J Electr Eng Technol.2015; 10(4): 1836-1842
http://dx.doi.org/10.5370/JEET.2015.10.4.1836

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

A Clustering Scheme for Discovering Congested Routes on Road Networks

He Li*, Kyoung Soo Bok, Jong Tae Lim**, Byoung Yup Lee*** and Jae Soo Yoo†**

Abstract – On road networks, the clustering of moving objects is important for traffic monitoring and routes recommendation. The existing schemes find out density route by considering the number of vehicles in a road segment. Since they don't consider the features of each road segment such as width, length, and directions in a road network, the results are not correct in some real road networks. To overcome such problems, we propose a clustering method for congested routes discovering from the trajectories of moving objects on road networks. The proposed scheme can be divided into three steps. First, it divides each road network into segments with different width, length, and directions. Second, the congested road segments are detected through analyzing the trajectories of moving objects on the road network. The saturation degree of each road segment and the average moving speed of vehicles in a road segment are computed to detect the congested road segments. Finally, we compute the final congested routes by using a clustering scheme. The experimental results showed that the proposed scheme can efficiently discover the congested routes in different directions of the roads.

Keywords: Location based service, Trajectory data, Road network, Clustering, Vehicle

1. Introduction

In today's world, with the increase of the use of mobile devices, the location-based services are becoming increasingly popular. Since the rapidly increased satellites and GPS (Global Position System) technologies have developed, it is possible to collect a large amount of trajectory data of moving objects such as the vehicle position data, hurricane track data, and animal movement data [1, 2, 16, 17]. The analysis over these trajectory data is becoming important for many applications, such as meteorological observation and forecast, animal habits observation, road traffic situation analysis, and navigation in transportations [3-6, 8]. According to the recorded trajectory data and road networks, the moving pattern, traffic situation and road recommendation services can be supported [1, 2, 12, 15, 18]. Recently, with the continuously increasing mobile devices and vehicles, the route recommendation service is becoming more and more important [1, 5, 6, 9, 17, 19, 20]. For road network based applications, the mobility of the vehicle is road network constrained.

Most of the existing schemes try to monitor and forecast the traffic by using the recorded history trajectory data of vehicles equipped with GPS devices. The index based schemes construct an index by adopting the trajectory data of the moving objects [3, 4]. And then, the routes are recommended according to the history trajectory data of the related moving objects. Some schemes generate the density routes of the road networks by analyzing the trajectory data of vehicles [7-10]. The density regions of the road networks are evaluated by considering both the location and time of the moving objects. According to the trajectory data, the number of the moving objects within a specific road segment and timestamp is be used to identify the density regions of a road. After that, the small density regions of each road are clustered and the final density regions in the road networks are generated. However, there are three major problems of the existing schemes to be applied in real road networks: (1) the directions of the roads in the road networks are not considered; (2) the widths and lengths of the road segments are not considered; (3) the average moving speed of vehicles within a road segment is not considered. In the real road network environments, each road is divided into two directions: positive direction and negative direction. The vehicles in the road toward to different directions do not affect each other. The width and length of each road segment are different in a road network, which will also affect the accuracy of the congested routes. Furthermore, the average moving speed of vehicles within a road segment can identify the congestion of the road.

To overcome these problems, we propose a clustering method for congested routes discovering, the directions, width, and length of roads are considered in real road network environments. The proposed scheme can be divided into three steps. First, it divides each road network

† Corresponding Author: Dept. of Computer and Communication Engineering, Chungbuk National University, Korea. (yjs@chungbuk.ac.kr)
* School of Software, Xidian University, China. (heli@xidian.edu.cn)
** Dept. of Computer and Communication Engineering, Chungbuk National University, Korea. ({ksbok, jtlim}@chungbuk.ac.kr)
*** Dept. of Electronic Commerce, Paichai University, Korea. (bylee@pcu.ac.kr)
Received: March 10, 2014; Accepted: February 11, 2015

into segments with different width, length, and directions. Second, the congested road segments are extracted by considering the average moving speed of the vehicles and the saturation degree of each road segment in the road networks. Third, we compute the final congested routes by using a clustering scheme. The experimental results showed that the proposed scheme can efficiently discover the congested routes in different directions of the roads.

The remainder of the paper is organized as follows. Section 2 discusses related work in section 2. Section 3 presents the details of the proposed scheme. Section 4 contains experimental evaluation that demonstrates the superiority of our proposed scheme. Finally, section 5 concludes this paper.

2. Related Work

Most of the existing schemes try to monitor and forecast the traffic by using the recorded history trajectory data of vehicles equipped with GPS devices. [1] proposed the MPR scheme for discovering the popular route between two locations by observing the traveling behaviors of many previous users. The maximum probability product algorithm is used for discovering the MPR from a transfer network based on the popularity indicators in a breadth-first scheme. [7] proposed a new density-based algorithm called FlowScan. It is a robust algorithm that can handle the complexities in the data and we verify through extensive experiments. Instead of clustering the moving objects, road segments are clustered based on the density of common traffic they share. [20] studied the problem of vehicular traffic density estimation, utilizing the information cues present in the cumulative acoustic signal acquired from a roadside-installed single microphone.

Vehicles which are located in a congested area try to move to a non-congested area. [12] proposed a route discover method for alleviating traffic congestions to provides a driving route whose trip time becomes short. The proposed method does not need global traffic information but regional traffic information for each vehicle. The vehicle calculates a route for a destination where a summation of evaluation values for roadway segments in the route becomes minimal. Given a spatial range and a user preference of depth/breadth specified by a user, [15] processed a Pattern-Aware Trajectory Search (PATS) to retrieve the top K trajectories passing through popular regions-of-interest (ROIs). PATS support trip planning without requiring prior knowledge of ROIs in the specified spatial range. PATS used a user movement graph to capture travel patterns hidden in trajectories and develop an algorithm to determine the attractive scores of the ROIs and proposed an algorithm BTS for efficiently retrieving the top K trajectories.

[11] proposed a fast path algorithm of finding the best shortest paths in the road network to solve the path planning problem in route guidance systems in terms of accuracy and speed. [13] presented a routing algorithm which uses the road hierarchy and pre-computed areas to limit the search space. This improves trip duration by using upgraded roads whenever beneficial, and finds routes that take into consideration both speed and driving patterns. [14] studied the problem of finding reasonable alternative routes in road networks.

[9] presented NETSCAN which carries out the clustering of dense sections and incorporates them by forming dense routes. NETSCAN cluster the road sections based upon the network density statistics. This clustering takes into account the orientation of the trajectory. Besides, this method utilizes the network topology to create relevant clusters. To propose a model to assess the evolution for dense route pairs at two consecutive time intervals, DENSITYLINK algorithm is presented. DENSITYLINK allows the characterization of the evolution of the dense road network. [10] proposed a time-based clustering algorithm called Tk-means that adapts the k-means algorithm for trajectory data. Tk-means cluster the objects based on the time intervals of different trajectory's motions. If an object spans different time intervals, it will eventually belong to different clusters. Tk-means used two approaches, an exact method and an approximate method. The exact method computes the actual clusters visited by the object throughout its life time and the approximate method exactly computes some of the actual visited clusters and based on those computed clusters along with the clusters generated from the remaining data set, it predicts the future motion pattern of the query object.

However, none of them consider the features of each road segment such as width, length and directions in a road network, they are not suitable for some real road networks. In this paper, we propose a clustering method for congested routes discovering in real road network environments. The congested road segments are extracted by considering the average moving speed of the vehicles and the saturation degree of each road segment in the road networks.

3. The Congested Routes Discovering Scheme

3.1 Road network and trajectories

Definition 1 *Road network*, we assume that the road network is represented by a graph $G(N, E)$, where N denotes the node which is the intersection between different road segments and E denotes the edge which is used to connect two adjacent nodes in the road network.

Definition 2 *Road segment*, E_i is used to denote a segment of the road network. '+' and '−' are used to represent the different directions of vehicles in a road network. Moreover, since the length and the number of lane of each road are different, the length and width of each road segment are stored. Therefore, each road segment is

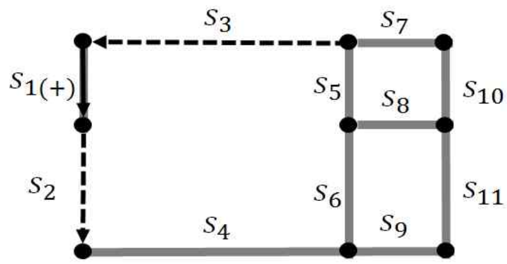

Fig. 1. The data model of road network

represented by $S_i(\pm) = \{N_i, N_j, length, width\}$, where *length* is the length of a road segment and *width* is the number of traffic lanes.

As shown in Fig. 1, S_2 and S_3 are the neighbor segments of S_1 in a road network *G*. In a road network *G*, each road segment S_i stores the information of its directly connected road segments. This information is used for the following clustering evaluations.

Definition 3 *Trajectory*, the trajectory of a vehicle is represented by *Tr*. Each node N_i in the road network is represented by a point $\{x_i, y_i\}$. Suppose that the trajectory *Tr* of each vehicle is defined as follows:

$$Tr_n =< (S_1(\pm),T_1),(S_2(\pm),T_2),...,(S_k(\pm),T_k) > \quad (1)$$

, where S_i denotes the segment *ID* and T_i is the timestamp. According to T_i, the location of each vehicle can be retrieved easily.

Since the vehicles may move continuously or stay in a position, it is necessary to have the location knowledge of each vehicle according to the timestamps.

3.2 Congested routes detection

Definition 4 *Congested road segment*, the road segment is considered as congested region if the evaluated complexity value is higher than the predefined threshold value. The road segments with low moving speed and high number of vehicles are determined as congested road segments.

Fig. 2 shows the procedure of computing congested routes. The initial road information and trajectory data can be used to determine the existence of vehicles in each road segment of different directions. The road network is divided into road segments with different width, length, and directions. The trajectory data of moving objects on the road networks are analyzed to discover the congested routes. Here, the congested road segments are extracted by considering the average moving speed of the vehicles and the saturation degree of each road segment in the road networks. Finally, we compute the final congested routes by using a clustering scheme.

In this paper, the congested road segments are computed according to the different directions of the roads. The location and direction of each vehicle can be retrieved from

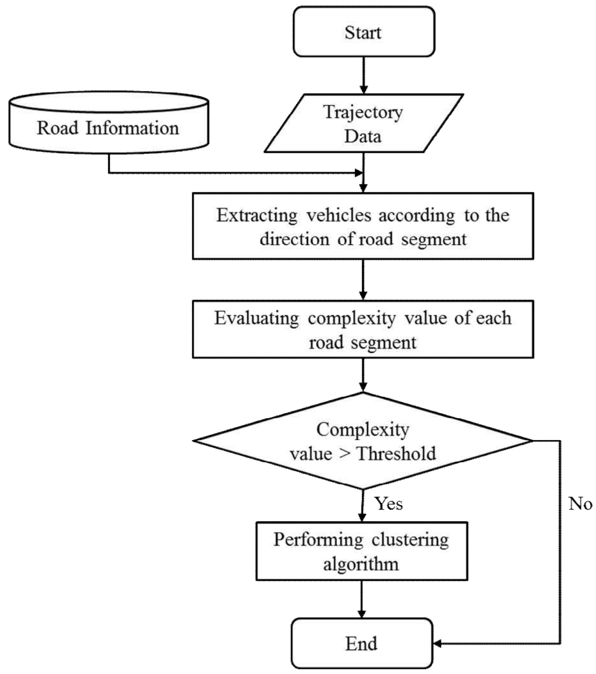

Fig. 2. The procedure of computing congested routes

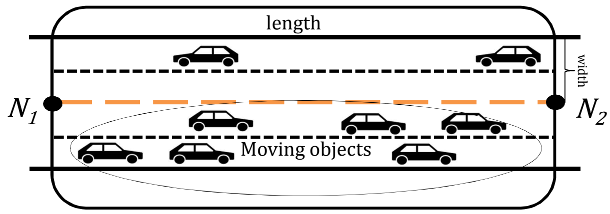

Fig. 3. The road segment between nodes N_1 and N_2 in a road network

the recorded trajectory data. The complexity value of each road segment is computed by considering the average speed of the vehicles in the road segment and the saturation degree of the road segment. The fast moving speed indicates that the congestion of the road segment is low. In contrast, the low moving speed indicates that the congestion of the road segment is high. The saturation degree is computed based on the number of the vehicles within a road segment and the length and width of a road segment, which are indicated in Fig. 3. We define that the congested road segments within a road network are the road segments with high complexity values.

The average moving speed (*Av*) of the vehicles in a road segment according to different directions is computed by the following Eq. 2, where $V(Ob_i)$ denotes the moving speed of a vehicle Ob_i. The saturation (*Sat*) according to the width (S_{width}) and length (S_{length}) of a road segment is computed in Eq. 3, where Ob_n denotes the number of the vehicles in a road segment. As a result, the complexity value of a road segment is computed by Eq. 4 which combines Eq. 2 and Eq. 3, α denotes the weight value between the average moving speed of vehicles and the

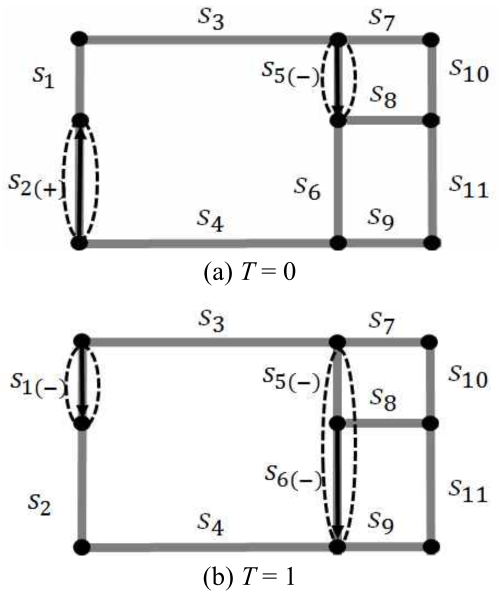

Fig. 4. The clusters according to different timestamps

saturation of a road segment.

$$Av_i(\pm) = \frac{\sum_{i=1}^{n} V(Ob_i)}{Ob_n(\pm)} \tag{2}$$

$$Sat_i(\pm) = 1 - \frac{Ob_n(\pm)}{S_{width} \times S_{length}} \tag{3}$$

$$Ri_i(\pm) = \alpha \times \frac{1}{Av_i(\pm)} + (1-\alpha) \times Sat_i(\pm) \tag{4}$$

The complexity value of each road segment is evaluated, and the clustering algorithm is performed among congested road segments. However, since the complexity values of road segments are changed according to different timestamps, it has to be computed periodically. The congested routes of a road network are computed according to the complexity values. Fig. 4 shows the congested routes (the dotted areas) of a road network in different timestamp T. According to the recorded trajectory data of vehicles at time T=0, the congested routes of the road network of different directions are generated, such as $S_2(+)$ and $S_5(-)$ in Fig. 4(a). As shown in Fig. 4(b), when time T=1, $S_1(-)$, $S_5(-)$ and $S_6(-)$ are evaluated as congested road segments. Since $S_5(-)$ and $S_6(-)$ are neighbor road segments and have the same direction in the road network, they are clustered together.

3.3 Congested routes detection algorithm

In this section, we present the algorithm of congested routes detection. The algorithm operates in two phases. In the first phase, the complexity value of each road segment in the road network according to different directions is computed. In the second phase, the congested routes of a road network are evaluated by clustering the congested road segments with each time interval. When the complexity value of a road segment is larger than the predefined threshold value the road segment is considered as a congested road segment. Finally, the congested road segments with same direction are clustered together. Fig. 5 shows the congested routes detection algorithm.

```
Congested routes detection algorithm
  Input :
  - Set of trajectories Tr = {tr1,tr2,...,trn}
  - Set of road segments S = {S1,S2,...,Sk}
  - Set of time intervals T = {t1,t2,...,tn} ∈ Tr
  - Threshold α : complexity value for determining congested road
    segments

  Output :
  - Set of cluster groups CH = {CH1,CH2,...,CHn}

  Algorithm :
  initialize CH, candidate list C;
  while there exists Tr in road segments list S
select segment Si in S according to each direction;
compute the complexity value v of Si ;
  end while

  for everyz time interval t ∈ Tr

while there exists Tr in segment list S
  foreach segment Si in S
     if complexity value v > α
        append Si to C;
     if there exists neighbor segment N in C
           create CH ← append N to candidate list C;
     else
           create CH ← candidate list C;
  end for
end while
end for
```

Fig. 5. The algorithm of the proposed scheme

4. Performance Evaluation

In this section, we introduce the performance evaluation by comparing the proposed scheme with the existing scheme NETSCAN [9]. The vehicles are generated by the network-based generator [21]. The complexity values and clusters are generated according to the number of vehicles in each time interval. All of the experiments are coded in Java and the experiments are performed in Intel i3 3.0GHz CPU and 4G memory. Table 1 summarizes the parameters for this performance evaluation.

Table 1. The values of parameters

Parameter	Value
Road network	Oldenburg city
The number of segments	7,035
The number of nodes	6,105
The number of vehicles	10,000~100,000
Velocity	0~120

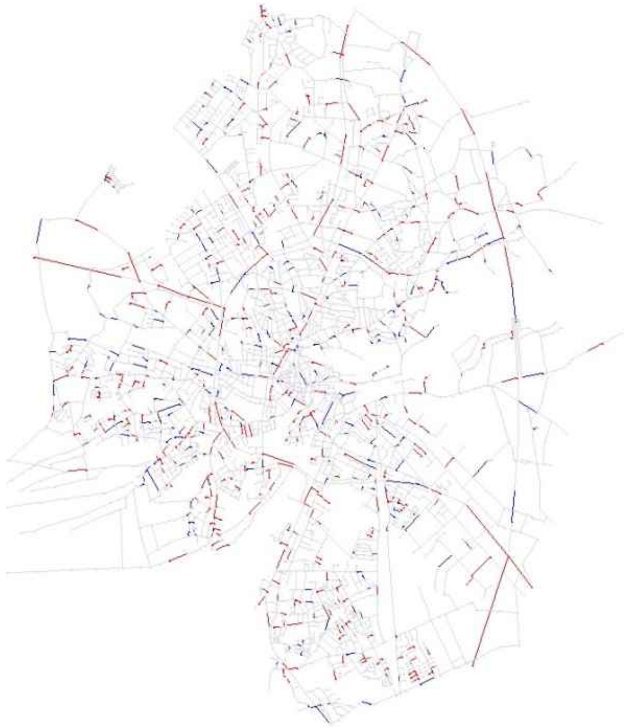

Fig. 6. The congested route of the Oldenburg city

In the first experiment, we show the congested routes road networks of Oldenburg city by using our proposed scheme. In this experiment, the total number of vehicles in the road network is set to 50,000 and the saturation of the road is set to 30%. As shown in Fig. 6, the results indicated that the congested routes of the road network are different according to the different directions of the road networks. The blue and red regions represent the congested routes of the road networks in positive direction and negative direction respectively.

In Fig. 7, we compare the NETSCAN scheme with our proposed scheme. The number of the congested routes of NETSCAN and the proposed scheme are evaluated according to the number of the vehicles. For the proposed scheme, the congested routes are evaluated in different directions (positive direction and negative direction) and same direction respectively. The results show that the number of the congested routes is increased when the number of the vehicles increases. The number of the congested routes of the proposed scheme is similar when the number of the vehicles between 20,000 and 30,000. This is because the saturation of each road segment is considered in the proposed scheme, when the width and length of a road segment is large the 20,000 and 30,000 vehicles is not large for the road. Therefore, most of the road segments are not identified as congested routes at first. For NETSCAN scheme, the number of congested routes is increased proportionally with the increase of the number of vehicles.

Fig. 8 shows the number of congested routes according

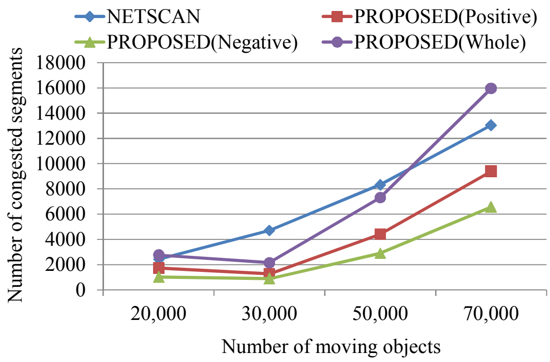

Fig. 7. The number of congested routes according to the number of vehicles

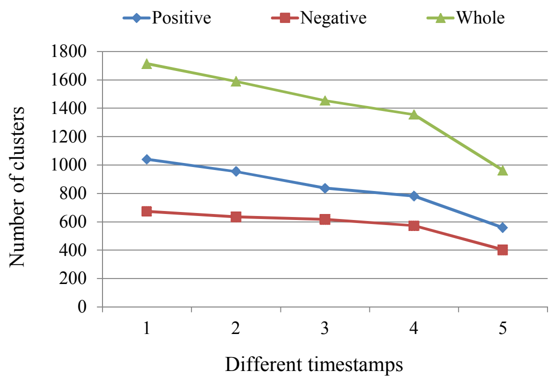

Fig. 8. The number of congested routes according to the different timestamps

to the different timestamps, the time interval is set as 1 hour. The congested routes are evaluated in different directions (positive direction and negative direction) and same direction respectively. From Fig. 8, we can see that the number of the congested routes of the positive direction is larger than that of the negative direction. And the number of the congested routes that without considering the direction of the road segment is larger than the direction considered scheme.

5. Conclusion

In this paper, we have proposed a congested routes discovering scheme in real road networks. The proposed scheme divides the road into segments with different widths and lengths. It extracts the congested road segments based on the average speed of the vehicles and the saturation degree of a road segment. The final congested routes are computed by performing clustering scheme. The experimental results showed that the proposed scheme can discover the congested routes in different directions over the existing schemes. In the future, we will conduct more performance evaluation of our approach by using the real trajectory data of vehicles.

Acknowledgements

This research was financially supported by the Ministry of Education(MOE) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation(No. 2013H1B8A2032298), by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2015-H8501-15-1013) supervised by the IITP(Institute for Information & communication Technology Promotion), by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIP) (No.2013R1A2A2A01015710), and by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No.2014R1A1A2055778).

References

[1] Z. Chen, H. S. Shen and X. Zhou, "Discovering Popular Routes from Trajectories," *in Proceedings of IEEE Data Engineering*, pp. 900-911, 2011

[2] P. C. Mauroux, E. Wu and S. Madden, "TrajStore: An Adaptive Storage System for Very Large Trajectory Data Sets," *in Proceedings of IEEE Data Engineering*, pp. 109-120, 2010

[3] J. W. Chang, M. S. Song and J. H. Um, "TMN-tree: New Trajectory Index Structure for Moving Objects in Spatial Networks," *in Proceedings of IEEE Computer and Information Technology*, pp. 1633-1638, 2010

[4] M. Huang, P. Hu and L. Xia, "A Grid based Trajectory Indexing Method for Moving Objects on Fixed Network," *in Proceedings of International Conference on Geoinformatics*, pp. 1-4, 2010

[5] J. I. Won, S. W. Kim, J. H. Baek and J. H. Lee, "Trajectory Clustering in Road Network Environment," *in Proceedings of IEEE Computational Intelligence and Data Mining*, pp. 299-305, 2009

[6] G. P. Roh, J. W. Roh, S. W. Hwang and B. K. Yi, "Supporting Pattern Matching Queries over Trajectories on Road Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1753-1758, 2011

[7] X. Li, J. Han, J. Lee and H. Gonzalez, "Traffic Density-Based Discovery of Hot Routes in Road Networks," *in Proceedings of SIAM International Conference on Data Mining*, pp. 441-459, 2007

[8] N. Pelekis, I. Kopanakis, E. E. Kotsifakos, E. F. Frentzos and Y. Theodoridis, "Clustering Trajectories of Moving Objects in an Uncertain World," *in Proceedings of IEEE Data Mining*, pp. 417-427, 2009

[9] A. Kharrat, K. Zeitouni and I. Sandu-Popa, "Characterizing Traffic Density and its Evolution through Moving Object Trajectories," *in Proceedings of International Conference on Signal Image Technology and Internet Based Systems*, pp. 257-263, 2009

[10] H. M. O. Mokhtar, O. Ossama and M. E. Sharkawi, "A Time Parameterized Technique for Clustering Moving Object Trajectories," *Journal of Data Mining and Knowledge Management Process*, vol. 1, no. 1, pp. 14-30, 2011

[11] A. Selamat, M. Z. Arokhlo, S. Z. M. Hashim and M. H. Selamat, "A Fast Path Planning Algorithm for Route Guidance System," *in Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp.2773-2778, 2011

[12] M. Kimura, S. Inoue, Y. Kakuda and T. Dohi, "A Route Discovery Method for Alleviating Traffic Congestion Based on VANETs in Urban Transportations Considering a Relation between Vehicle Density and Average Velocity," *in Proceedings of International Symposium on Autonomous Decentralized Systems*, pp. 299-302, 2011

[13] H. Gonzalez, J. Han, X. Li, M. Myslinska and J. P. Sondag, "Adaptive Fastest Path Computation on a Road Network: A Traffic Mining Approach," *in Proceedings of International Conference on Very Large Data Bases*, pp. 794-805, 2007

[14] I. Abraham, D. Delling, A. V. Goldberg and R. F. Werneck, "Alternative Routes in Road Networks," *ACM Journal of Experimental Algorithmics*, vol. 18, 2013

[15] L.Y. Wei, W.C. Peng and W.C. Lee, "Exploring Pattern-Aware Travel Routes for Trajectory Search," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 3, pp. 48, 2013

[16] E. Kanoulas, Y. Du, T. Xia and D. Zhang, "Finding Fastest Paths on A Road Network with Speed Patterns," *in Proceedings of International Conference on Data Engineering*, pp.10, 2006

[17] G. Gidófalvi, M. Kaul, C. Borgelt and T. B. Pedersen, "Frequent Route based Continuous Moving Object Location- and Density Prediction on Road Networks," *in Proceedings of ACM International Workshop on Advances in Geographic Information Systems*, pp. 381-384, 2011

[18] Y. Xu and J. Wang, "Optimal Path Solution of Urban Traffic Road," *in Proceedings of International Conference on Natural Computation*, pp. 799-802, 2011

[19] K. W. Min, J. W. Kim and J. H. Park, "Optimal Route Determination Technology Based on Trajectory Querying Moving Object Database," *in Proceedings of International Conference on Database and Expert Systems Applications*, pp.666-675, 2006

[20] V. Tyagi, S. Kalyanaraman and R. Krishnapuram, "Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1156-1166, 2012

[21] T. Brinkhoff, "A Framework for Generating Network-Based Moving Objects," *GeoInformatica*, vol. 6, no. 2, pp. 153-180, 2002

He Li He received his B.S. degree in the department of Computer Science from Yunnan University in 2006. He received his M.S and Ph.D degrees in the department of Information and Communication Engineering at Chungbuk National University in 2010 and 2014. His main research includes database, graph data, data mining, and social networks.

Kyoung Soo Bok He received his M.S. and Ph.D. in Computer and Communication Engineering from Chungbuk National University, Korea in 2000 and 2005. He is now a research professor in Information and Communication Engineering, Chungbuk National University, Korea. His research interests are social network services, mobile ad-hoc networks, and big data.

Jong Tae Lim He received his M.S. in Computer and Communication Engineering from Chungbuk National University, Korea in 2011. His research interests are database, social network services, and big data.

Byoung Yup Lee He received his M.S. in Computer Science and Ph.D. in Management Information System from the Korean Advanced Institute of Science and Technology, Korea in 1993 and 1997. He is now a professor in Electronic Commmerce, Paichai University, Korea. His main research interests include database and e-commerce.

Jae Soo Yoo He received his M.S. and Ph.D. in Computer Science from the Korean Advanced Institute of Science and Technology, Korea in 1991 and 1995. He is now a professor in Information and Communication Engineering, Chungbuk National University, Korea. His main research interests include sensor data management, big data, and mobile social networks.