ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DIVIDING AND CLUSTERING ALGORITHMS OF MOVING OBJECTS IN RFID TRACING SYSTEM

[1]KONGFA HU, [2]XIANGQIAN XUE, [3]JIANFEI GE, [4]YAN SUN, [5]LING CHEN
[1]College of Information Engineering, Yangzhou University, Yangzhou 225127,China
E-mail: [1]kfhu05@126.com

ABSTRACT

Trajectory clustering can predict moving trend of objects effectively. The traditional trajectory clustering algorithms take moving trajectory of a whole object as a research object, which will lose similar sub-trajectories. However, in practical applications, such as in RFID system, the users may only focus on some specific regions of trajectories. We propose PT-CLUS algorithms in this paper, according to coarse-fine algorithm, which first dividing a trajectory into a group of line segments and prunes by coarse-fine strategy, and then searching cluster in the sub-trajectories by checking neighborhood region of segments, using hierarchical clustering to accomplish the clustering of sub-trajectories. Experiment result shows that PT-CLUS algorithm can find the similar sub-trajectories from RFID trajectory database effectively.

Keywords: *Moving Objects; RFID; Line Segment; Sub-Trajectories; Clustering*

1. INTRODUCTIONS

Existing positioning services mostly include GPS system based on satellite positioning, positioning system based on infrared ray or ultrasonic and positioning system based on mobile network. The popularization of RFID[1] provides a new solution for space orientation and tracing service of person and objects. The solution utilize property， a tag is the only identification of a object and signal intensity of radio frequency communication between reader-writer and tag on object to measure the spatial location of object. These applications make it possible to collect the large quantity spate-temporal data of the moving objects trajectories, such as traffic control data, atmosphere information data, and animal moving data. One of the typical targets for data analysis is clustering, extracting moving characteristic patterns and predicting the behavior of moving objects by using the trajectory direction, movement, the relationship and the other typical characteristics of the moving objects trajectories.

In recent years the trajectory clustering problem has attracted the attention of many researchers, they put forward a lot of moving objects trajectory model and clustering method. In the literature [2, 3], Gaffney et al proposed a clustering method based on prototype of continuous trajectory. The method use mixed regression model for trajectory modeling, using a maximum likelihood principle to realize unsupervised learning. Especially the use of EM (Expectation-Maximization) algorithm to determine the cluster number of clustering .This algorithm clustering by taking track as a whole, that is take the object trajectory as a whole to deal with, so the basic clustering unit is a trajectory, which can not find similar sub tracks, a trajectory through the path is long and complex, many track may have a short similar, however, it is not similar in overall.

In the first step, we can use a coarse-fine level classification method to divide a trajectory into a group of sub-trajectories, which is coarse division of the trajectory according to the principle of MDL (Minimum Description Length) [4], pruning according to the distance between the sub-segments, and then take coarse-grained sub-segments to fine division. In the second step, we can use the density-based clustering algorithm to cluster a collection of trajectory segments. The PT-CLUS algorithm can not only dig out some similar segments of trajectory but also reduce the computational overhead by level division, improve the classification speed.

2. TRAJECTORY DIVISIONS

The purposed trajectory classification is to identify the significant changed points in the direction of trajectory, which called the feature points. To speed up the rate of clustering, this paper

Journal of Theoretical and Applied Information Technology
15th November 2012. Vol. 45 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

uses a segmentation method of coarse-fine level. First, according to the MDL principle, the trajectory is divided into the coarse sub-segments, which may meet the requirements of sub-clustering. So we can reduce the fine division of the data space, and then follow the basic unit, the coarse sub-segments will be further divided into the fine coarse segments, which cluster for the next stage.

2.1 Coarse-Grained Level-Coarse Division

In order to ensure the accuracy and simplicity, this paper uses MDL principle to find the best coarse classification.

Algorithm 1: Coarse_partition

Input: Trajectories set T={TR1,TR2,…,TRnumT}; TRi = p1p2p3⋯pi⋯pleni

Output: Sub-section trajectories set C={L1,L2,…,LK}

Method:

CP=NULL;

for each TRi ∈T do

{ StarIndex=1; length=1;

CP i={ p1};

while StartIndex+length≤leni do

{ curr=StarIndex+length;

compute L(H) and L(D|H) of line segments pstartIndex pcurrIndex for no partitioning;

s= L(H) + L(D|H);

compute L(H) and L(D|H) of line segments pstartIndex pcurrIndex for partitioning;

s'= L(H) + L(D|H);

if s< s' then

{ CP i= CP i+{ pcurrIndex-1 };

StartIndex=curr-1; length=1; }

else length=length+1; }

CP i= CP i+{ pleni } ; CP = CP +CP i; }

each line segment L in C denoted as pjpj+k*b in CP (b is the base unit);

2.2 A Fine-Grained Level - Fine-Division

At this stage, we advance a pruning strategy through judging the section distance between the upper and lower bounds which reduces the computational expense. Reference [6] defined the upper and lower bounds of distance function, if lb (Li, Lj , dist) > ε (ε is the distance threshold), when compared to coarse section Li and Lj which do not need to be fine-grained classification; if ub(Li, Lj , dist) ≤ ε , after dividing Li and Lj by the fine division, we do not need to calculate the distance between the fine segment.

Algorithm 2: prune_segments

Input: Coarse segments-set C={L1,L2,…,LK}.

Output: Fine segments-set F={l1,l2,…ln}.

Method:

F=null;

For each pair of Li, Lj ∈C(Li≠ Lj) do

{ if lb(Li, Lj , dist) >ε then break;

else fine partition Li ,Lj , and insert fine line segments into F;

if ub(Li, Lj , dist) ≤ε then

∀ li ∈Li and ∀ lj ∈Lj

N(li) = N(li) + {lj}; N(lj) = N(lj) + {li };

else for each pair of li ∈Li , lj ∈Lj do

if dist(li,lj) ≤ε then

do as line [6],[7]

3. TRAJECTORY CLUSTERING

On the trajectory segmentation segment clustering by using hierarchical clustering method, the Chameleon algorithm is a two class through the merger with higher standards to improve the quality of clustering algorithm, which considers both the interconnect, and the degree of approximation. In particular cluster internal features, which can automatically adapt to the merged cluster the interior features, therefore it find the arbitrary shape and arbitrary size cluster capacity. The algorithm is firstly constructed as a K-nearest neighbor graph Gk by the data set, then a graph partitioning algorithm for map Gk is divided into many sub graph, each sub-graph represents an initial sub-variety, finally find the real results cluster by an agglomerative hierarchical clustering algorithm iteratively merging sub cluster .

Application of agglomerative hierarchical clustering method on a stage before the algorithm generated initial sub-variety merged to form the final clusters. Decide the similarity between the two

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

clusters through the sub clusters Ci and Cj relative interconnectivity and the relative degree of approximation, choose two sub clusters which have the maximum by using the similarity function RI(Ci,Cj)*RC(Ci,Cj) to merge.

Definition 1 (Relative interconnectivity RI (Ci,Cj)). Sub-clusters Ci and Cj are absolute connectivity on the internal standardization of connectivity between two clusters. That is

$$RI(C_i,C_j)=\frac{|EC(C_i,C)|}{\frac{1}{2}(|EC(C_i)|+|EC(C_j)|)} \quad (1)$$

EC(Ci) is the sum of the weight of the edges necessary to be removed when Ci make minimum truncation, EC(Ci,Cj)is the sum of the weight of the edges connecting sub-clusters Ci and Cj.

Definition 2 (Relative approximation RC(Ci,Cj)). Sub-clusters Ci and Cj are absolute approximation on the internal standardization of approximation between two clusters. That is

$$RC(C_i,C_j)=\frac{TC(C_i,C_j)}{\frac{|C_i|}{|C_i|+|C_j|}\times TC(C_i)+\frac{|C_i|}{|C_i|+|C_j|}\times TC(C_j} \quad (2)$$

TC(Ci) is the average weight of the edges necessary to be removed when Ci make minimum truncation.

Algorithm 3 : PT-CLUS

Input: a collection of trajectories T={TR1,TR2,…,TRnumT}; TRi = p1p2p3⋯pi⋯pleni.

Output: a collection of clustering O= {C1,…,CnumC}.

Method:

For each TRi ∈T do

execute coarse_partition; get a set of coarse line segments C;

execute prune_segments; get a set of fine line segments F;

CreatKnnGraph();

Partition (CLUSTERING C)

{Bisection(C, C1, C2, min_ctw, min_sum) }

Merge(C[ ],C1[ ])

{compute similarity(C1,C2)

{abs_conclose(C1,m11,m12);

abs_conclose(C2,C21,m22);

inter_conclose (C1,C2,n1,n2)

k1=len(C1); k2=len(C2);

r1=2*n1/(m11+m21);

r2=(k1/(k1+k2))*m12+(k2/(k1+k2))*m22;

sim=r1*r2;

return sim; } }

4. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental hardware platform is Pentium IV, CPU 1GHz, memory for 512MB, Windows XP Professional SP2 operating system, using TRACLUS [6] comparative experiments. Because of practical data acquisition is difficult; we use a set of simulated data, i.e., a series of spatial position on sample data.

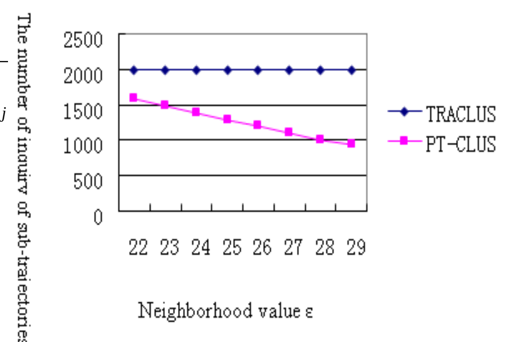

Figure 1. The Number Of Queries Of Neighborhood Sub-Segments

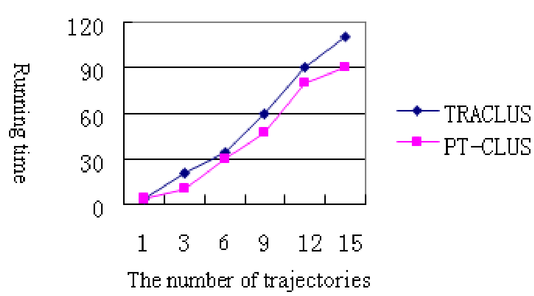

Figure 2. The Algorithm Running Time Analysis

Figure 1 shows that PT-CLUS algorithm greatly reduces require for TRACLUS neighborhood segmented queries when we keep MinLns unchanged. Because TRACLUS needs to query each line segment neighbor, so no matter how to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

change the value, ε neighborhood segment query frequency remains constant. But for PT-CLUS algorithm, due to use the coarse-fine segmentation method, in the coarse level query segmentation and pruning to reduce the query frequency on neighborhood segment of fine-grained, with the epsilon value increasing, the frequency for PT-CLUS neighborhood query operation is linearly reduced and scalability is good.

Figure 2 shows PT-CLUS algorithm; the running time is less than TRACLUS algorithm. As in the stage of Subsection, PT-CLUS algorithm uses coarse-fine pruning strategy, reducing the number of segment of the distance between the computational overhead, narrowing track segment set and also making it easy to cluster, reducing the running time of the algorithm. But TRACLUS uses DBSCAN algorithm to cluster similar sub tracks, more sensitive to parameters, the clustering results are influenced seriously by parameter. PT-CLUS uses clustering method based on hierarchical, clustering results are affected by parameter lowly, also can find high quality clusters of arbitrary shape.

5. CONCLUSIONS

RFID technology has been applied to the moving object tracking system, by clustering the trajectory of moving objects and extract based on the motion trajectory clustering and extract the moving characteristic patterns and predict the object's motion behavior. Due to the efficiency for the existing rail clustering algorithm is not high, therefore facing the low granularity and large orbital data set, the algorithm usually can not work effectively. This paper proposes the trajectory clustering algorithm PT-CLUS for moving objects based on the analysis of TRACLUS algorithm.

In order to find out the similar sub-trajectories, we propose a kind of classification and clustering method for object trajectory. Firstly, a long path is partitioned into a set of segments in the feature point, and then we adopt coarse-fine segmentation method to reduce the search space and accelerate clustering phase velocity. Then, according to the dynamic modeling for determining similarity among clusters, we use a hierarchical clustering method to cluster.

ACKNOWLEDGMENT

The research in the paper is supported by the National Natural Science Foundation of China under Grant No. 61070047, 61003180; the "Six Talent Peaks Program" of Jiangsu Province of China; the "333 Project" of Jiangsu Province of China.

REFERENCES

[1] D. Roozbeh, E. O. Maria and X. Li. Retrieving frequent walks from tracking data in RFID-equipped warehouses. Proceedings. of the 2007 IEEE International Conf. on RFID Gaylord Texan Resort. Grapevine, TX., USA, 2008: 175 - 182.

[2] S. Gaffney and P. Smyth. Trajectory Clustering with Mixtures of Regression Models. Proceedings of the 5th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, San Diego, California，1999: 63-72..

[3] S. Gaffney , A. Robertson, P. Smyth, S. Camargo and M. Ghil. Probabilistic Clustering of Extratropical Cyclones Using Regression Mixture Models, Technical Report UCI-ICS 06-02, University of California, Irvine, Jan. 2006.

[4] P. GrÄunwald, I. J. Myung and M. Pitt. Advances in Minimum Description Length: Theory and Applications, MIT Press, 2005.

[5] J. Chen, M. K. H. Leung and Y. Gao. Noisy Logo Recognition Using Line Segment Hausdorff Distance. Pattern Recognition, 2003, 36(4): 943-955.

[6] L. Jae-Gil, J. Han and W. Kyu-Young. Trajectory clustering: A partitionand-group framework. Proceedings. of the ACM SIGMOD Conf. On Management of Data, Beijing, China, 2007: 593－604.