 0.000 & 0.410 & 0.000 & 0.556 & 0.000 & 0.271 \\ 0.000 & 0.270 & 0.000 & 0.116 & 0.3870 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.531 & 0.000 & 0.000 & 0.324 & 0.000 \\ 0.000 & 0.266 & 0.192 & 0.000 & 0.000 & 0.445 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.215 & 0.102 & 0.000 & 0.514 \\ 0.224 & 0.268 & 0.000 & 0.000 & 0.000 & 0.691 & 0.000 \end{bmatrix}$$

At the proceeding step, all the data contained in the matrices are extracted to obtain the density fluctuations along a particular road segment throughout the observational time period. Here, each matrix is converted to a data frame initially. The data frame represents the connection with the direction, the density of respective connection and the time frame to which the data belongs. The main aim of the intended study is to recognize the variation of areal density along a particular road segment, with in the pre mentioned time period. Thus, it is consequential to extract road segments separately and further analyze them. The ultimate goal of this extraction procedure is to recognize areal density variation along a particular road segment with time.

Given below is data frame constructed through extraction of data for the time frame T1. At the end of the extraction process, there is expected to be 36 such data frames, for each five-minute interval.

Once the necessary data have been extracted, determination of the congestion states is determined. Since, this is a simulated study, exact real data are not being used. Data are used with logical reasoning. Techniques of graphical representation are used primarily in order to recognize the traffic congestion states.

Once the graphical representation is obtained, data contained in the matrices it is used to recognize the range of areal density variation. This aids in recognition of the traffic states, which is crucially essential in the construction of transition matrices.

Table 1: Areal density of the road network at time frame T_1

	Connections	Areal density
1	2-1	0.715
2	7-1	0.701
3	1-2	0.733
4	3-2	0.987
5	5-2	1.250
6	7-2	0.631
7	2-3	0.757
8	4-3	0.313
9	5-3	0.905
10	3-4	1.292
11	6-4	1.065
12	2-5	0.692
13	3-5	0.495
14	6-5	1.009
15	4-6	0.600
16	5-6	0.268
17	7-6	1.237
18	1-7	0.615
19	2-7	0.604
20	6-7	1.450

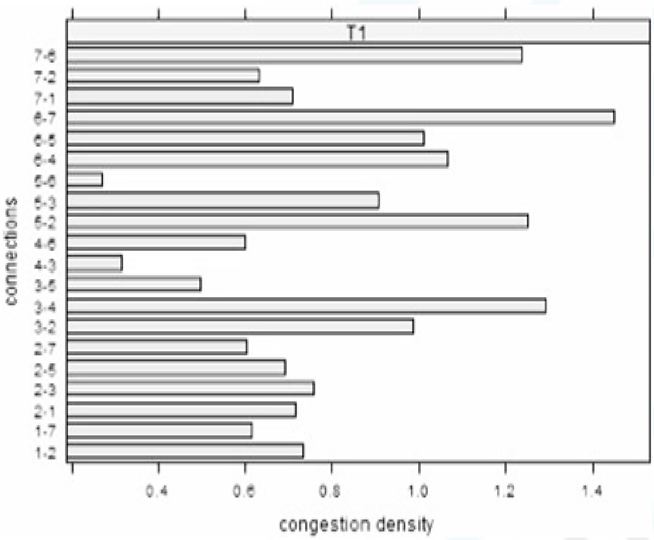

Figure 2: Areal density along each connection at time frame T1

Next, each connection is extracted separately and the variation of traffic congestion during the whole three-hour period is obtained. At the initial stage, graphical representation of variation of traffic congestion along each connection is thoroughly observed. 20 scatter plots for each connection is secured. Transition matrix is built based on this extracted data. Following is a scatter plot representing the variation of traffic congestion along a connection.

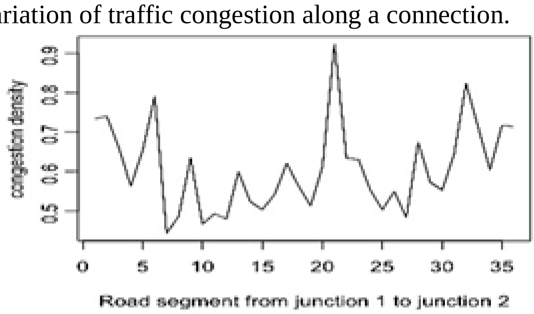

Figure 3: Areal density variation along the connection from junction 1 to junction 2

Next, transition matrices for each road segment are procured using the cut points obtained. At this stage, there would be 20 transition matrices, one for each connection of the road network under study. Following are a small sample of transition matrices.

Transition matrix of road segment connecting junction 2 and junction 1:

$$\begin{bmatrix} 0.857 & 0.143 & 0.000 & 0.000 \\ 0.571 & 0.429 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 \end{bmatrix}$$

The transition matrices demonstrate the conditional probability. In the above example, the segment that is being chosen is the segment between junction 1 and junction 2. Since the segment is two ways, a separate matrix for vehicles travelling from junction 1 to junction 2 and a separate matrix for vehicles travelling from 2 to 1 is obtained. It should be noted that, the rows, which are not having any positive transition probability, it indicates that the traffic congestion didn't reach those states. Thus, there were no transitions from that par particular state to any other state. Thus, in the transition matrix of connection 1-2, the traffic congestion didn't reach state 3 and 4. Thus, the row sums corresponding to rows 3 and 4 do not have any positive transition probability. However, this doesn't violate any effects in further calculations. This is clearly explained in the next chapter.

They depict the probability of the particular road segment being in a congestion state given its previous congestion state. This is in par with the Markovian property. Once the transition states have been obtained, we take the experiment another step forward. Here, we attempt to prognosticate the probabilities of a particular road segment be in different congestion states. Thus, if the initial probabilities of a road segment being in any of the four states is given, we triumph to determine the probability distribution of the congestion states of that given road segment.

Prior to obtaining the transition matrices, the traffic congestion states are defined. Under the simulation study in consideration, transition states are determined using the scatter plots plotted previously. The range of variation of areal density is recognized and thereby probable classification of transition states are firmed about using logical reasoning.

The optimum number of states can be statistically analyzed through clustering. However, under the simulation study, we adhere to a more simplified methodology. We obtain the range of Areal densities and simply obtain four sub equally sized ranges. Clustering is a statistical technique that is highly sensitive to the data set. Since we carry out a simulated study with simulated data, clustering is not used.

Next, transition matrices for each road segment are procured using the cut points obtained. At this stage, there would be 20 transition matrices, one for each connection of the road network under study. Following are a small sample of

ISSN (Online): 2347-3878
Index Copernicus Value (2015): 56.67 | Impact Factor (2017): 5.156

transition matrices.

4. Application and Results

Like most of other countries, the problem of traffic congestion has assumed to reach alarming proportions particularly in major cities of Sri Lanka. As a matter of fact, Kandy city traffic has recently become a major issue, which still remains unanswered entirely. The ever-growing number of motorists and inadequate road infrastructure has led to frequent traffic jams, queuing problem and environmental degradation. There seems to be a significant escalation in the travelling times during the peak hours due to heavy traffic. The city limits of Kandy city get overly congested during morning peak hours and evening peak hours. As it was clearly elucidated in the introduction chapter of the research study, there are many adverse situations that have surfaced due to heavy traffic. These adverse effects can be clearly observed in city of Kandy too. Thus, it signifies the need to study the traffic scenario in Kandy city. The previously proposed methodology is applied to the traffic scenario of Kandy city, and results are discussed under this chapter. However, due to lack of availability of sufficient amount of data, validation of the model is not discussed under this chapter. The future improvements of the model include validation of the model using sufficient amount of data.

Under this chapter, we aim to apply the method introduced under the study to a real world traffic scenario in Kandy city on a regular weekday and discuss the results. The study focuses on the Kandy city limits.

4.1. Introduction to the data set

There are four main entrances to the Kandy city, Sri Lanka. William Gopallawa road, Katukelle road, Katugasthota road via Mahaiyawa and Boowelikada. Under the study we consider a closed road network comprising of three of the main four entrances namely, William Gopallawa road, Katukelle Road and Katugasthota via Mahaiyawa. The date chosen is 26th July 2007.

The primarily obtained data includes counts of five vehicles types. The types of vehicles are three-wheelers, motorcycles, cars and jeeps, busses and lorries and vans. It should be noted that the classification of the types of the vehicle is based on the approximate size of the vehicle. The data belonging to the counts of vehicles that travelled along the defined road segments either ways are collected. The time period through which the data are collected is from 6.00 am to 9.00 am. There are assumed to be the peak time periods under the study. As the initial step, graphical representation of the road network is clearly illustrated.

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font).

While carrying out the study, roads leading to Kandy city from Hanthana and Ampitiya are ignored, as the vehicles coming from those directions have no alternative routes they could use and have only the main roads they always use. Hence those two routes leading to Kandy city do not affect the final answer that is to be obtained using the above method mentioned.

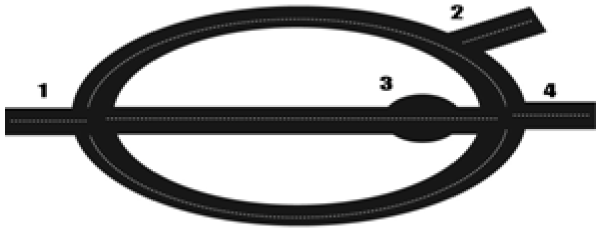

Figure 4: Graphical Representation of Road network

Junction 1 to junction 3: Katukelle Road
Junction 1 to Junction 4: William Gopallawa Mawatha
Junction 1 to Junction 2: Gohagoda -Katugathota Road
Junction 2 to Junction 4: Katugasthota - Kandy Road

The graphical representation illustrated above is a simplified version of the Kandy city road network. Here, junction 1 represents Getambe junction. Junction 2 represents Katugasthota town, while junction 3 and 4 collectively represents Kandy city. Both roads William Gopallawa Mawatha and Katukelle Road start from the same junction and end from the same junction. This causes a complication when inserting data into the matrices of the developed model. To overcome such a complication we introduce a dummy junction (junction 3). It should be noted that, implications of introducing such a dummy junction doesn't violate the concept of the model. Both junctions are theoretically the same.

The data belonging to Gohagoda – Katugasthota road is not considered under the study. The particular road segment doesn't get congested in reality. Thus, the significance of studying the particular road segment is immaterial. Our main focus is the rest of the three road segments that in reality are the core areas of congestion in the Kandy city. Their analysis is very much material.

Adjacency Matrix of the road network in consideration:

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

Since there are five types of vehicles as mentioned above, five separate sets of matrices for each time interval of five minutes is obtained. Each vehicle type is given a code, for convenience of analysis procedure. The code for each vehicle type is shown below.

Table 2: Vehicle types

Code	Vehicle Type
1	Three-wheelers
2	Cars/Jeeps
3	Van
4	Bus/Lorry
5	Motor Cycle

Thus, data representations in matrices are obtained as shown below.

The counts of three-wheelers that have travelled during T1:

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{bmatrix}$$

The counts of cars and jeeps that have travelled during T1:

$$\begin{bmatrix} 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix}$$

Similarly, data belonging to all the 36 of five-minute time intervals ranging from 6.00 am to 9.00 am are stored in these matrices. Next is to determine the road occupancy.

4.2. Determination of road occupancy

An assumption is being made the number of lanes per a headway is a constant. Thus, the effect if the Line matrix (L) is even. Hence, it is not specifically considered here for the simplicity of calculations. Road surface area matrix (R) is the next to be obtained. It should be noted that the Road Surface Area (R) matrix obtained under the study is a robust measure. However, the accuracy of the Road surface area of the road network has to be studied more closely at the time when validation of the model takes place. Validation of the model introduced can be done using any road network, thus, the road surface area of the road network considered for validation procedure has to be calculated with accuracy. Other factors, which can affect the road surface area too can be studied. Discussions regarding other factors to consider when calculating the surface area for a more accurate value are presented in conclusion.

Road Surface Area Matrix for the road network considered is provided below. Unit of measurement is square meters (m^2). The Kilometer value is rounded off in to the nearest 100. It should be notes that approximate values are being used

Table 3: Surface area of Road segments

Road Segment		Road type	Lane	Road	Effective
Start	End	(Standard)	width (m)	Distance (m)	Surface Area
1	4	A	4	4200	16800
4	1	A	4	4200	16800
1	3	A	4	4300	17200
3	1	A	4	4300	17200
2	4	A	4	4500	18000
4	2	A	4	4500	18000

The matrix depicting the surface area of road segments of the road network

$$\begin{bmatrix} 0 & 0 & 17200 & 16800 \\ 0 & 0 & 0 & 18000 \\ 17200 & 0 & 0 & 0 \\ 16800 & 18000 & 0 & 0 \end{bmatrix}$$

Thereafter the study is taken to the next level. At this level, sum of surface areas of all the vehicles during a five- minute interval is calculated. This results in 36 matrices, comprising of the total surface area occupied by vehicles along each segment. The total surface area occupied by vehicles during each time interval is calculated. The total surface area occupied by the vehicles during first two time intervals is shown below.

Surface area occupied by vehicles during Time frame:

$$\begin{bmatrix} 0.00 & 0.00 & 2974.26 & 2732.75 \\ 0.00 & 0.00 & 0.00 & 3364.29 \\ 1140.01 & 0.00 & 0.00 & 0.00 \\ 1870.96 & 1946 & 0.00 & 0.00 \end{bmatrix}$$

Once the set of matrices containing the sum of the surface areas of the vehicles are obtained, it is represented as a ratio of the road capacity area. Here after, the measure is known areal density. The density matrices depict the road occupancy by vehicles during a given time interval. This, step is very crucial to our study, since our Markov Model is built around this variable. It is noteworthy that since it is still at the simulation level, Surface area of the road is arbitrarily taken.

Areal density during T2 among junctions:

$$\begin{bmatrix} 0.00000000 & 0.0000000 & 0.1729221 & 0.1626637 \\ 0.00000000 & 0.0000000 & 0.1729221 & 0.1626637 \\ 0.00000000 & 0.0000000 & 0.1729221 & 0.1626637 \\ 0.11136667 & 0.1081111 & 0.0000000 & 0.0000000 \end{bmatrix}$$

At the proceeding step, all the data contained in the matrices are extracted to obtain the density fluctuations along a particular road segment throughout the observational time period. Here, each matrix is converted to a data frame initially. The data frame represents the connection with the direction, the density of respective connection and the time frame to which the data belongs. Given below is data frame constructed through extraction of data for the time frame T1. At the end of the extraction process, there is expected to be 36 such data frames, for each five-minute interval.

According to the methodology, the next step is the representation of variation of traffic congestion state. A graphical representation similar to the one shown in the simulation study is obtained. It is shown below.

Table 4: Areal density of the road network at time frame T1

	Connections	Time frame	Areal Density
1	3-1	T1	0.1678
2	4-1	T1	0.16781
3	4-2	T1	0.2644
4	1-3	T1	0.2644
5	1-4	T1	0.2644
6	2-4	T1	0.0546

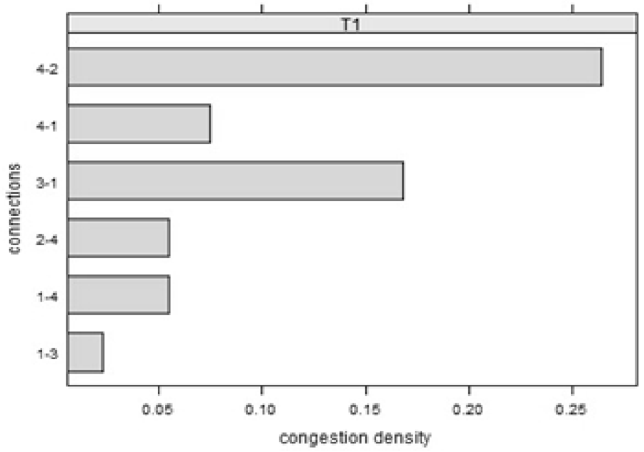

Figure 5: Areal density along each connection at time frame T1

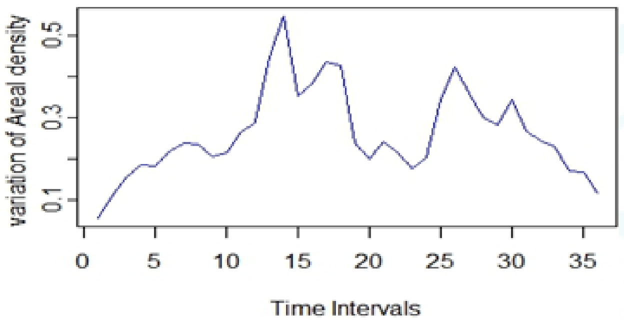

Figure 6: Areal density variation along the connection from junction 1 to junction 4

Next, as explained in the methodology each connection is extracted separately and the variation of traffic congestion during the whole three hour period is obtained. At the initial stage, graphical representation of variation of traffic congestion along each connection is thoroughly observed. 6 scatter plots for each connection is secured. Transition matrix is built based on this extracted data.

Next, transition matrices for each road segment are procured using the cut points obtained. At this stage, there would be 6 transition matrices, one for each connection of the road network under study. It should be significantly noted that definition of traffic states are user defined.

This study uses cut points for definition of states as below
Areal densities equal or below 0.0662790 indicate state 1
Areal densities equal or between 0. 111366indicate state 2
Areal densities equal or between 0.250418indicate state 3
Areal densities above 0.250418 indicate state 4

Following are the set of transition matrix obtained for connection 3 to 1.

$$\begin{bmatrix} NaN & NaN & NaN & NaN \\ NaN & NaN & NaN & NaN \\ 0.000 & 0.0000 & 0.6363636 & 0.3636364 \\ 0.000 & 0.0000 & 0.1666667 & 0.8333333 \end{bmatrix}$$

The transition matrices demonstrate the conditional probability. In the above example, the segment that is being chosen is the segment between junction 1 and junction 2. Since the segment is two ways, a separate matrix for vehicles travelling from junction 1 to junction 2 and a separate matrix for vehicles travelling from 2 to 1 is obtained.

As explained in chapter three, the transition matrices depict the probability of the particular road segment being in a congestion state given its previous congestion state. This is in par with the Markovian property. Once the transition states have been obtained, we take the experiment another step forward. Here, we attempt to prognosticate the probabilities of a particular road segment be in different congestion states. Thus, if the initial probabilities of a road segment being in any of the four states is given, we triumph to determine the probability distribution of the congestion states of that given road segment.

Suppose we consider connection from 1 to 4. This indicates a reference being made to the fifth transition matrix under the study. Transition matrix [, , 5]. Here, we check for the rows containing "NaN" values first, and then the altered matrix by dropping the rows containing "NaN" is obtained. As explained later , removing rows containing "NaN" values does not violate accuracy of prognostication.

The transition matrix after adjustments:

$$\begin{bmatrix} 0.000 & 0.000 & 1.0000000 & 0.000 \\ 0.000 & 0.000 & 0.8947368 & 0.1052632 \\ 0.000 & 0.000 & 0.1333333 & 0.8666667 \end{bmatrix}$$

Next, a matrix defining the probabilities of the initial conditions of the segment is declared.Matrix declaring the initial probabilities of traffic state conditions being 1, 3 and 4 respectively of the segment:

$$\begin{pmatrix} 0.3 & 0.5 & 0.2 \end{pmatrix}$$

Next, the matrix representing the probabilities of each traffic state being the traffic state (1, 2, 3, and 4) of the road segment in the next 5-minute interval is obtained. This is done through the matrix multiplication.

$$\begin{pmatrix} 0.0000000 & 0.0000000 & 0.77403551 & 0.2259649 \end{pmatrix}$$

The probabilities of each traffic state being the next traffic state of the particular road segment is obtained. The row sum is equal to 1 and the Markovian property is conserved. However, the probability of the particular segment being in state 2 was anyway assumed to be 0 in the calculation. Thus, we can declare the probability matrix of traffic state conditions being 1, 3 and 4 respectively of the segment as below:

$$\begin{pmatrix} 0.00000000 & 0.77403551 & 0.2259649 \end{pmatrix}$$

It should be noted that the values "NaN" represents that there hasn't been any transition from the current state to the next state during the time period considered. However, these values do not violate the concept behind the model. When assigning initial state of traffic condition, it should be noted that probabilities should not be assigned to the states having "NaN" values. Because the model assumes that if there weren't any transitions from state 'x' to any other state, it implies that the segment didn't ever reach state 'x' during the time period considered. Thus, there is no point of assigning a non.

5. Conclusion

The main objective of the study was to introduce an effective traffic analysis model to be presented as a solution to model traffic in a road network system. Under the literature reviews thoroughly studies, it was found out that stochastic models were more preferable than deterministic models due to the capability of stochastic models in dealing with the inherent probabilistic nature of traffic. A crucial observation made about almost all the models that were put forth was that, they had shortcomings when it came to considering the road network at its entirety.

The model presented under this study addresses to the above mentioned issue. It is able to consider a whole road network, including junctions at once. Further, the model addresses to the effect of the size of the vehicle to the traffic congestion.

Yet another significance of the model is that it can predict the probabilities of a state being in a given traffic congestion state given the initial state. This option is very useful. Since the traffic conditions may vary from one day to another, prognosticating traffic condition based on a given initial traffic condition is very much convenient. The model presented under the study succeeds in calculating the probabilities of a particular road segment being in a given traffic congestion state. Due to this model being a simple one, when improved, it can be utilized even by the general public to analyze the traffic congestion in their local road network. They can first chose the route they have to take and then check the traffic states in a future time interval with the use of transition matrices established. There after they can decide, the time at which they will have to leave home to reach their respective destinations facing a minimum traffic. This is advantageous to the general public in many aspects. A major advantage is the time factor. They can save their valuable time by leaving the house at the best possible time. Further, this can reduce the fuel cost since they can avoid large waiting times at traffic congestion. Moreover, reduction in sound pollution, reduction in accidents due to frustrated stuck up drivers can also be considered as advantages to the general public though they are in effect indirectly.

Another significant advantage is that, the model, when improved further can be used by traffic control units as well. They can observe the traffic congestion variation using the methodology put forth under the study and recognize which segments would be highly congested in the near future and then take precautions. They can either deploy traffic control units to the nearby areas of high congestion or else, provide a solution of by roads to specific types of vehicles.

Since the model allows to study about the effect of different types of vehicles to the traffic congestion, the traffic control units can analyze how traffic congestion changes with different counts of different vehicle types. This is very useful in developing solutions of by roads to the city. As an example, since the busses cannot change their routes, the counts of busses can be kept constant while, counts of cars, vehicles can be changed to check for the optimal state of traffic congestion.

Thus is can be clearly stated that the model introduced under the study, is proven to have fruitful results as a productive model.

However, the model put forth under the study still needs to accurately validated using an adequate data set. The model validation section couldn’t be carried out smoothly due to the unavailability of a real world data set that fits for all the requirements demanded by the designed experiment. Once the data are collected for a considerable period of time, at least for three months, then it would definitely bring about a more comprehensive overlook of the traffic congestion analysis. Due to the availability of data for all seven days of the week, more meaningful patterns could be derived. This could also result in several transition matrices for the same road segment depending on the day of the week. This could increase the accuracy of the prognostication process dramatically.

Further, in calculating the road surface area, it has to be noted that a standard and an approximate value has been obtained. It has been assumed that the width of the road stays constant although out the experiment. However, the model can be improved by improving the accuracy of this variable. Other factors that affect the road surface area such as vehicles parked on either sides of the roads, the areas under constructions can also be included. However, a noteworthy point is that, due to the flexibility of the Markov model introduced under the study, all the factors can easily be incorporated to the model. Thus, once the initial graphical representation of the road network in consideration is obtained and initial counts of different types of vehicles are stored, the model can be continuously improved. Types of vehicles under the initial study were four. And the real world scenario was five. This shows that, the model can be implemented for any number of types of vehicles. The unrestricted and flexible nature of the model signifies the importance and the specialty of the model over the other models introduced.

References

[1] Chen X, Li L.(2010), A Markov Model for Headway/Spacing Distribution of Road, Traffic, IEEE transactions on intelligent transportation systems, vol. 11, No. 4, 773-785
[2] Zhang B., Li B., Liu D. and Zhang S.(2011), Application of Short-term Traffic Flow Forecast in Selecting the Optimal Route, 2011 Fourth International Conference on Intelligent Computation Technology and Automation, 769-77.
[3] Antoniou C., Koutsopoulos H.N. and Yannis G. (2007), Traffic state prediction using Markov chain models, Proceedings of the European Control Conference 2007, 2428-2435.
[4] Qi Y. (2010), Probabilistic Models for Short Term Traffic Conditions Prediction, Louisiana State Univ.
[5] Metroul S.A., Intermodality (2015), A Solution To The Problems Of Large Urban Agglomerations.
[6] Dodge, Yadolah (2006). The Oxford Dictionary of Statistical Terms. Oxford, England: Oxford University

Press. p. 335.
[7] Lawler, G. (2006). Introduction to Stochastic processes (2 ed.). CRC Press. p. 1. “A stochastic process is a random process evolving with time. More precisely, a stochastic process is a collection of random variables indexed by time.”
[8] Thiele. T.N. and Steffen L. Lauritzen (1880), Time Series Analysis in 1880: A Discussion of Contributions Made, International Statistical Review, Vol. 49, No. 3 (Dec., 1981), pp. 319-331
[9] Norris. J.R. (1997), Markov Chains, Cambridge University Press.
[10] Reiter J. (2015), Application of a Markov Chain Traffic Model to the Greater Philadelphia Region, Concept: An interdisciplinary journal of graduate studies, vol. 38.