Markov Chain Modelling and GIS Interface to Analyze Traffic Congestion

Madduri Swathikiran, R. Archana Reddy, G. Sahithi, N. Prabhanjan, G. Swamy Yadav, D. Swetha

Abstract: Traffic congestion is the major problem that major metropolitan cities are facing. From the study it is observed that the main cause of the delay is because of not following the actual design of the transportation rules of the particular area. There are many factors which are influencing the congestion are taken in the study like road density, road geometrics and traffic composition. Data is collected at Gochibowli, Kothagudem and Cyber towers intersection regarding road geometrics, road density and volume count using ArcGIS and manually. The Markov chain is a mathematical model is used. Markov chain is a transition probabilistic model which can be used in any stream to find out percentage of chances for an event to take place. Empirical model is developed for delay prediction by considering the values from data collection. Penetrating the values into the transition probabilistic method and linear equations to find out the factors influencing delay and calculating the delay errors.

Keywords: Traffic congestion, Markov chain analysis, Transition probability, Delay studies.

I. INTRODUCTION

Traffic congestion occurs due to many factors, which have several consequences including economic loss, air pollution and traffic delay. Congestion can be called as a result of improper land use, growth of vehicle volume, limited capacity of roads, undeveloped infrastructure and many more external factors (Ren G, et al 2014). For the improvement in level of service, improved infrastructure of road, management of traffic and planned strategies, and live traveler information systems with ITS communication technologies can be used where traffic flow rapidly increases and results in congestion (Andrews 2002). Centerlines with topology, attributes, and functional class can be drawn by the a roadway spatial database. For most of the big cities traffic congestion is an unsolvable issue. With the increase in the privacy car urban road transportation has a rapid growth and many road intersections approach to saturated limit (Sun Ye 2012). Several studies made many conclusions about traffic congestion such as, it occurs due to insufficient area, less time and mostly in a single directed roadways especially at peak hour. In some metropolitan areas congestion happens throughout the day. (Orski, C.K 1987)

Humans need to access many activities in their daily life. This also includes mobility from one place to the other. And transportation plays a major role to access all the activities (Arasan 2012)

Urban transportation provides mobility and access to the goods and people within the cities. It includes non-motorized transport such as pedestrians, public transit and cyclists that is collective transport. Urban transportation systems are important to have high quality of life, increase the economic activity and opens up different opportunities to access essential services including social activities which improves the quality of living of people. (Arasan and Rodrigue 2009)

Urban transportation is the key factor for development of the urban area. This includes good traffic management, best transport network, monitored control system and most importantly effective and efficient mass transit (Eddington R. 2006).

Congestion of traffic may occur due to many factors. Sometimes the behavior of the road user is an important factor to be considered. Apart from that on road networks if there is any use of road infrastructure more than the capacity of the road. The excessive infrastructure may result in the effect like low speeds of the vehicles, increase of travel time and the queuing of vehicles which indirectly results to increase of waiting time of the vehicles (Rodrigue and J.P. 2009).

Traffic congestion occurs when the road capacity reaches the saturation point. It happens when the volume of vehicles is more than the capacity of the road. These situations occur when the traffic is reduced for a particular length of road or increased vehicular number for the movement of people, goods and services (Downie 2008).

When road capacity doesn’t match the vehicle demand then delays in travel time occurs which increases traffic congestion (Haribandhu Panda and RS Pandir 2002).

Among many factors of congestion environmental conditions are also the major one. In rainy season we have more congestion when compared to other season in Hyderabad. Rainy season result in failure of under drainage systems, moreover it also results in accumulation of excess runoff water on roads which may considered as the factor for obstruction of traffic flow. (B. Suresh et al. 2018 Amin 2013)

The estimated population in Hyderabad is 11.572 million. (as per 2018 census).

Deformations, failures in Pavements, corrugations, depressions, rutting and shaving etc leads to traffic congestion. Deformation may generally occur in both flexible and rigid pavements. (N Venkat Rao and B Suresh 2018)
Markov Chain Modelling and GIS Interface to Analyze Traffic Congestion

Street vendors create a lot of problem to the pedestrians. Foot paths are provided through out the road lengths, but these are occupied by street vendors which leaves no place for pedestrian to use the foot paths. In such conditions pedestrians tend to use roads to walk indirectly this tends to problems to the vehicles as low speeds, congestions and delays (Shulin He 2012).

Traffic jams also occur due to bottle necks. The bottle necks may be due to heavy traffic trying to exit from a small entrance. Or this maybe due to heavy vehicles trying pass through narrow openings (William Zhang and Wen Long YUE, 2001).

Huge traffic is stuck for longer duration on roads when the passage is narrow for the vehicles to cross (Mr. William Zhang and Dr. Wen Long YUE, 2001)

Lane driving is the best solution to avoid unnecessary traffic jams and obstructions. Many strict rules are to be implemented to force the drivers to follow the lane management. Here different classifications are given based on the factors as vehicle type. Lane driving successfully minimizes traffic congestion (Robakasunshmer and Mohammad nayeem Abdullah 2015).

Another main important characteristic of the traffic congestion is the absence of lane-discipline which results the disturbance in the traffic. In the mixed traffic condition or in homogeneous traffic, the slow moving non motorized vehicles interrupt the speed of fast moving vehicles. (Saikiran and Ashish 2016).

Rapid urbanization in cities has resulted in more increase in urban infrastructure requirements. Impact and Intensity has been significant on traffic and urban travel. Along the growth in population rising traffic, the trip rates on road network and vehicular delay and traffic congestion etc. are the normal phenomenon in our metropolitan cities. Congestion is one of the major topic for traffic engineers and travelers. (Aathira K Das et la. 2013).

Digitization data are under consideration, using software’s and methods which are used to map and update the highway maps and centerline maps. There are many influencing factors which are responsible for the uncertainties in the calculation of length of the road ways (Noronha et al 2000).

Stochastic models help to explain a sequence of possible events. Markov chain methodology is used to find the probability of each event, the sequence of events always depend on the state attained at the previous event (GagniucPaul, 2017).

Reversible jump Marco Chain Monte Carlo frame is used to develop random frog algorithm. Consider a design matrix X, of size n x p. Here the matrix consists of n samples in rows and p variables in columns. Matrix X, the class label vector of size n x 1, with the elements equal to 1 or -1 in binary classification (Hong Dong li, 2012).

Random frog model is developed in the process of Markov chain, for road density, geometric and traffic composition. With all these components it becomes a complicated stochastic process. Stochastic processes are good compared to Deterministic processes.

Markov Chain Model has two versions they are discrete time version and continuous time version in which the discrete time version is considered in this study. It explains about the evolution of process through a sequence of state s which is having equal time interval. such as S(0) --> S(1) --> S(2) -->…… S(t), where S(t) denotes the state of the system at time t. S(t) controls S(t+1) through transition probability pij, where pij given by pij=P(S(t+1) = j | S(t) = i). Transition probabilities can be explained in a neatly compacted way by arranging them in a square matrix P known as transition matrix, which looks like:

$$P = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix}$$

In Markov chain model, the prediction of probable future distribution is done. It is generally used for the four zone Markov chain model. Of few applications include the socio-economic, land use geographic and transportation research and then a transition matrix is estimated by the movement of firms from one zone to other (Lever W, F 1972)

The analysis further involves rising of powers to the initial probability matrix. After some iteration the values in the matrix will be stabilized.

Firstly we need to define the variables of interest to identify the model. Traffic composition of different places in the study area is exposed with the road network characteristics as road length, road density and geometrics of road as width of the carriageway and width of the shoulder. Here this is done to determine a relationship between the components delay and explanatory variables.

Delay is considered as the independent variable (Y) and the other components are independent variables represented as X0, X1 and X2. These independent variables are road geometrics, road density and traffic composition. Coefficients of the variables measure directly or indirectly marginal effects of the independent variables on delay in the study area. The most general form for a model is:

$$Y = f (X)$$

Where,

- Y measures delay in secs, (dependent);
- f specified function;
- X is explanatory variable;

In the specific form,

$$Y = a_0 X_0 + a_1 X_1 + a_2 X_2$$

Where,

- Y is dependent variable (delay in seconds);
- X1, X2, and X3 are independent variables i.e., road density per one square kilometer, road geometrics of road (width of the carriageway, shoulder width, median) and traffic composition (volume counts) respectively; a1, a2, and a3 are the coefficients which determine the contribution of the independent variables.

The basic data obtained from the observations are analyzed for different values of road density, road geometrics, and volume counts. Markov chain method is used in the analysis. From the methodology the iterations values are obtained and these values are substituted in the linear equation to get the delay.

II. OBJECTIVES OF STUDY

Develop an empirical model for delay prediction by considering various influencing factors like road density, Road geometrics, and traffic compositions Identifying...
various factors leading to delay. To determine the impact of various factors like road geometrics, network which are congestion. Identifying the delay errors. With the help of results suggestions and recommendations are summarized.

III. METHODOLOGY

The following steps are involved in the present study.
- Selection of study area
- Data collection (vehicle counts, road widths etc.)
- Road geometrics and volume studies
- Delay studies using Markov chain
- Validation

Study area:
Hi-tech city, hub for the leading Information Technology (IT), Bioinformatics, Health sector and Engineering is situated in Hyderabad. There are more than 400 IT companies in the city and 90% are located in Hitech city, Madhapur, Kothagudem and Gochibowli. Around 4 lakh IT employee travel through this IT corridor of the city. From the last few years there has been a tremendous and rapid growth in commercial standard which resulted in many negative impacts on transportation systems. Some of them are congestion, delays and accident risks, for this necessary steps are to done to the factors influencing delay. The three intersections where there is heavy traffic congestion are considered. They are cyber towers junction, Kothagudam junction and Gochbowli junction.

IV. DATA COLLECTION

Digitization
In this process data from other sources are converted to digital format in GIS. This data may be maps, images or other sources of data. Digitization is done by using the tools for road line. Road map is used from the study area. First the spatial data is collected such as road network, road maps where the base map is included (Mohammad Abousaeidi 2015). Network tracking explains about a path through the network. For the development of the transportation applications, the help is taken from the ArcGIS model data (Satish Kumar 2017). The digitization is done to find out the length of the road.

Graph 4 shows the road density values per km2. Digitization is done using ArcGIS to calculate the length of the road. The estimated growth of road density is considered and then compared to intersections at cyber towers, kothagudam and Gochbowli. Through the observation the values obtained are less when compared to actual road density value which is one of specific reasons the occurrence of traffic congestion.

Length (portion or percentage) of the road network in the metropolitan cities plays a significant role in traffic congestion. Digitization is done using ArcGIS to calculate the length of the road. Table shows the road density values when compared to the estimated value.
Markov Chain Modelling and GIS Interface to Analyze Traffic Congestion

Classification of Urban roads

Roads are classified based on their purpose and are constructed for the future use and to reach the future expectation of new trends. The strength characteristics are important to improve the life of the roads (Venkat Reddy 2017). Pavement strength is directly dependent on the materials used. Some recent researches are done to improve the strength of the materials used in the construction of roads using different admixtures (Shaik 2017, Ravindran 2016, Venkat Reddy 2018). Urban roads are classified into four main categories (IRC 86-1983). This is done for the purpose of geometric design.

(i) Arterial
(ii) Sub-arterial
(iii) Collector-Street
(iv) Local Street

Design speed

Some specific speed is to be taken into consideration for designing a road and this speed is taken as design speed. Based on the road type some design speeds are already recommended by IRC.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Design speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arterial roads</td>
<td>80 km</td>
</tr>
<tr>
<td>Sub arterial roads</td>
<td>60 km</td>
</tr>
<tr>
<td>Collector streets</td>
<td>50 km</td>
</tr>
<tr>
<td>Local streets</td>
<td>30 km</td>
</tr>
</tbody>
</table>

Table 1: Design speeds of roads as per IRC 86-1983

Road geometrics:

Geometrics design deals with the elements which can be seen. Considering and using proper geometric standards facilitates economical and safe operation of vehicles. Geometrics design is influenced by a number of factors among which nature of terrain, operation speed, composition and volume of traffic, aesthetics and land use characteristics are important. (IRC)

Figure 5 shows the Road geometrics of the road obtained from the field are composed of carriage width; median and side walk (IRC 86-1983). The actual values are taken and compared to the field values. The field values are less when compared to the IRC values which is the reason for the traffic congestion. The side walk is less than the people are intended to walk on the road which leads to the increase in delay and accidents.

Table 2: Recommended Land Widths for Roads in Urban Areas as per IRC 86-1983

<table>
<thead>
<tr>
<th>Classification</th>
<th>Recommended land width in meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arterial roads</td>
<td>50-60</td>
</tr>
<tr>
<td>Sub Arterial roads</td>
<td>30-40</td>
</tr>
<tr>
<td>Collector streets</td>
<td>20-30</td>
</tr>
<tr>
<td>Local streets</td>
<td>10-20</td>
</tr>
</tbody>
</table>

Right of way

The Right of way recommended for the various categories of urban roads are given in Table 2

Using traffic flow parameters and knowledge of normal ranges of vehicle behaviors, traffic engineers analyze and evaluate a proper plan to improve traffic facilities (Jain. et.al. 2014).

Volume count:

Volume count studies at the intersections of cyber Cyber towers, Kothagudem and Gochibowli intersections. Number of vehicles are counted and then converted into passenger car unit (PCU)

Volume count: volume count studies are done at Cyber towers, Kothagudem and Gochibowli intersections for one peak hour. (IRC 106-1990).

Traffic Composition

Graph 6 showing the total no of vehicles in terms of PCUs is collected at the intersection for three intersections. For different categories of urban roads, the design service volumes which are considered in pcu’s are given in IRC 106-1990. Comparing with the IRC the obtained value is many times more than actual which is the main reason the traffic congestion. As per the clash in between the road capacity limit and traffic demand there exist severe traffic congestion.
Over time probability:
Graph 7 shows the factors road density, road geometrics and traffic composition that cause the traffic congestion which are observed in the study are added in a place to get the over time probability.

Delay Studies:
Travel time is considered as a fundamental measure in transportation analysis. The time taken by any user to reach the required destination from their origin in a specific route in all the extreme traffic conditions is considered as travel time. Here delay can be stated as the time taken more than the stipulated duration in that specified route (Reigna Jewel Ritz et al, 2011).

Delays may also be understood as the actual time expected to reach the destination and the experienced travel time on that particular situation. The experienced travel time may be more due to unexpected incidents like accidents; overlapping the design capacity, improper signal controls etc. Delay may occur at anywhere of the section i.e., either at midblock of the section or at starting or ending of that particular section. Congestion and delay are indicating that demand that is greater than the available capacity of the road network. At intersections, due to the lack of lane discipline there is significant lateral movement of vehicles and vehicles tend to use in between gaps to reach a bit further and to stand in front of the queue. Under such conditions, the delay models including the classical delay formula suggested by Webster (1958) which were developed under homogeneous traffic conditions will not produce realistic estimate if directly applied to heterogeneous traffic conditions. Consequently, there arises a need to develop a model for the realistic estimation of delay under heterogeneous traffic conditions.

Linear equation
In general since 1970 there is a lot research attention has been paid to algebra particularly in equation solving. Many researches are done in solving linear equations; this is always based on wide range of techniques and different perspectives of the researcher. Firstly we need to create a equation which has some theoretical structure is required to organize all the variable pieces for correct analysis. Variable in this equation means the values which are to be found. We need to find the unknown variables from the known ones. Secondly one needs to understand the equations having the variables and the relationship between the variables (John E. Bernard 2006.) After creating all the needed calculations in the Markov chain process for each parameter i.e., traffic characteristics, road density, road geometrics. Based on our objectives, a linear model created and different variables defined and the data was consequently measured. From the Markov chain process values they are given into linear equation modeling to find out the delays that are taking place in each intersection, Whether the functionality of the road is good or not. The general form of the model is:

\[ Y = a_0X_0 + a_1X_1 + a_2X_2 + a_3X_3 \]

Where \( Y \) is the Delay, \( X_0 \) is road density, \( X_1 \) is the road geometrics, \( X_2 \) is the traffic composition.

**Table 3: Observed delays and Overtime Probabilities.**

<table>
<thead>
<tr>
<th>Y(Delay) Observed</th>
<th>Road Density X0</th>
<th>Road Geometrics X1</th>
<th>Traffic Characteristics X2</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>150</td>
<td>0.44</td>
<td>0.22</td>
<td>0.33</td>
</tr>
<tr>
<td>180</td>
<td>0.29</td>
<td>0.49</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Markov Chain Modelling and GIS Interface to Analyze Traffic Congestion

Validation:
Deriving the data which predicts another variable is widely general. Deriving statistic models is to predict data of one variable from one or many variables or it can be called as a predictive modeling. It is important and necessary to determine the quality of the model and then we need to quantify the values and then report the predictive validity of the derived models.

quantify and report the predictive validity of the models derived. Validating the predictive models provides the important information to the researchers about the specific model. In many cases the validation predicted can be reduced nearly to zero. The estimation of predictive validity of models and the methods describing to estimate the validity of prediction which is called as cross validation can be seen here. Based on the methods and the accuracy of the expected outcome it should all the time be reported to the new models which are predictive. R2 estimates the total variance in the outcome measurements that is explained by the prediction model. When this is expressed as a ratio, then R2 ranges from 0 to 1. For a predictive regression model when R2 is closer to 1 then it is considered as the better prediction (Andrada E. Ivanescu, et al 2016).

To estimate to the power of prediction of a Quantitative structure–activity relationship models regression models are used. use of statistical characteristics and the tests, we need to follow some steps, correlation coefficient of both the observed variables and the predicted activities is taken as R coefficients of determination (R2 ) R2 is taken as the slope of the regression line, through the origin. (Fadilah et al, 2018) International Journal of Drug Design and Discovery Volume 2 model is predictive, if the following conditions are satisfied14: R2 prediction > 0.6, r2 – 0.2 or /r2 < 0.1, r2 –.2 or / r2 < 0.1 and 0.85 < k < 1.15 or 0.85 < k’ < 1.15. The predictive ability of the selected model was also confirmed by external R2 prediction. A value of R2 prediction is greater than 0.6 may be taken as an indicator of good external predictability. (Ravichandran et al 2011)

For the model to be acceptable the difference between the observed and model values should be less than 15%. R-square is a statistical measure of how close the data are to the fitted linear model and it shows the goodness of fit. Generally, the higher the R2 value, the better the model fits the data. The difference between the observed value of dependent variable (Y) and predicted value (Y’) is called the residual (e) that it should behave normal. Table 5 represents the goodness of fit for both of the above mentioned models by calculating $R^2$.

V. RESULTS ANALYSIS

<table>
<thead>
<tr>
<th>Place</th>
<th>Delay Observed (Y)</th>
<th>Delay Calculated (Y’)</th>
<th>Residual (e)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyber towers</td>
<td>140</td>
<td>129.41</td>
<td>10.59</td>
<td>0.99</td>
</tr>
<tr>
<td>Kothaguda m</td>
<td>150</td>
<td>138.591</td>
<td>11.409</td>
<td></td>
</tr>
<tr>
<td>Gochibowli</td>
<td>180</td>
<td>165.181</td>
<td>14.819</td>
<td></td>
</tr>
</tbody>
</table>

VI. CONCLUSION

Modeling view has facilitated to identify the cause for traffic delay in dominance. It is observed that the prevailing road geometrics and road network characteristics are the real constraints for poor level of service. The road side activities can be controlled by well-planned system in a town. The street vendors can be moved to separate selling zones at every 1 or 2 km stretch. The vertical occupancy reduces the land use. The transport hubs should be interlinked so that no public could get down and come on to roads interchanging vehicles or routes.

REFERENCES


8. Sun Ye Department of economics, Shandong Jiaotong University Jinan, China Research on Urban Road Traffic Congestion Charging Based on Sustainable Development 2012.


21. Traffic congestion modelling under mixed traffic conditions through fuzzy logic approach: an Indian case study of arterial road Aathira K Das1, Krishna Saw2, B K Katti3


29. satish kumar july 2017 analysis of road network using remote sensing and gis data nainital district (uttarakhand)