Travel Pattern Mining Assisted Solution for Vehicle Routing Problem (VRP)

Sudha Chaturvedi[1], Tapsi Nagpal[1], Vishnu Shankar Tiwari[2]
[1]Department of Computer Science and Information Technology, Lingayas University, India
[2]Department of Computer Science and Engineering (AI&ML), PES University, India

ABSTRACT

Vehicle routing problem (VRP) is a well researched optimization in the field of route planning and recommendation system for transportation and logistics in supply chain. Existing solutions uses mathematical techniques to solve combinatorial problem over digitized road network considering factors like shortest distance, traffic condition etc., individually or a combination of these factors. Almost all the solutions use this road network graph to compute feasible shortest paths and then do the optimization for minimizing the distribution cost. But it's not necessary that feasible shortest paths selected by weighing scheme be the best path. Many factors influence the determination of the shortest best path between any two given places, for example - security, lighting conditions condition or any unseen scenarios etc. These are known by local drivers and local residents and so mostly these people drive through most optimal and best paths. We propose a framework in which spatial data mining is performed on location data captured from devices like GPS, Mobile phone location etc., for finding the frequently user traveled paths between each pair nodes. After applying the mathematical optimization additional weighing is done based on travel pattern model to select best path. This model can be applied on any existing VRP solution to provide more practical route planning. Proposed framework is demonstrated with practical scenario and evaluated on real datasets.

Keywords: Vehicle routing problem (VRP), Road Network, Travel Pattern, Route Optimization, Geo Spatial

INTRODUCTION

This problem exists in many different variations like VRPPD, LIFO, VRPTW and CVRP etc. Solution to Vehicle Routing Problem with Pickup and Delivery (VRPPD) deals with a large number of logistics need to be transported from various source locations to multiple destinations following optimal routes with constraints. The objective is to compute optimal routes for a fleet of vehicles to pickup and drop-off locations whereas, Vehicle Routing Problem with (LIFO) solves same issue with considering additional constraint on the loading of the vehicles at source and delivery of the most recent item picked up. VRP with Time Windows (VRPTW) deals with constraint on delivery destinations have pre-defined time window for delivery. Capacitated Vehicle Routing Problem (CVRPTW) considers the constraint of limited capacity for carriage when scheduling delivery route [4, 15].

Fleet of vehicles with limited capacity needs to be planned route in order to visit set of destinations at a minimum cost which is objective of Vehicle Routing Problem (VRP). [15, 16]. All algorithmic solutions uses digitized road Network. These algorithms assign some weight to edges of Road Network graph. Weights may be dependent on factors like shortest distance, traffic condition etc., individually or a combination of these factors. The solutions in the literature can be categorized as Static and Dynamic. In the static Vehicle Routing Problem, all the contributing factors are known a priori [1]. Dynamic Vehicle Routing Problems (DVRP) deals with issues which have arisen like road blocks, water logging etc. obtained by using communication and information technologies and processed in real time. Static solutions are quick in computation but less practical in dealing with surprising situations for e.g, road blocks, while dynamic solutions are harder in computation time but are more practical in handling unseen situations.

Almost all existing solutions use road network graph to compute feasible shortest paths and then do the optimization for minimizing the distribution cost. But it's not necessary that feasible shortest path selected by distance weighing scheme be the best path. Many factors influence the determination of the best path between any two given places. For example, security, lighting conditions, condition of the road etc. These are known by local drivers and local residents and so mostly these people drive through most optimal and best paths. Researchers found that the taxi drivers/local residents follow the best and optimized paths to minimize travel time taking into consideration various static and dynamic factors involved in routing.

Selection of the best path while scheduling cannot only be found by querying the road network but instead it a data mining problem. In this paper we formulate a framework which is based on spatial data mining. Our solution depends upon the mining of location data received from GPS device installed on the vehicles. Over a period of time large amount of location data becomes available. These location data can be broken down in the form of trips. These trips with the help of map matching can be associated with road network edges [5]. Then between each destination frequently followed path can be found [6]. Optimal routes are computed using mathematical combinatorial approach and they further assisted by users travel pattern to avoid any unseen issues.

BACKGROUND

It is often observed that route recommendation systems are biased towards finding shortest path to minimize cost and travel time and hence maximize profits. Existing algorithms considers static factors like road traversal time and sometime other factors like traffic conditions. But that may not be always the best path. Other factors also do come into picture when best route is to be constructed like security, road conditions, lighting conditions etc. Fig 1 shows route suggested for a source and destination and Fig 2 shows bad road conditions on suggested route.

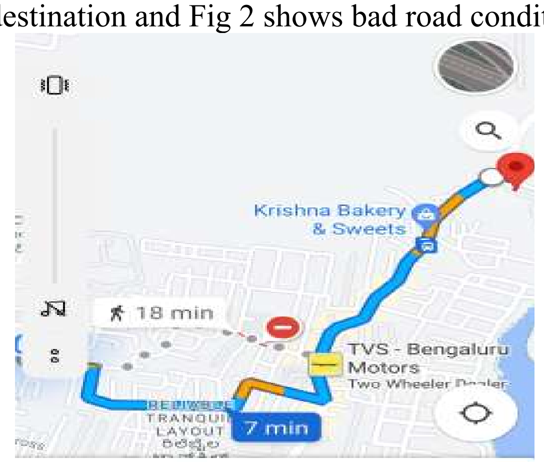

Fig 1. Route suggested by App

Fig 2: Suggest route has bad road condition

Taxi drivers for example know the city roads and when choosing the route from source to a destination considers shortest path in order to maximize the profit but not at the cost of compromising important hidden factors which may lead them in trouble [17]. Similarly, local residents know the area roads better and do follow the best path considering aspects like security and route quality. Fig 3 shows local residents and taxi drivers following better route with better quality for same destination.

Fig 3: Taxi drivers and local residents are following better paths for same destination

DATASET AND PREPROCESSING

Digitized Road Network

Digitized road network is digital representation of real network of any given area. Set of vertices in network are road intersections or important locations in the area. Road segments in the real road network represent the edges of the roads. Road network edges are assigned weights as the length of road segment [8, 9]. Digitized road network is stored in relational database which is spatial enabled. In this work storage is done in Postgres database with PostGis layer enabled to handle geometrical data type. Vertices are represented in database by a point object and road edges are represented by Multiline String. An example of road network data obtained from Open Street Maps (OSM) and stored in database Fig. 4.

User Location Data

User Location data is a point data objects consisting of coordinates as longitude and latitude [13]. It is stored in database as 1-dimensional point data object. Data obtained from Microsoft Geolife project [19, 20, 21] stored in database and its visualization is as shown in Fig. 5.

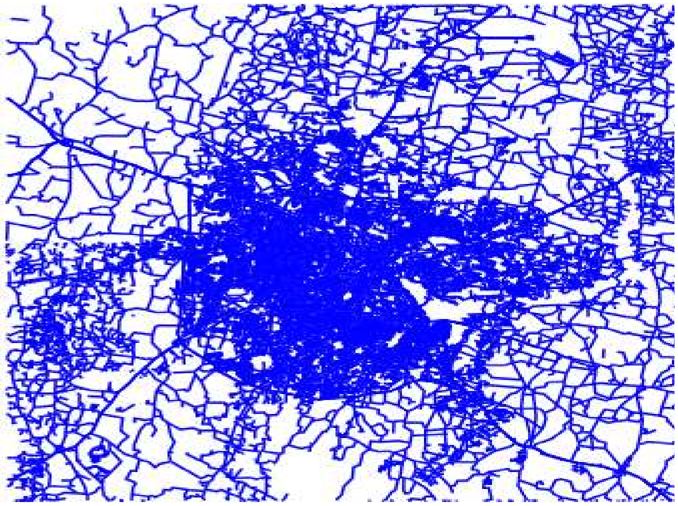

Fig 4. Road network data stored in database

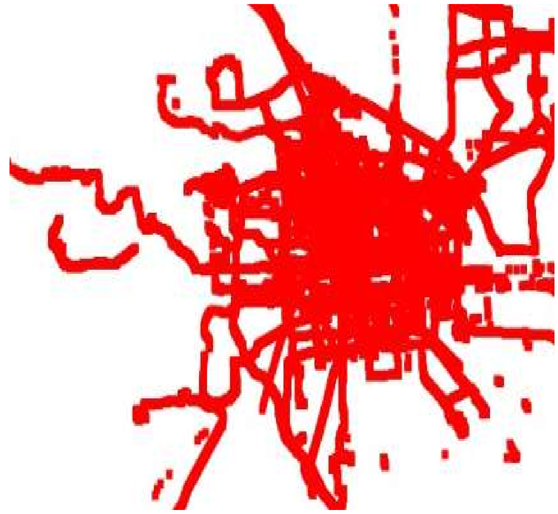

Fig 5. User Locations stored in database

Trips and Trip Segmentation

GPS devices log the continuous positional information in log files. Decomposing these data into smaller units of trips is known as Trip segmentation [5, 7]. For example a user travel log contains continuous location ordered sequence p_1, p_2 …..p_1, p_{i+1} …….p_n representing user has traveled from p_1, p_2 …..p_i and halts there for some time and then travels p_i, p_{i+1} …….p_n then this represents two trips. One realistic scenario representing this is user travels to office in morning and end of day goes to a restaurant is two trips. Further user goes back home after dining then that will be considered as another trip. Fig 6 represents a visualization of a raw trip made by user.
Fig.7 represents a user trip visualized on OSM.

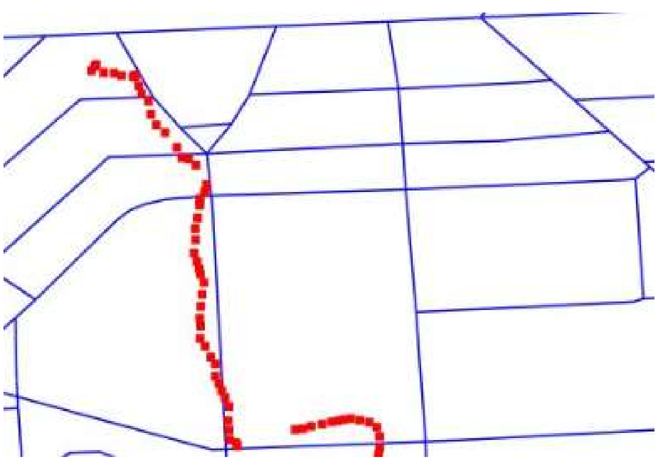

Fig 6. Visualization of a raw trip made by user

Fig.7 User trip visualized on OSM

Map Matching

Map matching is a well-established in literature which deals with locating user's location on the road network and has many use cases in different applications which depends on moving object data like route recommendation, traffic prediction, route prediction etc. [10, 11, 12]. When user travels on a road network the GPS location captured may have some inaccuracies which may be because of various factors like error in reading obtained from satellite, hardware limitation of receiving device and in many cases error in digitization of road network. This makes difficult to locate exact position of user on road network and many a times introduces ambiguity in determination process. Map matching process make use of GPS location data and digitized road network and maps user's exact position on road network [5, 6]. In this work, user trips represented by sequence of GPS data points are converted to sequence of road network edges using Map Matching. Fig. 8 shows GPS points in Fig 7 mapped to network and Fig. 9 edges of road network for mapped locations.

Fig. 8 GPS points in Fig 7 mapped to network

Fig. 9 Edges for mapped locations

TRAVEL FREQUENCY FOR USER TRIPS

As discussed in above sections continuous user travel log are decomposed in smaller units called as trips and is composed of sequence of longitude and latitude pair. A sample log file for user travel obtained from Geolife project is as below.

```
Geolife trajectory
WGS 84
Altitude is in Feet

39.984702,116.318417,,2008-10-23,02:53:04
39.984683,116.31845, 2008-10-23,02:53:10
39.984686,116.318417, 2008-10-23,02:53:15
39.984688,116.318385, 2008-10-23,02:53:20
39.984655,116.318263, 2008-10-23,02:53:25
39.984611,116.318026, 2008-10-23,02:53:30
```

Field 1 represents Latitude and field 2 represents and Longitude of location coordinates respectively in decimal degrees. Field 3 is all set to 0 and is for use in future. Altitude in feet is displayed in Field 4. Date and time is represented in in Field and 6 respectively.

These trips are Map-Matched to road network to convert the user trips in the form of sequence of road network edges.

$$MapMatching(Trip - GPS\ points\ sequece) \rightarrow Trip{:}\ Sequence\ of\ Road\ network\ edges$$
$$MapMatching(point_0, point_1, point_2, \ldots .\ point_n) \rightarrow edge_0, edge_1, edge_2, \ldots .\ edge_m$$

Algorithm 1: User trip frequency calculation
Input: Set of user trips T composed of ordered sequence of road network edges
Output: Map M of user trips and travel frequency

1. *Initialize empty map $M(String{:}\ route, Numeric{:}\ frequency)$*
2. *For each trip $T_i \in T$ Do*
   If $T_i \notin M$ insert $(T_i, 1)$ in M
   Else Increase value of route in M i.e., set $M(T_i) = M(T) + 1$

For illustration, one sample dataset Map for user trips applied on road network may look like be data in Table 1.

Trip Id (i)	Trip (T_i)	$Frequency\ (f_i)$
1	$edge_1, edge_2, edge_3, edge_4$	$550k$
2	$edge_1, edge_2, edge_4, edge_3$	$850k$
3	$edge_1, edge_3, edge_2, edge_4$	$952k$
4	$edge_1, edge_3, edge_4, edge_2$	$600k$
5	$edge_1, edge_4, edge_2, edge_3$	$900k$
6	$edge_1, edge_4, edge_3, edge_2$	$300k$

ROUTE SUGGESTION FOR VRP

Given is the set of destinations to be visited $e_1, e_2 \ldots .\ e_n$ find the optimized travel order for the vehicle [14]. Approach now is to select between each pair of location points the set of edges and their combined length. Any of the shortest path algorithm can be used for example Dijkstra's algorithm can be used for finding the shortest path from a source and a targeted destination. Similarly, if one is interested in calculation of shortest path from all the sources to all the destinations then Floyd Warshall's algorithm can be used [18]. It depends on the nature of the requirement for application.

In this work we demonstrate by an example where a user has to travel from a particular source to all other destination in shortest possible time. So all possible combination will explored and whichever visiting sequence takes less time will be considered. This is for illustration purpose but this solution can be generalized and applied for routs calculated by any shortest path algorithm for example single source shortest path, all pair shortest path etc. Adjacency matrix for input road network is as given in Table 2.

	$edge_1$	$edge_2$	$edge_3$	$edge_4$
$edge_1$	0	2	5	7
$edge_2$	2	0	8	3
$edge_3$	5	8	0	1
$edge_4$	7	3	1	0

Table T2: Input adjacency matrix for VRP

All combinations for traveling to all other nodes starting from node $e_1$along with distance covered are shown in Table 3. All frequency of route travelled is sourced from Table 1 is mentioned.

Trip Id (i)	Trip (T_i)	Cost c_i	$Frequency\ (f_i)$
1	$edge_1, edge_2, dgee_3, edge_4$	11	$550k$
2	$edge_1, edge_2, edge_4, edge_3$	6	$10k$
3	$edge_1, edge_3, edge_2, edge_4$	16	$952k$
4	$edge_1, edge_3, edge_4, edge_2$	9	$600k$
5	$edge_1, edge_4, edge_2, edge_3$	18	$900k$
6	$edge_1, edge_4, edge_3, edge_2$	16	$300k$

Algorithm 2: Best path selection
Input: Set of user trips T, Adjacency Matrix, Travel Frequency F for each $T_i \in T$, Threshold Frequency f_t, Source S
Output: Trip representing best path

1. *Calculates shortest path cost from source S to all destinations*
2. *Remove Trips from T for which Frequency $f_i < f_t$*
3. *Output the trip for which cost is minimum from remaining Trips in T*

Can be observed from above example e_1, e_2, e_4, e_3 is the shortest path with total cost of 6. Say if the threshold for frequency is set for $500k$ i.e, if number of users traveled is below then this frequency is not advisable. Hence, Trip Id 2 and 6 are not advisable. From remaining, the next best path is Trip Id 4 with cost of 9.

PERFORMANCE AND IMPLEMENTATION

Proposed framework has two phases, in first phase give an source and all required destinations to be covered. It requires road network data in digital form. Road network data is sourced from Open Street Map (OSM). OSM is open platform from where digital data like road network, land usage, water bodies, national and international boundaries can be downloaded. In this work road network data from India and China is used. Data is stored in geo-Spatial enabled data relation base. For this purpose Postgres database with PostGis data layer enabled for handling geometrical data types. Road network data is preprocessed in Sql format and imported in data base. On this road network data shortest path algorithm is run to calculate the distances between source and destinations.

Second phase is to construct travel pattern model. It requires Road network data and historical travel data. Road network data is same which is used in previous phase. Fig. 10 shows a sample road network data stored in database.

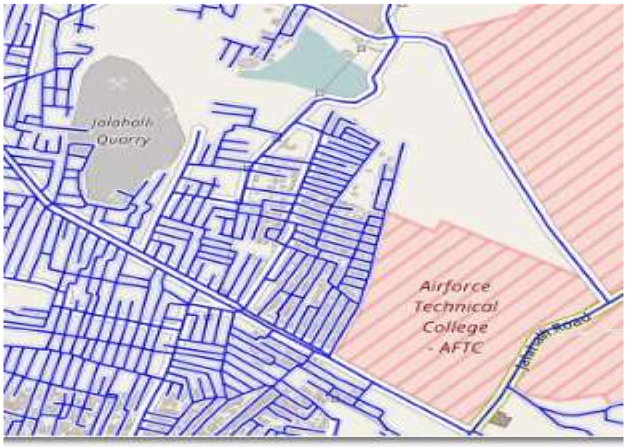

Fig. 10 Sample road network

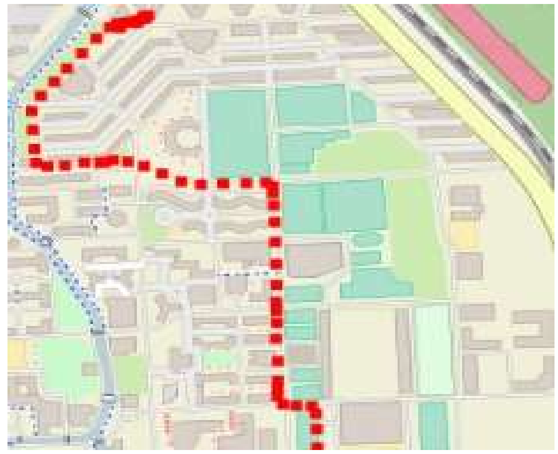

Fig. 11 Shows sample GPS data

Historical location traces are sequence of GPS traces collected from users. This data for evaluation is sourced from Microsoft's Geolife Project [19, 20, 21]. This corpus has millions of historical GPS data collected over time and released for research purpose. Location traces are decomposed in smaller units and stored in PostGis database which was used for storage of road network data as well. Fig. 11 shows sample GPS data sourced and loaded in database. All implementation is done on real datasets. All evaluation and implementation was carried out on a system having Inte7 processor with 16 GB internal memory and 500 Hard Disk. Travel pattern model was constructed from historical data consisting of corpus of huge historical data. Running time for model construction is as shown in Fig. 12. In second phase input was a source and multiple destinations. After shortest path was computed filter and selection of best path was done. Fig. 13 represents total time to response including shortest path and filtering both.

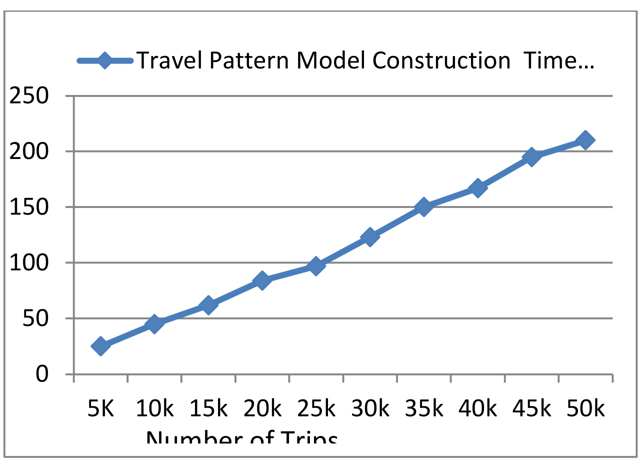

Fig. 12 Model construction time

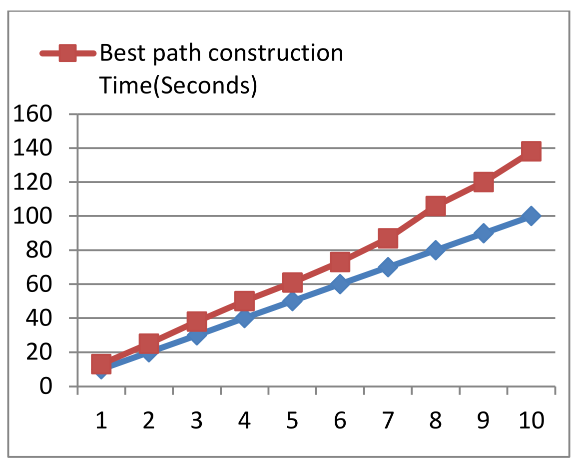

Fig. 13 Total time to response

REFERENCES

[1] Liong, C.-Y, Wan, I., Omar, K., 2008, Vehicle routing problem: Models and solutions. Journal of Quality Measurement and Analysis. 4. 205-218.

[2] Kallehauge, B., Larsen, J., Madsen, O.B.G., Solomon, Marius M.M., 2005, Vehicle Routing Problem with Time Windows, Book Chapter, Column Generation, 2005, ISBN : 978-0-387-25485-2.

[3] Yeun,L.C., Ismail,W.R., Omar,K., Zirour ,M., 2008, Vehicle Routing Problem: Models and Solutions, Journal of Quality Measurement and Analysis, JQMA 4(1) 2008, 205-218.

[4] J. Vasilev, 2015, Solving the Traveling Salesman Problem with the All different, In International Conference on Application of Information and Communication Technology and Statistics in Economy and Education, Sofia.

[5] Tiwari,V.S., Arya,A., 2018, Distributed Context Tree Weighting (CTW) for Route Prediction, Open Geospatial Data, Software and Standards, Springer Journal, 2018.

[6] Tiwari,V.S., Arya,A., 2017, Horizontally scalable probabilistic generalized suffix tree (PGST) based route prediction using map data and GPS traces, Journal of Big Data, Springer Journal, 2017.

[7] Nawaz, A.; Zhiqiu, H.; Senzhang, W.; Hussain, Y.; Naseer, A.; Izhar, M.; Khan, Z., 2020, Mode Inference using enhanced Segmentation and Pre-processing on raw Global Positioning System data, Meas. Control 2020

[8] Ekim, B., Sertel, E., & Kabadayı, M. E., 2021, Automatic road extraction from historical maps using deep learning techniques: A regional case study of turkey in a German World War II Map. ISPRS International Journal of Geo-Information, 10(8), 492

[9] Saeedimoghaddam, M., Stepinski, T. F., 2020, Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. International Journal of Geographical Information Science, 34(5), 947-968.

[10] Quddus, M. A., Noland, R. B., Ochieng, W. Y., 2006, A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport, Journal of Intelligent Transportation Systems, 10: 3, 103 — 115

[11] Yin, H., Wolfson,O., 2004, A Weight-based Map Matching Method in Moving Objects Databases, SSDBM '04 Proceedings of the 16th International Conference on Scientific and Statistical Database Management IEEE Computer Society Washington, DC, USA, ISBN:0-7695-2146-0

[12] Lou,Y., Xie,X., Zhang,C., Wang,W., Zheng,Y., Huang,Y., 2009,Map-Matching for Low-Sampling-Rate GPS Trajectories , ACM GIS ISBN 978-1-60558-649-6/09/11

[13] Nawaz, A., Huang, Z., Senzhang, W., Yasir, H., Izhar, K., Zaheer, K., 2020, Convolutional LSTM based transportation mode learning from raw GPS trajectories. IET Intelligent Transport Systems. 14. 570-577. 10.1049/iet-its.2019.0017.

[14] Srivastava, K., Pandey P.C., Sharma, J.K., 2020, An approach for route optimization in applications of precision agriculture using UAVs. Drones 4(3), 58 (2020)

[15] Toro, E. M. O., Escobar, A. H. Z. and Granada, M.E., 2015, Literature Review on the Vehicle Routing Problem in the Green Transportation Contest, Luna Azul, vol. 42, pp. 362- 387.

[16] Goel,A., Gruhn,V., 2006, A General Vehicle Routing Problem, European Journal of Operational Research, Volume 191, Issue 3, 16 December 2008, Pages 650–660.

[17] Liu, L., Andris,C., Biderman,A., Ratti,C., 2009, Uncovering Taxi Driver’s Mobility Intelligence through His Trace, IEEE Pervasive Computing (2009).

[18] Kairanbay, M., Jani, H.M., 2013, A Review and Evaluations of Shortest Path Algorithms, International Journal of Scientific & Technology Research 2(6):99-104

[19] Zheng, Y., Zhang, L., Xie X., Ma, W. Y., 2009, Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800

[20] Zheng., Y., Li, Q., Chen, Y., Xie, X., Ma, W.Y., 2008, Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea. ACM Press: 312-321.

[21] Zheng, Y., Xie, X., Ma, W.Y., 2010, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in IEEE Data Engineering Bulletin. 33, 2, 2010, pp. 32-40.