Estimation of Road Traffic Congestion System using GPS Data on Mobile Cloud

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

Abstract

Road traffic jams continue to remain a major problem in most cities around the world, especially in developing countries.Congested roads can be avoided by estimating the real time road traffic conditions. In this paper, we estimate the traffic congestion conditions using hidden markov model. To get the real time data, we use GPS data from mobile phone on vehicles for cheaper estimation. We accept the user's query including GPS data from the mobile phone and communicate with the cloud and present result to user. When we get the real time GPS data from mobile phone in vehicle, we use Map Matching algorithm to match these GPS data to road network to know which vehicles are on which roads. Framework based on Hidden Markov model is used to estimate the traffic congestion on each road by considering both historical traffic and present traffic flow data.

Keywords*:Traffic Estimation, GPS, Hidden Markov Model.*

1. Introduction

Urban traffic congestion is a severe problem that significantly reduces the quality of life in particularly metropolitan areas. However, frequently constructing new roads is not realistic and untenable in social and economic aspects. Therefore, the urban traffic management such as traffic congestion estimation systems has been conducted by many researchers.Traditionally, estimating travel times 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. Since road networksare ever changing and traffic volumes fluctuate, travel-time estimates must be recalculated occasionally or continually using current data to reflect these changes.

The increasing popularity of mobile phones embedded with positioning functionality such as GPS is allowing users to easily acquire theirown locations and collect their own trajectories, which can be used for various purposes such as location-based service applications. Therefore, new possibilities have opened for cheaper travel-time prediction [1,2] using this properties.This techniques can provide more accurate location information, and thus more accurate traffic data such as speeds and travel times. Additional quantities can potential be obtained from these devices, such as instantaneous velocity, acceleration, and direction of travel.For this reason, we use GPS data from mobile for the efficient and effective methods.

Previous approaches to travel-time estimation include algorithms based 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 [2,1]. 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.

In this paper, we collect data from mobile GPS on the vehicles.We then map these GPS data with the road network to know which vehicles are on which road. We then use Hidden Markov model to estimate the traffic on a particular road ahead of time using these GPS data and historical data. This paper intends (1) to estimate travel time of a route for a vehicle driver to choose fastest path to the target destination, (2) to minimize travel time and cost by avoiding traffic congestion that leads to be wastes a lot of gas and fuel (3) to calculate the traffic data from both historical and recent dynamic traffic data and to send the accurate result to mobile users.

2. Related Work

Traffic modeling has gained significant interest among researchers.Kanolus et al. [14] propose a methodfor finding the fastest path through a road network given theconstraints of a time interval at either the start of or destinationof the trip. Ku et al. [16] propose an adaptivenearest-neighbour query based on travel time instead of Euclidianor network distance.

In [12], Nielsen presents methods for using data recordedby GPS devices mounted in cars to analyze congestion. Itis argued that using GPS data provides more knowledgethan traditional methods as routes can be inferred from thestream of GPS observations. Hansen presents methodsfor analyzing congestion continually over extended periodsof time, using GPS data.

In [7] LudgerHovestad, VahidMossavi 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.

Advances in GPS and tracking technology have motivatedlarge efforts in classifying trajectories. Rajput *et al.* [13], proposed a basic framework byintegrating the hypothesis of rough set theory (reduct)and k-means algorithm for efficient clustering of highdimensional data.

In [2, 12, 15] GPS equipped vehicles are used to collectsamples at regular intervals, which are then used forestimating travel times. Many different sampling intervalsare used30 seconds in [10] and onesecond in [9]. Different systems will provide data recordedwith different sampling rates, a solution independent of samplingrates has not been proposed to our knowledge. In [11]Quiroga et al. study the impact of changing sampling rateand road segment length using GPS.

Yan Qi [22] 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 [21] 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 [20] Jing Yuan, Yu Zheng, Xing Xie, Guangzhonpresents 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.

3. Estimation Traffic

The following figure shows the general architecture for traffic estimation.

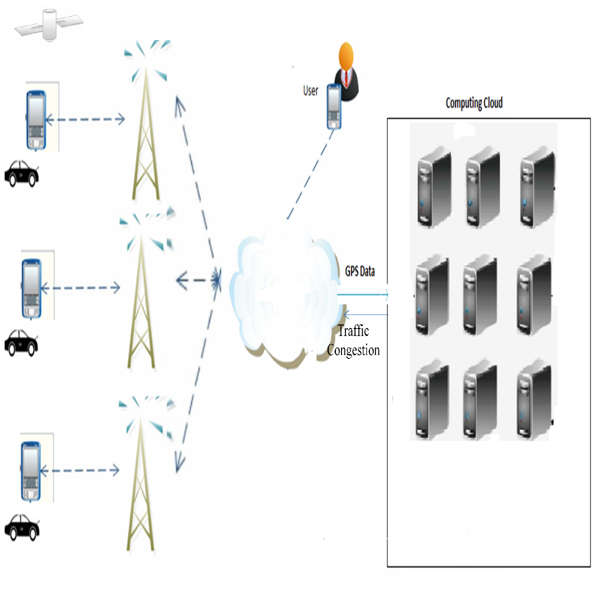

Figure 1.General architecture for Traffic Prediction.

Android device on vehicles is used for capturing GPS location data, source and destination and transmitting the data to data center. GPS data provide with longitude, latitude, a timestamp, direction, speed and a unique identification for the recording GPS device. The data transmitted by android devices deployed in vehicles is collected and in a computing cloud. These GPS data from mobile phone on vehicle is matched with the road network using map matching algorithm to know vehicle's location. We use decision rule map matching algorithm to do this. Traffic conditions are estimated using Hidden Markov model using these GPS data and historical data. To carry out these processes, we used cloud computing as back end server.

The processes of backend sever in cloud computing to estimate traffic congestion condition is described in following figure. It consists of two main parts map matching and traffic estimation.

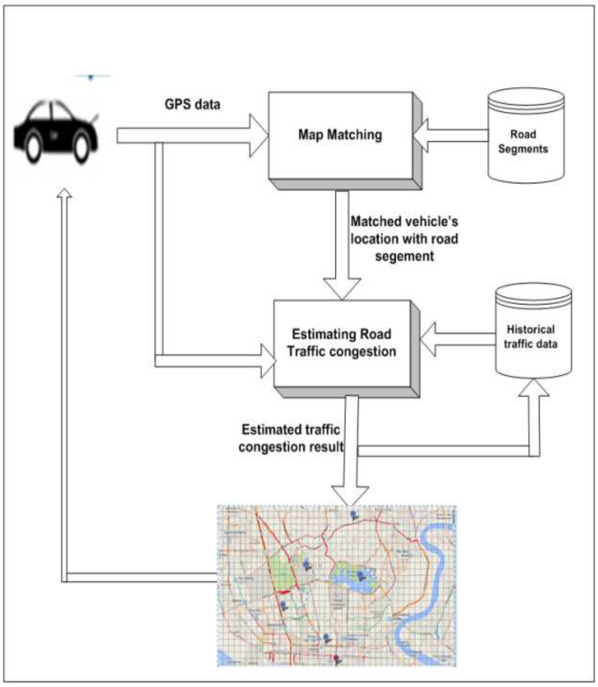

Figure 2. Process flow of the system

3.1. Map Matching

The increasing popularity of GPS-enabled device has facilitated users to track moving objects, such as vehicles and people. However, as the reading of a GPS sensor have positioning errors and sampling errors, the departure of the GPS tracking data from the actual trajectory can hardly be avoided. To match an original GPS tracking data to a digital map or a digital road network is often referred to as Map Matching. The general purpose of a map matching algorithm is to identify the true road segment on which a user (or a vehicle) is travelling.

3.1.1.Road Segment

We create the road segment database of Yangon to map the vehicle with road segments. A road segment *r*is a directed (one way or bidirectional) edge that is associated with two terminal points start point (*r.s*),end point (*r.e*), and a list of intermediate points describing the segment using a polyline. If *r.dir* = one way, r can only be traveled from start point *r.s* to end point *r.e*, otherwise, people can start from both terminal points, i.e, *r.s*→*r.e* or*r.e*→ *r.s*.To create road segment, we gather road networks of Yangon from Google Map and to get the latitude and longitude of each road segment we collect data in the open street map data. Then, we store the road segment data to the database with corresponding additional information for each road segment. The following figure shows some portions of road segment database in our system.

g_id	segment_name	traffic_lights	shape_file
1	hledan_sanyeiknyein	FALSE	010500000010000010200000800000092AD2EA704D4340CD902ABA406C3F1
2	hledan_marlar	FALSE	0105000000100000102000074000041E261D422C213042A801F590860E2CF1
3	sanyeiknyein_sinyaytwin	TRUE	0105000000100000102000001F0000E960FD9FC3D4304068924EFD4B085844
4	marlar_tadarphyu	TRUE	010500000010000010200000A30000B502D86481AC430758240FA3A143295
5	sinyaytwin_butaryoune	FALSE	010500000010000010200000240000076E09C11AD53040925A02d23908540E2
6	tadarphyu_kyakwae	FALSE	01050000001000001020000002C000036591A2449301482794 1E2A603572D53
7	butaryone_thuka	FALSE	010500000010000010200000D0000B579C3657D630402B81A5962C08584076
8	thuka_thanlan	FALSE	010500000010000010200001B0000A3F161411CD73040778A7D3521085840 8
9	bartar_okkin	FALSE	01050000001000001020000120000022CEBDE2CDD930404D6688E0F90758407
10	okkin_phayarlan	TRUE	0105000000100000102000480000E39283410BA74061579B1D37311558F453
11	hledan_sitepyoyay	TRUE	0105000000100000102000009100087 1E53A816C53601DA8041B9C5074C2E

Figure 3.some portion of road segments database

3.1.2. Matching GPS Data with Road Segments

Map matching is the process of matching a raw GPS tracking data to road network. To get the real time estimation of traffic congestion, we use the GPS data collected from the mobile phone on vehicles. GPS data for a vehicle trajectory(*Tr*) consists of a sequence of GPS points pertaining to one trip.Each point p consists of a longitude, latitude, speed, timestamp *p.t* and a unique identification for the regarding GPS device.The information we can get from GPS is so rich for exploitation. The data collected from GPS can give us information about the network status, which can turn traffic status. The following figure show the collected GPS point on the vehicles.

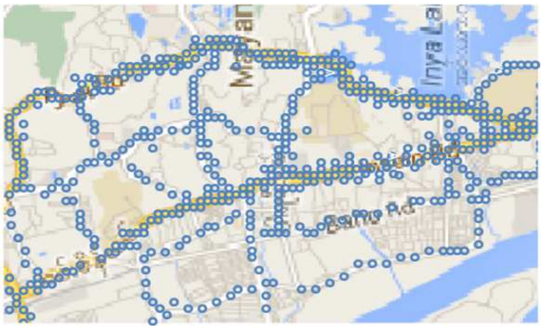

Figure 4. GPS points on digital map

According to the information of input GPS tracking data used, existing methods can be categorized into four groups: geometric, topological, probabilistic and other advanced techniques. To know which vehicles are on which road, we map these GPS data with the road segments. However, there is a problem of correctly matching a sequence of GPS sampling points to the roads on a digital map. To solve this ambiguity, we use decision-rule topological map-matching algorithm.A map matching algorithm which makes use of the connectively and contiguity information is referred to as a topological map matching algorithm. Topological methods use the topology of map features to constrain the candidate matches for a sampling point. Decision rule topological map matching algorithm determines the correct roadway centerline for vehicle travel by obtaining feasible shortest paths between snapped GPS data points in post-processing mode. The algorithm selects all roadways within a buffer around a GPS data point and snaps the point to the closest roadway by obtaining the minimum perpendicular distance from the data point to each roadway. 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 50 meters with the GPS point into the buffer and then snap the GPS point (p1) to the closet roadway within the buffer.

Algorithm 1 Snap GPS Point

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[] $\leftarrow r_i$
end if
end for

2.for each roadsegmentr_i in bf[]
Calculate_distance(k_i.lat,k_i.lang,r_i.lat,r_i.lang)
$r_{min} \leftarrow$ min(distance)
3. snap point (k_i) to r_{min}

Figure 5. Algorithm for snap GPS points with the road network

When we finish snapping this GPS point with road network we determine whether this vehicle is exist in this road using the next point. 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 centerline 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 centerline within its buffer around point 3 that have not already been used in a feasibility path check. The following diagram show the steps of the map matching algorithm that use to solve the ambiguity of GPS points in this paper.

Input: set of road segments(r_s), GPS trajectories (k_0,….k_j)

Output: map matching results

1. snap point (k_i) to r_{min}
2. ifk_i and k_j are in same road r_{min}
Match points with r_{min}
End
3. Else if k_j and k_{j+1} are in same road
4. If k_i has another nearest road segment in buffer[]
Snap point k_i to another r_{min1}
$K_s \leftarrow k_i$
If k_{s-1}, k_s, k_{s+1} are in same r_{min1}
Match points with r_{min1}
End
Else
Mismatch point
End
End if
5. Else if k_{i-1} and k_i are in r_{min1}
Match points with r_{min1}
End
Elseif find k_{i-1} has another nearest road segment in buffer[]
Snap k_{i-1} to another r_{min2}
$K_s \leftarrow k_{i-1}$
If k_{s-1}, k_s, k_{s+1} are in same r_{min2}
Match points with r_{min2}
End
Else If k_{i-2} and k_j are in same road r_{min3}
Match point with r_{min3}
End
Else
Mismatch point with r_{min3}
End
End If
End If
6. Else if k_j has another nearest road segment in bf[]
Snap the k_j to r_{min4}
$K_s \leftarrow k_j$
If k_{s-1}, k_s, k_{s+1} are in same r_{min4}
Match points with r_{min4}
End
Else
Mismatch point
End
End if
7. Else if $k_i$and k_{i+1} are in same road r_{min}
Match point with r_{min}
End
Elseifk_i has another nearest road segment in buffer[]
Snap point k_i to another r_{min1}
$K_s \leftarrow k_i$
If k_{s-1}, k_s, k_{s+1} are in same r_{min1}
Match points with r_{min1}
End
Else
Mismatch point
End
End if
End If
End If

Figure 6.the step sequence of the Map-Matching Algorithm

In the following figure 7 describes the result of map matching of the road network using the decision rule topological map matching algorithm.

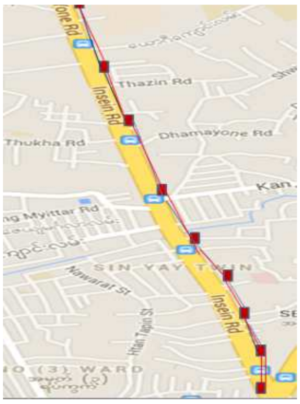

Figure 7.A screen shoot of map matching result

3.2.Estimation Traffic

When we have the map matching result of vehicle trajectories data with road segment and we know the vehicle's location, we estimate the traffic congestion conditions. In order to make the real-time traffic estimation a success, we need information about the road networks state in the past and present. Therefore, to estimate the real time traffic condition at a time (T) on each road network, weuseboth historical data on each road network (H) and the recent GPS trajectories data of mobile phone on vehicle (R).

To get the historical data for each road segment on each time, we collect the data based on weekends and weekdays. We collect the traffic congestion probability of each road segment in each minute of time window and marked which road segment is the most traffic congestion area in the city and traffic lights exist on which road segment and how long it is. For real time data, we can get it from the GPS enabled phone from the vehicle.There are many available models to estimate the traffic and these models can be categorized based on their way of computation to different kinds of statistical models, simulation models and artificial intelligence models. In this paper, we use the Hidden Markov model to estimate the traffic congestion of time (T). It is a statistical Markov model in which the system being modeled is assumed tobe a Markov process with unobserved (hidden) states. Ina 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.Markovchains 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 matchesthe basic structure of HMMs, and therefore, an HMM is suitable for traffic modeling.

Hidden Markov Model is represented by

$$\lambda = (A, B, \pi),$$

M is the number of states in the model. The individual states are denoted as $S=\{S_1,S_2,..S_M)$ and the state of the system at time t as S_i.

- A set of prior probabilities $\Pi = \{ \pi_i\}$ where $\pi_i = P(q_1 = S_i), 1 \le i \le M$
- A set of state transition probabilities $A=\{ h_{i,j}\}$ where $h_{i,j} = P(q_{t+1} = S_j | q_t = S_i),\ 1 \le i,j \le M$
- A set of output distributions $B=\{b_j\}$ where $b_j(o_t) = P(O_t = o_t | q_t = S_j) = N(O_t, \mu_j, \sum j)$, $1 \le j \le M$ where μ_j *and* $\sum j$ are the mean and covariance of the Gaussian of state S_j.

Where q_t and O_t are respectively the state and observation at time t. Wedefine three states: traffic jam, traffic heavy and traffic smooth and three observation symbols: speed, peak hour and traffic lights. 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.

$$P(x|y) = \frac{P(y|x).P(x)}{P(y)} \tag{1}$$

Where x is thecurrent stateof the traffic congestion (traffic jam, traffic heavy, traffic smooth) and y is the historical states of the traffic congestion states of each road segment in time (t).$P(x|y)$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 peak hour and traffic lights 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 state using the number of speed thatcan

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 start state.After the symbols and states are defined, the unknown model are then determined using training data. For the purpose of traffic congestion state prediction, the trained HMM model is used to find the optimal traffic state sequence corresponding to a given traffic speed observation sequence. The traffic processes were broken down into peak periods and non-peak periods because the traffic during peak periods experiences more breakdowns and congestion. 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 traffic smooth) of each road segment. We identify traffic jam as red color, traffic heavy as yellow color and traffic smooth as green color based on theseprobabilities of these traffic congestion results. We integrate these results with the Google Map and show the traffic estimation results to the user asin the following figure.

Figure 8.a sample traffic congestion in a selected part of the Yangon; Red for jam, Orange for heavy and Green for light traffic.

4. Evaluation of Traffic Estimation

To quantify the accuracy of the traffic inference, we compare the real time traffic condition of an observed road segment to the estimated traffic condition using the root mean square error (RMSE) define as:

$$RMSE = \sqrt{\frac{1}{N}\sum_i (x_i - \widehat{x_i})^2} \qquad (2)$$

whether traffic jam, traffic heavy or traffic smooth in a time window; we count the number of speed that can

Where x_i is the real travel time and $\widehat{x_i}$ is the estimated travel time and N is the number of estimation. The following figure shows the RMSE of a road segment (hledan_sanyeitnyein) of real traffic conditions and estimated traffic conditions using real time data real time and historical data of our system using the time slot 8am – 10am of the test day.

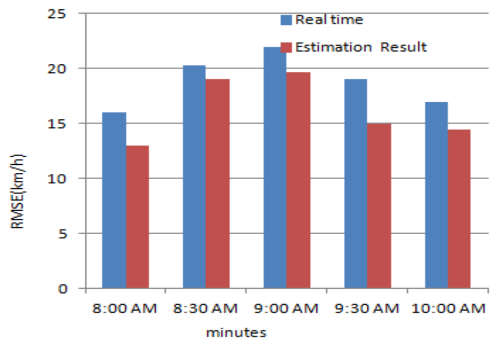

Figure 9. RMSE of a road segment

5. Conclusions

This paperdescribes a system for drivers to compute shortest-time driving routes using traffic information. Weusereal time traffic data on mobile phone in vehicles' GPS and match these GPS data with the road segments using the decision rule map matching algorithm to know the location of vehicles. In this paper we use Hidden Markov Model to estimate traffic congestion for more accurate result using real time GPS data and historical data.

5. References

[1] B.S. Yoo, S.P. Kang, and C.H. Park. Travel time estimation using mobile data. In Proceedings of theEastern Asia Society for Transportation Studies, Vol.5, pages 1533–1547, 2005.

[2] Barbara Frith, David Pearce, and Tom Sutch. Thehighways agency journey time database. RoadTransport Information and Control, pages 98–105,2004.

[3] Baibo Zhang, Changshui Zhang, Xing Yi, ompetitive EM Algorithmfor Finite Mixture Models, *Pattern Recognition*, Volume: 37,Issue: 1, January, 2004, pp. 131-144.

[4] Isra_lTamim, Asif Khan ,Emdad Ahmed, Muhammad Abdul Awal Multiple Parameter Based Clustering (MPC): Prospective Analysis for Effective Clustering in Wireless Sensor Network (WSN) Using K-Means Algorithm Wireless

[5] Jyoti R. Patole, "Clustering the Wireless Sensor Network using K-Means and MapReduce algorithm", 2012.Sensor Network,4, 18-24,2012.

[6] Lasgouttes J M ,Furtlehner C and Fortelle A D L 2007 A belief-propagation approach to trafficprediction using probe vehicles Proc. 10th IEEE Conf. of Intelligent Transportation Systems pp10221027

[7] LudgerHovestad, VahidMossavi ,"Modeling Urban Traffic Dynamics in Coexistene with Urban Data Stream" Augest 11,2013 , Chicago,USA

[8] Ming-Hui Chen, Qi-Man Shao, Joseph G. Ibrahim, *Monte CarloMethods in Bayesian Computation*, Springer, New York, 2000.

[9] 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.

[10] NectariaTryfona, DieterPfoser, and Agnes Voisard. Dynamic travel time maps - enabling efficient navigation. In SSDBM '06: Proceedings of the 18^{th} International Conference on Scientific and Statistical Database Management (SSDBM'06), pages 369–378, 2006.

[11] N. Quiroga Cesar A and Darcy Bullock. Travel timestudies with global positioning and geographic information systems: an integrated methodology. Transportation research. Part C : Emerging technologies, pages 101–127, 1998.

[12] Otto Anker Nielsen. Analyse of traengseloghastighedervha. gps-data. Trafikdage, pages 1–21, 2003.

[13] Rajput D., Singh P., and Bhattacharya M., "An Efficient and Generic Hybrid Framework for High Dimensional Data Clustering," *in Proceedings of International Conference on Data Mining and Knowledge Engineering*, *World Academy of Science, Engineering and Technology*, Rome, pp. 174-179, 2010.

[14] Tian Xia, EvangelosKanoulas, Yang Du, and Donghui Zhang. Finding fastest paths on a road network with speed patterns. In ICDE '06: Proceedings of the 22^{nd} International Conference on Data Engineering(ICDE'06), pages 1–10, 2006.

[15] W. Carola A. Blazquez, Chatpter 9 in "Global Navigation Satellite Systems-Signal,Theory and Application "

[16] Wei-Shinn Ku, Roger Zimmermann, Haojun Wang, and Chi-Ngai Wan. Adaptive nearest neighbor queries in travel time networks. In GIS '05: Proceedings of the13th annual ACM international workshop onGeographic information systems (GIS'05), pages210–219, 2005.

[17] X.DavidDoria," Expectation-Maximization: Application to Gaussian Mixture Model Parameter Estimation", April 23, 2009

[18] Xing Yi, Baibo Zhang, Changshui Zhang, Competitive EM Algorithm for Finite Mixture Models, *Pattern Recognition*, Volume: 37, Issue: 1, January, 2004, [19] Xing Xie, Jing Yuan, Yu Zheng, Guangzhon, "An Interactive- voting based Map Matching Algorithm",In Proc.MDM 2010.

[19] Xing Xie, Jing Yuan, Yu Zheng, Guangzhon, "Driving with knowledge from physical world", Feb, 2013

[20] Yannis. George, Constantinos Antoniou and Haris N. Koutsopoulos"Traffic state prediction using Markov chain Model", Julu, 2007

[21] Yan Qi ".probabilistic models for short term traffic conditions predictions"