

Estimation of Road Traffic Congestion using GPS Data

Hnin Thant Lwin¹, Thinn Thu Naing²

University of Computer Studies, Yangon, Myanmar¹

Professor, Software Department, University of Computer Studies, Yangon, Myanmar²

Abstract: Road traffic congestion is headache major problem in urban area of both developing and developed countries. In order to reduce this problem, traffic congestion states of road networks are estimated so that congested road can be avoided. In this paper, we estimate the real time traffic congestion states of user's desired source and destination and present the estimated results in Google Map. We use Hidden Markov model (HMM) for estimating the traffic condition states of these road network using both historical and real time data. To get these traffic data we use GPS trajectories data collected from mobile phones on vehicles. We evaluate our estimating system using dataset generated by collect data from phone-equipped vehicles over a period of 4 months in Yangon.

Keywords: traffic, HMM, GPS, estimation.

I. INTRODUCTION

Traffic congestion remains a major societal concern across the world. Therefore, the urban traffic management such as traffic congestion estimation systems has been conducted by researcher through processing specific sources of information such as public transport management and priority, traffic management in urban areas, and real-time traffic light signal management.

Traditionally, estimating traffic conditions has relied on slow and costly methods such as loop detectors, observations vehicles or automatic vehicle identification or floating car observers. Although dedicated moving observer or floating car vehicle-based methods can provide precise estimations, they require that an instructed driver collected the data needed. This is both time consuming and costly as the driver must be paid. This method also provides less data as a relatively small number of vehicles are usually used.

The increasing popularity of mobile phones embedded with positioning functionality such as GPS is allowing users to easily acquire their own locations which can be used for various purposes such as location-based applications. Moreover, traffic estimation systems based on GPS-enabled mobile phones are particularly suitable for developing countries, where there is a lack of resources for traffic monitoring infrastructure systems, and where the penetration rate of mobile phones in the population is rapidly increasing. Therefore, we use GPS data from mobile phone on vehicles to collect traffic data in this system.

Previous approaches to travel-time estimation include algorithms based solely on more or less educated guesses calculated from the permitted speed on a particular road segment, on finding weighted average given single observations, on data collected using expensive moving observer methods, or on the experience of traffic experts [3,2]. There are many approaches for estimation traffic in

urban areas such as data fusion, fuzzy control theory and microscopic traffic simulation, etc; but all these techniques need costly traffic data.

This system uses Hidden Markov Model (HMM) to estimate the real time traffic congestion states of user's desired source and destination. Then, the estimated traffic state results are presented on Google Map with colours on user's desired road network with traffic jam state for red, traffic heavy state for blue and traffic normal state for green. An application of this framework uses data from urban network in Yangon. This paper intend (1) to estimate traffic congestion condition of a route for vehicle driver to choose less congested path to the target destination (2) to estimate traffic condition using GPS data from mobile phone equipped vehicles for cheaper and real time traffic data (3) to get accurate map matching result with short time from ambiguous GPS data and (4) to calculate the traffic congestion states with data from both historical and recent dynamic traffic data and to send the accurate result to mobile users.

II. RELATED WORK

Several papers are concerned with prediction travel- time estimation and actual path finding. Kanolus et al. [11] propose a method for finding the fastest path through a road network given the constraints of a time interval at either the start of or destination of the trip. Ku et al. [12] propose an adaptive nearest- neighbour query based on travel time instead of Euclidian or network distance.

In [5] Ludger Hovestad, Vahid Mossavi proposed a conceptual data driven traffic modeling framework, which is mainly based on the application of Markov chain in a continuous coexistence with data stream from GPS data on taxi cabs.

In [2] GPS equipped vehicles are used to collect samples at regular intervals, which are then used for estimating

travel times. Many different sampling intervals are used 30 seconds in [8] and one second in [7]. Different systems will provide data recorded with different sampling rates, a solution independent of sampling rates has not been proposed to our knowledge. In [9] Quiroga et al. study the impact of changing sampling rate and road segment length using GPS.

Yan Qi [17] presented probabilistic models for short term traffic conditions predictions and compared the traffic prediction using HMM based model and one step stochastic model. He derived traffic features from embedded magnetic loop in the road.

In [16] Yannis. George, Constantinos Antoniou and Hoaris N. Koutsopoulos describe a methodology for the identification and short-term prediction of traffic state. This methodology comprises the components such as model-based clustering, variable length Markov chains and nearest neighbor classification.

In [15] Jing Yuan, Yu Zheng, Xing Xie, Guangzhon presents a Cloud-based system computing customized and practically fast driving routes for an end user using historical and real time traffic conditions and driver behaviour. The cloud builds a model incorporating day of the week, time of day, weather conditions and individual driving strategies. This paper infer the future traffic conditions on a road using an m^{th} -order Markov model and this condition is integrated into the proposed routing service. to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. ROAD TRAFFIC CONGESTION ESTIMATION SYSTEM

The process flow of proposed road traffic congestion estimation system is shown in the following figure 1. This system shows the estimated traffic congestion condition results of user desired source and destination. Firstly, we create the road segment data with their respectively latitude, longitude and road condition. Then we map these road segments with the collected GPS data using map matching algorithm. After knowing the GPS data of each road segment from map matching, we calculate the historical traffic data depending on these data and store them on cloud. When user searches the traffic conditions of desired source and destination, we collect these source and destination and real time GPS data from user. Then, we match these real time GPS data with road segment to know the vehicles' location using map matching algorithm.

Then, we estimate the traffic congestion conditions of user's desired road segments using Hidden Markov Model with these historical traffic data and real time traffic data of these road segments. After getting the traffic congestion probability results of user's desired source and destination, we present these results on Google Map by coloring on road with red for traffic jam, blue for heavy and green for normal.

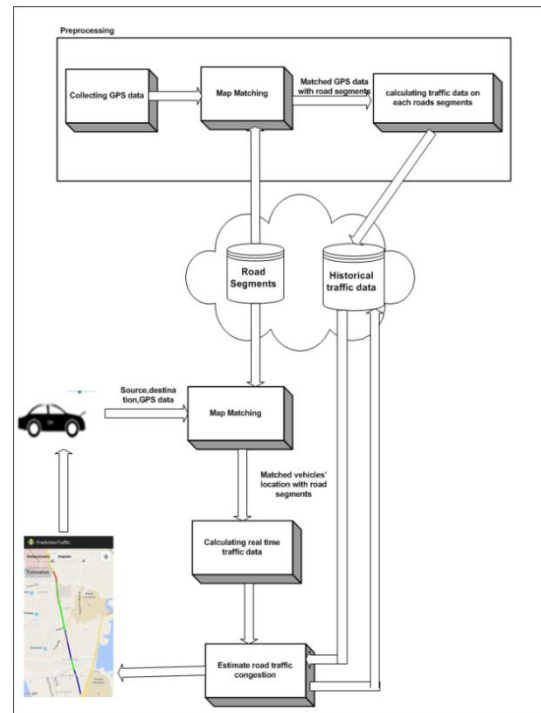


Figure 1. Process flow of the System

A. Pre processing

Historical traffic data are calculated from collected GPS trajectories data over four months. These data consists of latitude, longitude, speed, date, time and direction of each vehicle. To know the vehicle's location we match these GPS data with road network. We partition the road networks to road segments with their respective latitude, longitude that are collected from Google Map. These road segments data are store in Cloud with their respective additional information such as whether this road segment is in jam zone or not, whether this road segment has traffic lights or not etc. When we mach collected GPS data with these road segment, we use decision rule topological map matching algorithm because GPS data can have ambiguity. If we mismatch these GPS data with wrong road segment, it can reduce the accuracy of the estimation system.

Therefore, we use the decision rule topological map matching algorithm that can solve this problem. This algorithm determines the correct roadway centerlines for vehicle travel by obtaining feasible shortest paths between snapped GPS data point and snaps the point to the closest roadway by obtaining the minimum perpendicular distance from the data point to each road segment. To calculate the closest roadway with the GPS point of the vehicle we use the Spherical Law of Cosine formula because it gives well-conditioned results down to distances as small as a few meters on the Earth's surface. To reduce the calculation time we first load the road way from the database within 150 meters with the GPS point and then snap the GPS point (p1) to the closet roadway within the buffer.

The following figure 2 shows the algorithm of snap GPS point with the road segments and the figure 3 show the results of snapped GPS points on road segments.

```

Input: set of road segments( $r_s$ ), GPS trajectories
( $k_0, \dots, k_j$ )
Output: snapped GPS point

1. for each road segments in  $r_i$  in  $r_s$ 
    getBoundingBox( $k_i$ .lat,  $k_i$ .lang, distance)//
    If((minlat <  $r_i$ .lat < maxlat)
        and (minlang <  $r_i$ .lang < maxlang))
        bf[]  $\hat{C}$   $r_i$ 
    end if
end for

2. for each roadsegment  $r_i$  in bf[]
    Calculate_distance( $k_i$ .lat,  $k_i$ .lang,  $r_i$ .lat,  $r_i$ .lang)
     $r_{min} \hat{C}$  min(distance)

3. snap point ( $k_j$ ) to  $r_{min}$ 
    
```

Figure 2: Algorithm for snap GPS points with the road network

Then, we calculate the shortest path between this snapped point and the newly-snapped GPS data point (p2). If the path between these two points is not feasible, then we determine if feasible routes exist between the preceding and subsequent points bounding the GPS data points of concern. Therefore, we look ahead by snapping next newly-snapped GPS data point (p3) to nearest roadway centreline within its buffer and determine if the shortest path between snapped points (p2) and (p3) is possible. If the tested path is not feasible, we snap point (p2) to the next nearest roadway centreline within its buffer around point 3 that have not already been used in a feasibility path check. There is no feasible path between GPS point (p2) and other point p1 and p3 in the following figure 3, so the algorithm snaps the point p2 to road segment s1. This way the decision rule topological map matching algorithm solves the ambiguity of GPS data.

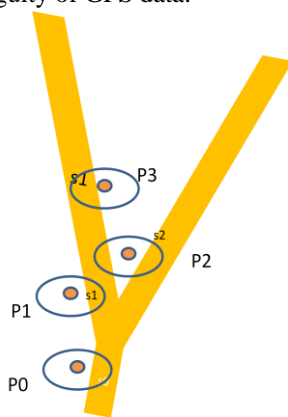


Figure 3 Snapped GPS point on road segment

The following figure 4 shows the example map matching result of the system.

After knowing the vehicle's location with the road segment from map matching result, we calculate the historical traffic data of each road segment. Historical traffic data of each road segment are calculated from collected GPS trajectories data over four months depending on speed, rush hour, whether the road segment is in jam zone or not, whether traffic lights exist in road segment or not for every minute depending on weekdays and weekends.

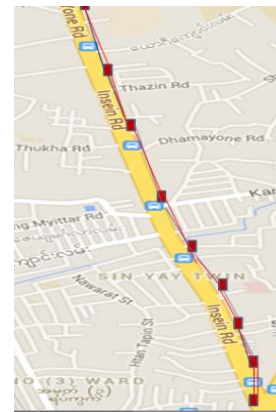


Figure 4: A screen shoot of a map matching result

B. Estimation of Road Traffic Congestion

This system shows the real time estimated results of traffic congestion states of user desired source and destination. GPS-enabled android devices equipped vehicles are used for capturing GPS trajectories data, source and destination and then these data are map with road segments using map matching algorithm. Then, real time traffic congestion states of these desired road segments are estimated with Hidden Markov model using both historical data and real time traffic data. It is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. In a regular model, the state is directly visible to the observer and therefore the state transition probabilities are only parameter. In a hidden Markov model, the state is not directly visible, but output, dependent on the state is visible.

Markov chains have some other interesting features that can be used for specific task such as finding critical urban segments, empirical expected travel times, community detection, road engineering and traffic management. Traffic conditions are considered as hidden states and traffic parameters observations are symbols. It matches the basic structure of HMMs, and therefore, an HMM is suitable for traffic modelling.

Hidden Markov Model is represented by $\lambda = (A, B, \pi), N$, the number of states in the model, M , the number of distinct observation symbols.

- The transition matrix for state distribution is initialized randomly and is represented by $A, A_{N \times N} = (a_{ij}), a_{ij} = P(s_i | s_j)$,
- The initial state of matrix of observation probabilities is represented by B ,

$$B_{N \times M} = (b_i(v_m)), b_i(v_m) = P(v_m | s_i) \text{ and}$$

A vector of initial probability is represented by $\pi, \pi_i = P(s_i)$.

These initial HMMs are then further refined by running the Baum-Welch re-estimation procedure [5] until it converges to the local minima.

We define three traffic congestion states: traffic jam, traffic heavy and normal traffic and four observation symbols: speed, traffic jam zone and traffic lights and rush

hour. We also have to define the transition probability matrix (A) and observation probability (B) and start state (π). To define the transition probability matrix we use the Baye's theorem.

$$P(x|y) = \frac{P(y|x).P(x)}{P(y)}$$

Where x is the current states of the traffic congestion (traffic jam, traffic heavy, traffic normal) and y is the historical states of the traffic congestion of each road segment in time (t) depending on weekdays and weekends. $P(x|y)$ is the probability of current states given historical state, $P(y|x)$ is the probability of historical state given current states and $P(x)$ is the probability of current state and $P(y)$ is the probability of historical states. We already predefined the observation probability for traffic jam zone and traffic lights and rush hours for each road segment in time (t). When we get the GPS data of vehicles in on the road network, we also get the speed of vehicles riding on the road. We define the speed as:

- traffic jam : if the speed is γ
 - traffic heavy : if the speed is ρ
 - traffic smooth : if the speed is υ
- where γ , ρ and υ are threshold values.

We calculate the probability of traffic congestion states using the number of speed that can whether traffic jam, traffic heavy or traffic smooth in a time window; we count the number of speed that can occur traffic jam in a time window (one minute). Then, we can model the traffic estimation with the Hidden Markov Model with the corresponding transition probability matrix, observation matrix and initial state for user desired road segments in real time. In the framework of HMMs the problem of traffic condition estimation can be expressed as an optimization problem. The Viterbi algorithm was used to search for the optimal states sequence, which is based on dynamic programming methods. Therefore, we use Viterbi algorithm to search the optimal states sequence of traffic estimation.



Figure 5 A screen shoot of estimated real time traffic congestion states

From this, we get the probability of traffic congestion states (traffic jam, traffic heavy and normal traffic). We identify traffic jam as red colour, traffic heavy as blue colour and normal traffic as green colour and present the result on Google Map. The following figure 5 shows the some estimated result of our system. These estimated traffic congestion probability results are added as the historical data that are stored in the Cloud.

C. Evaluation of the System

This section presents evaluation results of the system. This system shows traffic congestion as three states traffic jam, traffic heavy and normal traffic. The accuracy of each congestion state and all three states estimated by the system are calculated. The following figure 6 shows the traffic congested conditions of observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for jam state and accuracy of this state is show in table 1.

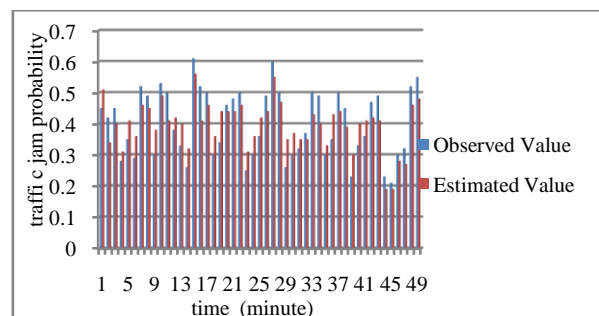


Figure 6 : observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for jam state

Estimated Result			Accuracy
Observed Result	true	false	
		27	3
	4	16	

Table 1 : Accuracy for jam state

And the following figure 7 shows the traffic congested conditions of observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for heavy state and accuracy of this state is show in table 2.

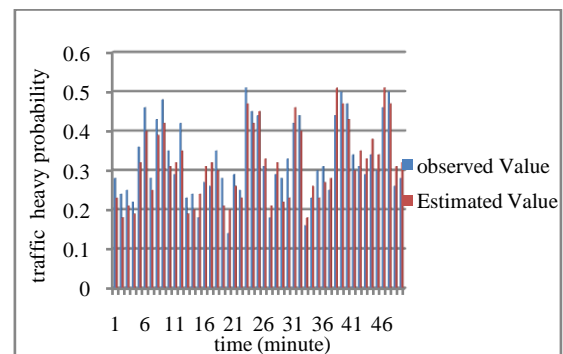


Figure 7 : observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for heavy state

Estimated Result				Accuracy
Observed Result		true	false	
	true	10	2	
	false	5	33	

Table 2 : Accuracy for heavy state

And the following figure 8 shows the traffic congested conditions of observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for normal state and accuracy of this state is show in table 3 and accuracy for all three states show in table 4.

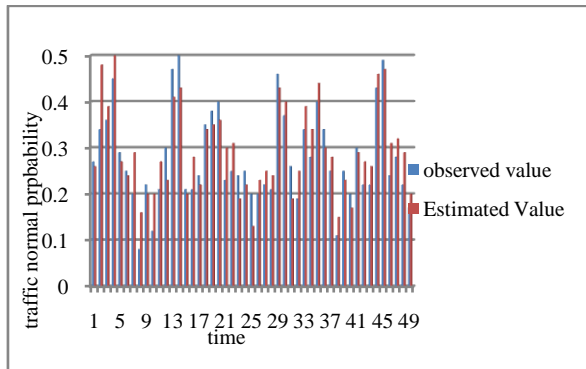


Figure 8 : observed value and estimated value in marlar_ tadaphyu road segment from 8:10AM to 9:00AM for heavy state

Estimated Result				Accuracy
Observed Result		true	false	
	true	7	3	
	false	5	35	

Table 3 : Accuracy for normal state

Estimated Result				Accuracy
Observed Result		true	false	
	true	45	9	
	false	12	84	

Table 4 : Accuracy for all three states: jam , heavy, normal

IV. CONCLUSION

This paper estimates the real time traffic congestion states for user’s desired source and destination and shows the estimated results in Google Map by coloring on road segments. We use GPS enabled mobile phone equipped vehicles to collect traffic data. By using decision rule map matching algorithm, it solves the ambiguity of GPS data that can occur mismatch when matching with road segment to know the vehicle’s location. Hidden Markov Model is used for estimating the traffic congestion states by using historical and real time data. Traffic data are collected over four months in Yangon.

REFERENCES

[1] Barbara Frith, David Pearce, and Tom Sutch. The highways agency journey time database. Road Transport Information and Control, pages 98–105, 2004.
[2] Baibo Zhang, Changshui Zhang, Xing Yi, ompetitive EM Algorithm for Finite Mixture Models, *Pattern Recognition*, Volume: 37, Issue: 1, January, 2004, pp. 131-144.

[3] Carola A. Blazquez, Chapter 9 in "Global Navigation Satellite Systems-Signal, Theory and Application"
[4] Lasgouttes J M ,Furtlehner C and Fortelle A D L 2007 A belief-propagation approach to traffic prediction using probe vehicles Proc. 10th IEEE Conf. of Intelligent Transportation Systems pp10221027
[5] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition Proc. IEEE, vol. 77, pages 257–286. Feb 1989
[6] Ludger Hovestad, Vahid Mossavi ,”Modeling Urban Traffic Dynamics in Coexistene with Urban Data Stream” August 11,2013 , Chicago,USA
[7] Ming-Hui Chen, Qi-Man Shao, Joseph G. Ibrahim, *Monte Carlo Methods in Bayesian Computation*, Springer, New York, 2000.
[8] Michael A.P. Taylor, Jeremy E. Woolley, and Rocco Zito. Integration of the global positioning system and geographical information systems for traffic congestion studies. Transportation research. Part C : Emerging technologies, pages 257–285, 2000.
[9] Nectaria Tryfona, Dieter Pfoser, and Agnes Voisard. Dynamic travel time maps - enabling efficient navigation. In SSDBM '06: Proceedings of the 18th International Conference on Scientific and Statistical Database Management (SSDBM'06), pages 369–378, 2006.
[10] N. Quiroga Cesar A and Darcy Bullock. Travel time studies with global positioning and geographic information systems
[11] Rajput D., Singh P., and Bhattacharya M., “An Efficient and Generic Hybrid Framework for High Dimensional Data Clustering,” 2010.
[12] Tian Xia, Evangelos Kanoulas, Yang Du, and Donghui Zhang. Finding fastest paths on a road network with speed patterns. In ICDE '06: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06), pages 1–10, 2006.
[13] Wei-Shinn Ku, Roger Zimmermann, Haojun Wang, and Chi-Ngai Wan. Adaptive nearest neighbor queries in travel time networks. In GIS '05: Proceedings of the 13th annual ACM international workshop on Geographic information systems (GIS'05), pages 210–219, 2005.
[14] X.David Doria,” Expectation-Maximization: Application to Gaussian Mixture Model Parameter Estimation”, April 23, 2009
[15] Xing Yi, Baibo Zhang, Changshui Zhang, Competitive EM Algorithm for Finite Mixture Models, *Pattern Recognition*, Volume: 37, Issue: 1, January, 2004,
[16] Xing Xie, Jing Yuan, Yu Zheng, Guangzhon, “Driving with knowledge from physical world”, Feb, 2013
[17] Yannis. George, Constantinos Antoniou and Haris N. Koutsopoulos “Traffic state prediction using Markov chain Model”, Julu , 2007
[18] Yan Qi “probabilistic models for short term traffic conditions”,2010