

A Road Network Construction Method Based on Clustering Algorithm

Xuelian Chen¹²

*Dept. of Information Engineering , Liaoning Provincial College of Communications
Shenyang, China
E-mail: xuelian_chen@163.com*

Qingchun Li

*State Grid Liaoning Electric Power Supply Co., Ltd.
Shenyang, China*

Based on clustering algorithm, a road network construction method is presented to solve the existing problems underlying the road network structure, such as low efficiency, high input, slow update, and so on. In this paper, the clustering algorithm is designed to detect the crossings of the road from the turning point of the vehicle by using the taxi GPS trajectory data. These crossings are connected with the line to form a road network topology. Results show that the proposed method can detect most locations of the road crossings, and is a kind of relatively reasonable extraction method for the road network construction.

*ISCC 2015
18-19, December, 2015
Guangzhou, China*

¹Xuelian Chen

²This study is supported by the general research project of Liaoning Province Education Department (L2014512) and the excellent talents development support project of Liaoning Province Communications College (LNCCRC201413)

1. Introduction

With the development of GPS devices, a large number of GPS trajectory data can be obtained from the assembly of vehicles and smart phones equipped with GPS devices. The GPS trajectory data include abundant road information, such as length of road section, moving speed, moving direction, staying time, geographical position of vehicle, and so on[1]. The road network structure can be extracted from GPS trajectory data, and the mobile object's behavior pattern can also be extracted.

Many researches have touched upon the automatical extraction of road network structure based on the GPS trajectory data, and presented fruitful results. Schroedl et al. first used the K - mean algorithm to solve the problem in the trajectory assignment[2]. The algorithm randomly selects some points as the center of the cluster. Then these GPS points are added to the cluster, in which the clustering center and direction are calculated. However, the initial clustering center method is adopted in the algorithm, and can not guarantee that the final clustering seeds can cover all the roads. Bruntrup et al. and Davics et al. used the density algorithm to realize the GPS trajectory data cluster, in which the point sequence is transformed into a serial sequence number[3-4]. Worrall and Nebot adopted the clustering method to allocate trajectories[5], but they use straight line segments and arc segments to simulate a section. Rakshit et al. used the clustering method based on the segment density[6], which divides the trajectory segments into different clusters and uses the clustering sequence to express the trajectory. Shi et al. and Cao et al. used the grid clustering method to cluster a region into some non-overlapping rectangular regions, and expressed the trajectory by using the region sequence[7-8]. Goh et al proposed a clustering algorithm based on the basic unit of the line segment between two GPS points[9].

These methods mostly focus on the road connectivity or topology, ignoring the road network's position in an important part of intersection. Crossings not only reflect the connectivity between the road, but the complex changes in the network, such as lane collection and separation, lane between the connected relationship, different sections of the passage length. In order to solve these problems, this paper presents a method of road network construction based on the clustering algorithm.

2. The Road Network Construction Method

The shape and topology of the road can be reflected according to the GPS trajectory data because the vehicles and other moving objects are usually moving along the road. The closer to the road those vehicles are, the denser the GPS trajectory data are. Therefore, it is feasible to create a road network topological structure by using the clustering algorithm to transfer the GPS trajectory data to different clusters.

In this paper, a new road network construction method is proposed based on the clustering algorithm. The basic idea of this method is: first, most of turning points are located nearby the crossings when the vehicles are running on the road. The turning points of those vehicles can be found from the GPS trajectory data, and these points are distributed into different clusters according to the clustering algorithm, so each cluster corresponds to one crossing. Second, the vehicle traffic flowing in the crossing can be detected based on the cluster. It is true that a road exists when the traffic flow value is greater than the certain threshold. The road section can be detected by connecting two crossings. Finally, the road network topological structure can be constructed by connecting all identified road sections.

2.1 Searching Cluster Point

In line with the characteristics of the GPS trajectory data, most vehicle turning points occur near to the crossings. The crossings location can be detected based on the clustering algorithm. The vehicle turning points in the GPS trajectory data are evaluated by computing the change of the direction angle between the two continuous GPS data in one vehicle. The point will be a

turing point if the direction angle is larger than the certain angle. Meanwhile, data of those turning points usually scatter around the crossing. However, these data are not at the center of the crossing for the characteristic of coarse-grained GPS data and traffic rules. In order to improve the processing precision of clustering algorithm, concentration of vehicle turning points needs to be performed. All the vector intersection point of turning points are located in the center of the crossing because the real turning location of the vehicle GPS trajectory should be located in the center of the crossing. The method of computing vector intersection point between turning points in the GPS trajectory data is shown in Fig. 1. It can be seen from Fig. 1 that the point C is the vector intersection of point A and point B, I.e., it is located in the center of the crossing.

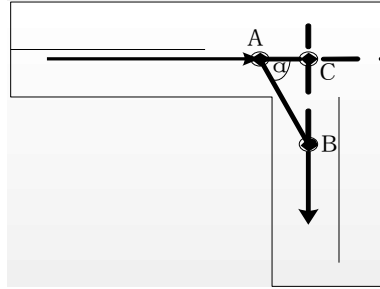


Figure 1: Extract Converging Points from the Turning Point

2.2 Detecting Crossings

There are lots of effective algorithms, such as the clustering algorithm based on the hierarchical and partitioning net, the clustering algorithm based on the the hypergraph model, the clustering algorithm based on the grid and the clustering algorithm based on density. Four questions need to be considered when the problem of crossing detection is solved by the clustering algorithm. Firstly, the selected clustering algorithm should automatically determine the number of clusters because the distribution of GPS trajectory data is not priori knowledge and the number of clusters in the target area is not determined. Secondly, for most turning points distributed around the crossings, the shape of each cluster should be similar to a circle or oval. Thirdly, the size of each cluster should not be too large in order to avoid covering multiple crossings. Finally, the distance of any two clustering centers can not be too close, because the two crossings are almost not adjacent in an actual road network.

The X-means clustering algorithm can gather the original complex data into one cluster, which is based on the similarity of the samples[10]. Then the whole data space is divided into plenty of local data regions by using X-means clustering algorithm. The X-means clustering algorithm is an improved one of the k-means algorithm, which only needs to be given a range of $[K_{min}, K_{max}]$. The optimal cluster number k can be automatically determined by using the BIC (Bayesian Information Criterion) model, which can greatly accelerate the clustering process. However, it is difficult to be consistent with the actual distribution of crossings by using the X-means algorithm. Therefore, this paper adopts the improved X-means algorithm to detect the crossings position. The algorithm steps are as follows:

- Step1:** computing initial clusters;
- Step2:** training the model;
- Step3:** if $K > K_{max}$ or there is no cluster to be split during the search, then go to Step4; else go to Step1;
- Step4:** merging clusters;
- Step5:** improving clusters;

In the improved X-means algorithm, Step 1 is an iterative process of the K-means algorithm, in which K_{min} is selected as the initial centroid, then every turning point is assigned to the closest centroid that becomes K_{min} initial cluster. Each clustering centroid is recomputed

until the clusters are not changed. Step 2 is the segmentation strategy for clusters with the BIS model to choose an optimal model. The segmentation rules are shown in the formula (2.1):

$$\begin{aligned} r_i > R_{up} \text{ and } d_i > D_{up} \\ R_{low} < r_i < R_{up}, D_{low} < d_i < D_{up}, \text{ and } b_{p_i} < b_{c_i} \end{aligned} \quad (2.1)$$

where r_i is the clustering radius of parent cluster; d_i is the distance of the two sub cluster centers; R_{low} and R_{up} are the upper bound and lower bound of r_i respectively; b_{p_i} and b_{c_i} are the BIC values of parent cluster and sub cluster respectively; D_{low} and D_{up} are similar to R_{low} and R_{up} that are used to ensure the segmented sub cluster with a crossing.

The parent cluster will be segmented when one condition is met in the above two conditions. There is a certain limit to the size of any crossings according to the common sense, so the cluster may cover multiple crossings if the clustering radius is larger than the certain value. It means that the cluster covers multiple crossings if the distance of two clusters is larger than the certain value. Some clusters may still cover multiple crossings if the values of R_{up} and D_{up} are too large. They may be segmented that convert the same crossings if the values of R_{up} and D_{up} are too small.

Step 4 will merge adjacent clusters. Some clusters may be adjacent in the procedure of clustering for the condition that the clustering distance should be larger than the certain value is not designated. All adjacent clusters should merge into one cluster.

Step 5 will remove all pseudo-crossings. The method of removing pseudo-crossings is to assess the number of trajectory points in each cluster. If the number is lower than the certain threshold, then the crossing is a pseudo-crossing and should be removed.

2.3 Constructing the Road Network Topology

It is essential to compute the number of trajectory points between any two crossings after the crossing locations are determined. The number is used to examine if there are sufficient common GPS trajectory data between two crossings. The crossings set is shown as follows:

$$p = \{p_i \mid i = 1, 2, \dots, n\} \quad (2.2)$$

All trajectories C_k passing the crossing are arranged in a sequence S_k according to their chronological order. S_k is presented as follows:

$$S_k = \{\omega_k p_k \mid k = 1, 2, \dots, n; p_k \langle \rangle p_{k+1}\} \quad (2.3)$$

Then the number of the common trajectories of C_k is added with 1 between the crossings of P_k and P_{k+1} . There is one road section between two crossings if the number of common trajectories is larger than the threshold. The road network topological structure can be constructed after all crossings between the road sections are connected.

There are some questions during the construction of the road network. In the Fig. 2, the crossing B is connected with the point A and C, and the length of the road section AB and AC is short. There are two GPS trajectory data located at A and C when the vehicle is running from A to C. The crossing B may be ignored when the common trajectory is computed. AC will be treated as one road section. Which road section can be treated as the real road section between

the road section AC and ABC? The method is to compare the vehicle flow of the two sections when the vehicle flow is larger and the distance is same. The road section AC can be take as one road section.

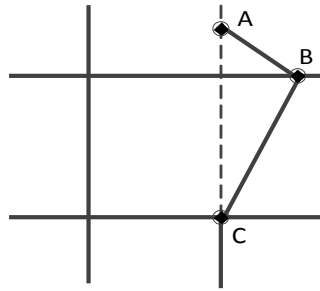


Figure 2: the Recognized Incorrect Road Section AC

3. Experimental Result Analysis

In order to validate the method of road network construction, some of taxi GPS trajectory data of Shenzhen city are used as the experimental data from 8am to 9am on Sep.1, 2010 with 7494 taxis and 857711 trajectory data. The improved clustering algorithm is used to analyze the GPS trajectory data.

The number of clusters is shown in Fig. 3 where k_{min} has different values. From Fig. 3, it can be seen that the number of the algorithm is stable at 191 when the number of different values is identified. The algorithm can identify the appropriate number of crossings and the stable crossings position. The road network topological structure is constructed with the method (see Fig. 4). In the constructed road network, some road segments are not recognized, which are shown in Fig. 5. In the Fig. 5, the region A and region B are the unrecognized road segments.

Results show the method is more suitable for the road network construction that contains the straight road. If there are most curves in the real road network, the direction of the vehicle will always change when it runs on the curve road. The turning points will be uniformly distributed on each curve road, and it will not be concentrated around the crossing. Under this condition, effectiveness of this method will be greatly reduced.

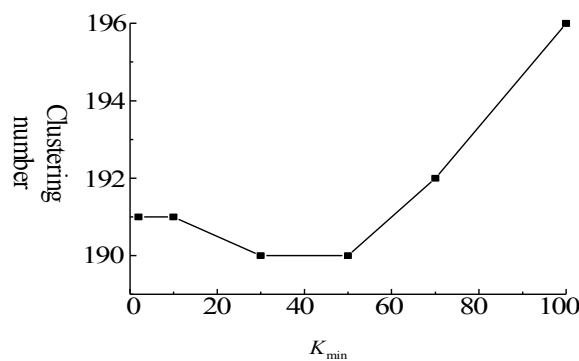


Figure 3: the Constructed Road Network

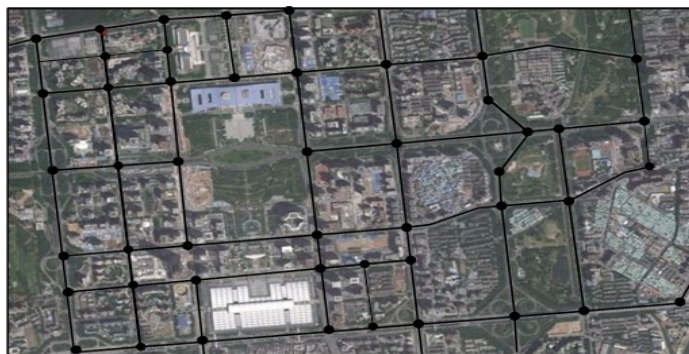


Figure 4: the Constructed Road Network



Figure 5: Unrecognized Road Segements

4. Conclusion

Based on the improved clustering algorithm, a new method for the road network construction is introduced in this paper. According to the experimental result, the optimal road network topoglical structure is constructed by using this method. Results show that this method is suitable for regular road network. Moreover, with this method,optimal topology and high precision of the road network can also be obtained when there are fewer curves, circles etc. in the actual road network, and iffERENCE in the road width is small and distance between crossings is nearly same.

References

- [1] Yuan J, Zheng Y, Xie X. *Discovering regions of different functions in a city using human mobility and POIs*[C]. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, USA, 2012: 186-194
- [2] Schroedl S, Wagstaff K, Rogers S, *et al.* *Mining GPS traces for map refinement*[J]. Data mining and knowledge Discovery, 2004, 9(1): 59-87
- [3] Bruntrup R, Edelkamp S, Jabbar S, *et al..* *Incremental map generation with GPS traces*[C]. Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE , Vienna, Austria, 2005: 574-579
- [4] Davics J J, Beresford A R, and Hopper A. *Scalable, distributed, real-time map generation*[J]. Pervasive Computing, IEEE, 2006, 5(4): 47-54

- [5] Worrall, S., and E. Nebot. Automated Process for Generating Digitised Maps Through GPS Data Compression[C]. Australasian Conference on Robotics and Automation, Brisbane, Australia, Australian Robotics and Automation Association, Sydney, Australia, 2007 :1-6
- [6] Rakshit S, Ghosh A, Shankar B U. *Fast mean filtering technique (FMFT)*[J]. Pattern Recognition, 2007, 40(3): 890-897
- [7] Shi W, Shen S, Liu Y. *Automatic generation of road network map from massive GPS, vehicle trajectories*[C]. Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on. IEEE, St. Louis, MO, USA, 2009: 48-53
- [8] L. Cao, J. Krumm . *From GPS traces to a routable road map*[C]. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, Seattle, Washington, 2009: 3-12
- [9] Goh C Y,Dauwels J,Mitrovic N,et al. *Online map-matching based on hidden markov model for real-time traffic sensing applications*[C].2012 15th International IEEE Conference on Intelligent Transportation Systems, IEEE,Piscataway,USA,2012: 776 - 781
- [10] Chen Z, Shen H T, Zhou X. *Discovering popular routes from trajectories*[C]. Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, IEEE, Hannover, Germany, 2011: 900-911