Estimation of Traffic Congestion States using GPS Data

Hnin Thant Lwin, Thinn Thu Naing
University of Computer Studies, Yangon
hninthantlwin@gmail.com

Abstract

Road traffic congestion is 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 system, we estimate the current traffic congestion states of user's desired source and destination and present the estimated results in Google Map. To get the traffic data we use GPS data from mobile phones on vehicles but this GPS points can have ambiguity. The decision support topological based map matching algorithm that can solve the ambiguity of GPS data is used to identify which vehicles are on which road by matching these GPS data with road networks. The historical traffic condition data of each road network on each time using the pre-collected data are utilized in this research. We use Hidden Markov model (HMM) for estimating the traffic condition states of these road network using historical traffic data. These estimated traffic probabilities states are presented by coloring (traffic jam for red, traffic heavy for blue and traffic normal for green) users' desired source and destination road segments on Google Map. We evaluate our estimating system using dataset generated by collect data from mobile phone-equipped vehicles over a period of 4 months in Yangon.

1. Introduction

Millions of people waste time waiting in car queues because of traffic congestion problem and it becomes late to work, miss social events, or become stressed. If there is a way to estimate current traffic congestion condition of desired roads, people could anticipate by leaving on a different time or by choosing an alternative route. Congestions are directly related to the dynamic interplay between traffic demand and capacity. If the traffic demand is high and the road capacity is low, it is likely that there will be congestion. Traffic situations can be measured with variables as speed, intensity or density.

In practice, the methodologies applied to estimate or predict traffic congestion states depend on the data available. Furthermore, the approaches that will be adopted differ in terms of provided data from sensors available on the urban network. We have to take into consideration that the urban networks are not completely covered with sensors. In order to make the real-time estimation road traffic congestion condition a success, we need information about the networks state in past and present. However, we do not have enough sensors on road network to get the complete and accurate information.

Nowadays, the information we can get from GPS is so rich for exploitation. This can be used through floating car in the urban network. The data collected can give us information about the network status, which in turn implies traffic status. Therefore, we use GPS data from mobile phones on vehicles to get the real time and historical traffic status in this paper. This system uses Hidden Markov Model (HMM) to estimate the current 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.

2. Related Works

Several papers are concerned with prediction traffic estimation. 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.

3. Process Flow of Estimation Traffic Congestion States

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.

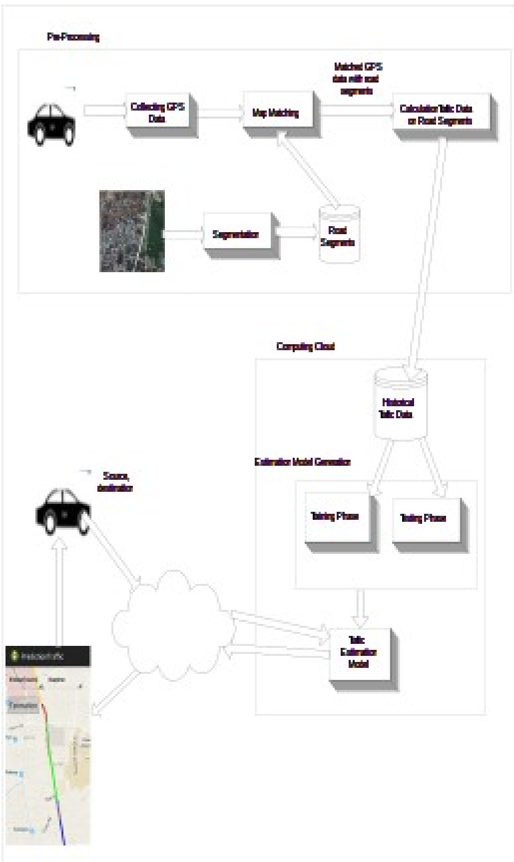

Figure1. Process flow of the System

3.1 Preprocessing

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 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.

seg_id	seg_name	length	latitude	longitude	traffic_light	congested_area
1	phayarlan	294.25	16.85789	96.12354	Yes	No
2	phayarlan	294.25	16.85823	96.12351	Yes	No
3	phayarlan	294.25	16.85864	96.12347	Yes	No
4	phayarlan	294.25	16.85888	96.12342	Yes	No
5	phayarlan	294.25	16.85915	96.12338	Yes	No
6	phayarlan	294.25	16.8593	96.12328	Yes	No
7	phayarlan	294.25	16.85938	96.12325	Yes	No
8	phayarlan	294.25	16.85993	96.12294	Yes	No
9	phayarlan	294.25	16.86029	96.1227	Yes	No
10	phayarlan	294.25	16.86082	96.12243	Yes	No
11	phayarlan	294.25	16.86077	96.12237	Yes	No
12	phayarlan	294.25	16.86047	96.12252	Yes	No
13	phayarlan	294.25	16.86021	96.12268	Yes	No
14	phayarlan	294.25	16.85993	96.12285	Yes	No
15	phayarlan	294.25	16.85968	96.12301	Yes	No
16	phayarlan	294.25	16.85961	96.12304	Yes	No
17	phayarlan	294.25	16.85913	96.12333	Yes	No
18	phayarlan	294.25	16.85864	96.1234	Yes	No
19	phayarlan	294.25	16.85842	96.12342	Yes	No
20	phayarlan	294.25	16.85788	96.12347	Yes	No
21	phayarlan	294.25	16.85789	96.12354	Yes	No

Figure 2: Some portion of road_segments data

The above Figure 3 shows some portions of road segments data. When we match collected GPS data with these road segments, 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 3 shows the algorithm of snap GPS point with the road segments.

Input: set of road segments(r_s), GPS trajectories (k_0,….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[]ć 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} ć min(distance)
3. snap point (k_i) to r_{min}

Figure 3: 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 snaps 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. This way the decision rule topological map matching algorithm solves the ambiguity of GPS data.

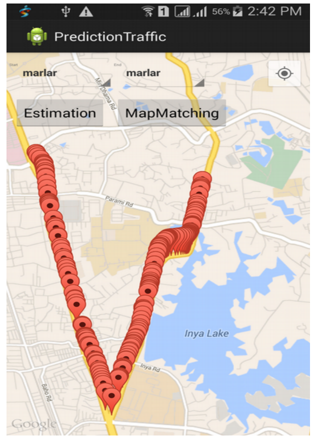

Figure 4: Screen shoot of the map matching result

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. The following figure 5 shows some portion of historical traffic data.

h_id	seg_name	jam_prob	heavy_prob	light_prob	jam_zone	traffic_ligh	rush_hour	sp_jam_prob	sp_heavy_pro	sp_light_prob	weekdays/weekend	time
1	thanlan_bartar	0.28	0.23	0.49	0	0	1	0.25	0.21	0.54	weekdays	8:01
2	thanlan_bartar	0.28	0.19	0.53	0	0	1	0.23	0.17	0.6	weekdays	8:01
3	thanlan_bartar	0.3	0.26	0.44	0	0	1	0.2	0.39	0.41	weekdays	8:01
4	thanlan_bartar	0.25	0.2	0.55	0	0	1	0.09	0.41	0.5	weekdays	8:01
5	thanlan_bartar	0.26	0.4	0.34	0	0	1	0.22	0.48	0.3	weekdays	8:01
6	thanlan_bartar	0.32	0.15	0.53	0	0	1	0.24	0.31	0.45	weekdays	8:02
7	thanlan_bartar	0.2	0.25	0.55	0	0	1	0.18	0.33	0.49	weekdays	8:02
8	thanlan_bartar	0.3	0.16	0.54	0	0	1	0.2	0.16	0.63	weekdays	8:02
9	thanlan_bartar	0.28	0.51	0.21	0	0	1	0.25	0.66	0.09	weekdays	8:02
10	thanlan_bartar	0.31	0.4	0.29	0	0	1	0.26	0.54	0.2	weekdays	8:02
11	thanlan_bartar	0.27	0.22	0.51	0	0	1	0.23	0.32	0.45	weekdays	8:02
12	thanlan_bartar	0.16	0.26	0.58	0	0	1	0.05	0.28	0.67	weekdays	8:03
13	thanlan_bartar	0.24	0.28	0.48	0	0	1	0.18	0.32	0.5	weekdays	8:03
14	thanlan_bartar	0.23	0.49	0.28	0	0	1	0.17	0.52	0.31	weekdays	8:03
15	thanlan_bartar	0.3	0.21	0.49	0	0	1	0.19	0.32	0.49	weekdays	8:03
16	thanlan_bartar	0.15	0.24	0.61	0	0	1	0.04	0.28	0.62	weekdays	8:04
17	thanlan_bartar	0.19	0.28	0.53	0	0	1	0.1	0.3	0.6	weekdays	8:04
18	thanlan_bartar	0.29	0.31	0.4	0	0	1	0.21	0.35	0.44	weekdays	8:04
19	thanlan_bartar	0.16	0.24	0.61	0	0	1	0.04	0.28	0.62	weekdays	8:04
20	thanlan_bartar	0.21	0.29	0.49	0	0	1	0.23	0.32	0.45	weekdays	8:04

Figure 5: Some portion of historical traffic data

3.2 Estimation of Traffic Congestion States

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 probabities 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, $A_{NxN}=(a_{ij})$, $a_{ij}= P(s_i \mid s_j)$,
- The initial state of matrix of observation probabilities is represented by B , $B_{NxM} = (b_i(v_m))$, $b_i(v_m) = P(v_m \mid s_i)$ and
- A vector of initial probability is represented by π, $\pi_i = P(s_i)$.

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: jammed traffic state, heavy traffic states and normal traffic states depend on four observation symbols: speed of vehicles, whether road segment is in traffic jam zone or not and whether road segment has traffic lights or not 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:

Jammed speed : if the speed is between 0 and 5

Heavy speed : if the speed is between 5 and 10

Normal speed : if the speed is above 10

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. 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 6 shows the some estimated result of our system.

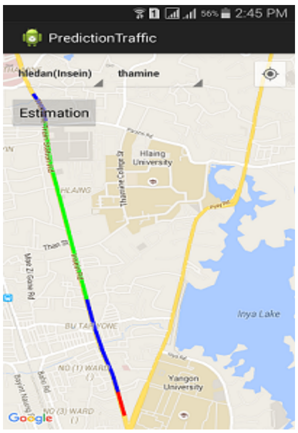

Figure 6 A screen shoot of estimated real time traffic congestion states

4. Evaluation Results

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. System's accuracy is calculated from the mean value of system's weekdays' accuracy and weekends' accuracy. Therefore, we calculate the accuracy of estimated traffic result of each road segment of the system for weekends and weekdays. We calculate the accuracy of each state in each road segment because the system estimates the traffic congestion condition with three states. Accuracy is calculated using the following equation.

$$Accuracy = \frac{Number\ of\ correct\ states}{Total\ number\ of\ test\ states}$$

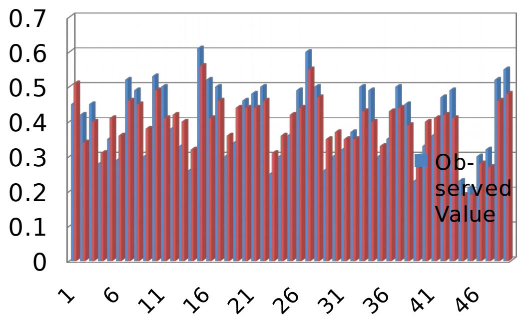

Figure 7 : observed value and estimated value in thanlan_batar road segment from 8:10AM to 9:00AM for jam state

Estimated Result				Accuracy
Observed Result		true	false	0.74%
	true	3	1	
	false	12	34	

Table 1: Accuracy for jam state in thanlan_batar road segment from 8:10AM to 9:00AM

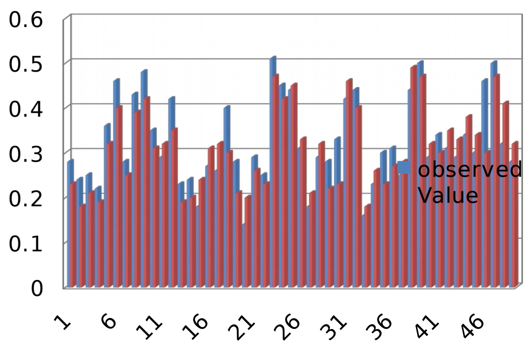

Figure 8 : observed value and estimated value in thanlan_batar road segment from 8:10AM to 9:00AM for heavy state

Estimated Result				Accuracy
Observed Result		true	false	0.78 %
	true	25	7	
	false	4	14	

Table 2: Accuracy for heavy state in *thanlan_batar* road segment from 8:10AM to 9:00AM

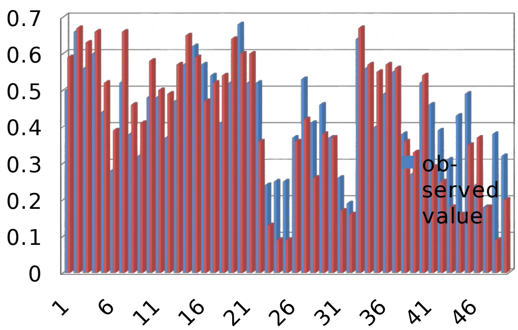

Figure 9 : observed value and estimated value in thanlan_batar road segment from 8:10AM to 9:00AM for normal state

Estimated Result				Accuracy
Observed Result		true	false	0.84%
	true	10	4	
	false	4	32	

Table 3: Accuracy for normal state in thanlan_batar road segment from 8:10AM to 9:00AM

Estimated Result				Accuracy
Observed Result		true	false	0.79 %
	true	38	12	
	false	20	80	

Table 4: Accuracy in *thanlan_batar* road segment from 8:10AM to 9:00AM

The above accuracy results are on *thanlan_bartar* road segment (one of normal road segments) that is calculated on rush hour of 50 minutes time frame between 8:10AM and 9:00AM. According to estimated results, normal traffic states occur than jam traffic state and heavy traffic state and accuracy is nearly 80 % in this time frame. Although the time (between 8:10AM and 9:00AM)is in rush hour, this road segment is normal road segment that has no traffic light or not jam zone, and normal speed probabilities occur than heavy and jam speed probabilities, so the system estimates most normal state results according to the system's traffic parameters.

	Accuracy
Insein Road	72%
Pyay Road	78%
Whole System	75%

Table 5: Accuracy of estimation system

The above table shows the accuracy result of the system. This result is calculated from the mean value of the accuracy results of each road segment of every minute for all weekdays and weekends. According to the result, Pyay road (consist of 7 road segments) can estimate more accurate than the Insein road (consist of 9 road segments). The accuracy result of whole system is calculated from the mean value of these accuracy results of two road networks.

5. Conclusions

Traffic estimation system in this paper estimate the traffic congestion states of user's desired source and destination road segment of current time. GPS-enabled cell phones can determine their position and velocity accurately. If this information is transmitted to a traffic center, it can provide a cost-effective augmentation of the available traffic data. This system estimate the traffic congestion states of source and destination of region of interest in Yangon using GPS data from mobile phone on vehicles. This type of system is particularly appropriate for developing countries, which lack resources and monitoring infrastructure, and where the penetration of mobile phones in the population is significant. This system estimates real time traffic congestion states with Hidden Markov Model using both historical data and real time traffic data. We estimate three traffic congestion states: traffic jam, traffic heavy and traffic normal of every minute in real time for user's desired source and destination. We use traffic parameters as speed, whether road segment is in jam zone or not, rush hour and whether traffic lights exist in road segment or not. This system also provides the accurate map matching for more efficient estimation results for traffic state from ambiguities of GPS data. This system uses the GPS traffic data in Yangon collected over four months. This road congestion states estimation system are more useful when sufficient data is provided because accuracy of the system is depend on data available. Given that GPS is becoming a standard feature of cell phones and the given penetration of cell phones in the population, it is likely that sufficient data will be available in the near future, making our estimation method attractive for traffic estimation purposes.

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, Chatpter 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" Augest 11,2013 , Chicago,USA

[7] Ming-Hui Chen, Qi-Man Shao, Joseph G. Ibrahim, *Monte CarloMethods 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