International Journal of Database Theory and Application
Vol.7, No.2 (2014), pp.37-44
http://dx.doi.org/10.14257/ijdta.2014.7.2.04

Mining a New Movement Pattern in RFID Database on Internet of Things

Kongfa Hu[1] and Long Li[2]

[1]*College of Information Technology of Nanjing University of Chinese Medicine, Nanjing, China*
[2]*College of Information Engineering of Yangzhou University, Yangzhou, China*
kfhu05@126.com

Abstract

In the last few years, RFID is commonly used in many related fields of new service application and new study as one of the significant technological advancements, such as science manufacturing, logistics transportation, traffic management, medical information, and so on. Those intelligent and automatic innovative products gradually take the place of manpower. Due to low cost and remote automatic identification advantages, RFID technology seems to be a popular current trend. In RFID database, there is a vast number of RFID trajectory data with the spatial-temporal characteristic. How to extract the traveling partners from these data is a difficult problem. For solving the difficult problem, we proposed an algorithm called MTP to discovery the traveling partners from RFID database, it uses a intersecting method to mine moving objects with moving together in a period of time. Meanwhile, we analyze the performance of MTP, the result of our experiment demonstrate that the MTP algorithm is suited to mine the traveling partners.

Keywords: *radio frequency identification, RFID trajectory data, traveling partners, MTP algorithm*

1. Introduction

In recent years there has been an increase in the location-aware devices (*e.g.*, RFID, GPS, and so on) and wireless communication networks. This has led to a large amount of trajectory data capturing the movement of animals, vehicles, and people [1]. Radio Frequency Identification (RFID) is a technology of automatic object identification that uses radio waves to transfer data from an electronic tag, called RFID tag or label, attached to an object through a reader for the purpose of identifying and tracking the object [2]. As the rapid development of RFID, all kinds of wireless sensor devices and positioning equipment have been widely used, such as scientific research, logistics management, traffic management, medical information, tracking management and security monitoring [3], because the low cost and easy deployment of RFID technology, it gets more people favor [4]. As RFID technology matures, widely deployed RFID devices produce large amounts of RFID trajectory data every day, how to effectively and efficiently extract important information for us from these RFID trajectory data in RFID database becomes extremely important.

In these information, people greatly concerns about the moving object groups with moving together in the existing life, called traveling partners, such as batches of goods movement and trend in the modern logistics system, the species migration paths in scientist's study, and so on. These existence methods such as flock method, but it limits the shape of moving object group is circular, and it is oversensitive for its radius parameter [5]; the model of Swarm is

ISSN: 2005-4270 IJDTA

based on the concept of frequent itemsets, it is difficult to detect the large size of frequent itemsets and is difficult for large data sets [6] in space and time are too large, so this method cannot be directly used in RFID trajectory data that are such incremental data. This paper studied the moving object group with moving together in the RFID database, and proposed a discovery problem of the traveling partners. We can cluster the RFID trajectory data to produce sub-trajectory clusters by using a sub-trajectory distance measure method, and discovery the traveling partners by intersecting method. The discovery of traveling partners has played a catalytic role in studying frequent path and its future movement trend for further research.

2. Related Concepts

Raw RFID records is made up of triples in the form of <EPC, location, time>, where EPC is the global unique electronic product code, location is the position of the reader that is reading the tag, time is time point of reading the tag by the reader. RFID trajectory data stream is generated by m moving objects, it is defined as follows:

$$S = (L_1^1 L_2^1 \dots L_i^1 \dots L_m^1 L_1^2 L_2^2 \dots L_i^2 \dots L_m^2 \dots L_1^j L_2^j \dots L_i^j \dots L_m^j \dots)$$

Where L_i^j is the position of reader that is reading the i-th moving object at the j-th time point.

A sample RFID trajectory data in RFID database is shown in Table 1.

Table 1. RFID Trajectory Data in RFID Database

EPC	location	time
epc1	loc1	9:00
epc2	loc2	9:00
epc3	loc1	9:00
epc4	loc5	9:00
epc5	loc3	9:00
	⋮	
epc1	loc7	9:09
epc2	loc8	9:09
epc3	loc4	9:09
epc4	loc2	9:09
epc5	loc6	9:09

Definition 1 (sub-trajectory). For EPC of moving object is n, setting L_n^i is the position at time point of t_i and L_n^{i+1} is the position at time point of t_{i+1}, so $\overrightarrow{L_n^i L_n^{i+1}}$ is called sub-trajectory. Shown in Figure 1, ls_i and ls_j are both sub-trajectories.

Definition 2 (directly density-reachable). Let D be a sub-trajectories set at time interval, ε is the distance threshold of two sub-trajectories, μ is the density threshold,

$N_{\varepsilon}(ls_i) = \{ls \mid ls \in D$ and distance $(ls, ls_i) \le \varepsilon\}$. If $ls_j \in N_{\varepsilon}(ls_i)$ and $|N_{\varepsilon}(ls_i)| \ge \mu$, then ls_j is directly density-reachable from ls_i.

Definition 3 (density-reachable). Let *D* be a sub-trajectories set at time interval, ls_j is density-reachable from ls_i, if there is a chain of sub-trajectories $ls_1, ls_2, ls, \ldots, ls_n$, such that $ls_1 = ls_i, ls_n = ls_j$, while $ls_k \in D(1 \le k \le n)$, ls_{k+1} is directly density-reachable from ls_k in relation to ε and μ.

Definition 4 (density-connected). Let *D* be a sub-trajectories set at time interval, ls_j and $ls_i \in D$ are density-connected in relation to ε and μ, if there is a sub-trajectory $ls \in D$, ls_j and ls_i are both density-reachable from *ls* in relation to ε and μ.

3. The Description of Method

3.1. The Process of Creating Sub-Trajectory Clusters

With these following proposed concepts, we formally define the traveling partners and the candidate set of traveling partners based on RFID applications as follows:

Definition 5 (traveling partners). For RFID trajectory data, let δ_s be the scale threshold, δ_t be the continuous threshold, the moving object set *O* is called traveling partners, it must be satisfied as follows:

(1) when $t \ge \delta_t$, sub-trajectories that are generated by members of the moving object set *O* are connected to each other at time period *t*;

(2) The scale of the moving object set *O* is more than δ_s.

Definition 6 (the candidate set of traveling partners). For RFID trajectory data, let δ_s be the scale threshold, δ_t be the continuous threshold, the candidate set of traveling partners *C* must be satisfied as follows:

(1) when $t < \delta_t$, sub-trajectories that are generated by members of the candidate set of traveling partners *C* are connected to each other at time period *t*;

(2) The scale of the candidate set of traveling partners *C* is more than δ_s.

Because our algorithm is based on a density clustering method, we should introduce the distance measure of sub-trajectory.

As shown in Figure 1, ls_i and ls_j of sub-trajectories are generated by two moving objects at same time interval, the distance of ls_i and ls_j is composed of the horizontal distance, the vertical distance and the angular distance, as shown in Formula 1.

$$distance(ls_i, ls_j) = d_{\perp}(ls_i, ls_j) + d_{\parallel}(ls_i, ls_j) + d_{\theta}(ls_i, ls_j) \qquad (1)$$

Where $d_{\perp}(ls_i, ls_j)$ is the vertical distance between two sub-trajectories; $d_{\parallel}(ls_i, ls_j)$ is the horizontal distance between two sub-trajectories; $d_{\theta}(ls_i, ls_j)$ is the angular distance between two sub-trajectories.

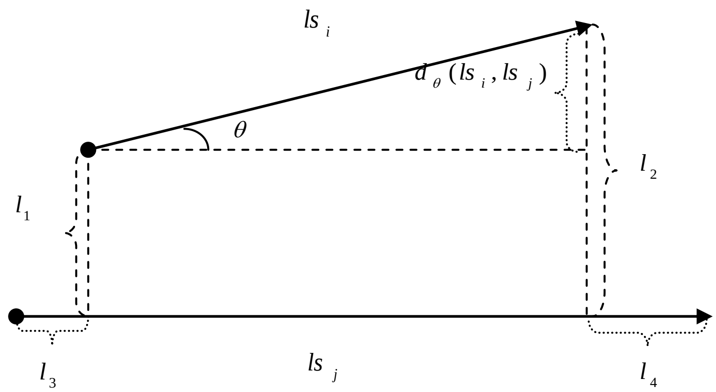

Figure 1. The Distance of Sub-trajectories

Their definitions are shown in formula 2-4, respectively.

$$d_{\perp}(ls_i, ls_j) = \min(l_1, l_2) \tag{2}$$

$$d_{\parallel}(ls_i, ls_j) = \min(l_3, l_4) \tag{3}$$

$$d_{\theta}(ls_i, ls_j) = \begin{cases} |ls_i| \times \sin\theta, 0 \le \theta < \dfrac{\pi}{2} \\ |ls_i|, \dfrac{\pi}{2} \le \theta \le \pi \end{cases} \tag{4}$$

Where θ is the inner angle between two sub-trajectories.

Algorithm 1: The creating algorithm of sub-trajectories cluster based RFID
Input: RFID trajectory data, distance threshold ε , density threshold μ .
Output: sub-trajectories cluster *sc*
Step:
(1) for all RFID trajectory data
(2) using RFID trajectory data to generate sub-trajectories set *D*
(3) all sub-trajectories in *D* are marked as “unvisited”
(4) do
(5) select randomly a sub-trajectory *ls* that is marked as “unvisited”
(6) Setting *ls* is marked as “visited”
(7) if there is at least μ sub-trajectories in the ε -neighborhood of *ls*


```
(8)     then create a new sub-trajectories cluster sc
(9)     add ls to sc
(10)      let N be sub-trajectories set in the ε-neighborhood of ls
(11)      select ls' ∈ N
(12)      if ls' is marked as "unvisited"
(13)         ls' is marked as "visited"
(14)         if there is at least μ sub-trajectories in the ε-neighborhood of ls'
(15)           add these sub-trajectories to N
(16)      if ls' is not a sub-trajectory member of any sub-trajectories cluster
(17)           add ls' to sc
(18)      if there is any "unvisited" sub-trajectory in N
(19)        goto (11)
(20)      output sc
(21)   until all sub-trajectories are marked as "visited"
```

3.2. The Algorithm of MTP

Because RFID trajectory data are incremental, in order to efficiently get moving objects based RFID technology with moving together at a certain period of time, we use a intersecting method to solve the problem of low efficiency in discovering the traveling partners. The system is updating when the new trajectory data is coming, and then intersect sub-trajectories corresponding moving objects with moving objects in the stored candidate set, so as to gradually get the result.

Closed property: In a candidate of traveling partners, it is not necessary to intersect the candidate with the rest of clusters, if there is more than $scale(r) - \delta_s$ moving objects existing in the intersected cluster.

Proof: Since in every unit time, every moving object only generates a trajectory, it also belongs to a cluster. If more than $scale(r) - \delta_s$ moving objects already exist in the intersected cluster, in best case, intersected result will be smaller than the scale threshold.

Definition 7 (the closed candidate of traveling partners). A traveling partner candidate is a closed candidate of traveling partners, if no candidate is existing that it contains the traveling partner candidate and its continuous time is less than one of the traveling partner candidate.

MTP algorithm first processes the new incoming trajectory data to generate sub-trajectories, and initializes the temporary candidate set, using algorithm 1 to get sub-trajectories clusters (line 1-3). MTP algorithm checks the remaining scale of the candidate set, if it is satisfied conditions, the algorithm will intersect candidate set with the set of moving objects that are corresponding new generated sub-trajectories clusters to improve candidate set (line 4-10), and then delete these intersected moving objects from candidate set (line 11). The result of enough scale is stored into new candidate set (line 12). Those of enough continuous time are reported as traveling partners (line 13). MTP algorithm always checks whether it is exist a candidate set that contains same moving objects and lasts for longer time, only those that pass the checking of closed property will be added new candidate (line 15-16). Finally, the candidate set C is updated, and algorithm waits for processing the new incoming trajectory data (line 17).

Algorithm 2: the mining traveling partners algorithm (MTP)

Input: new RFID trajectory data, distance threshold ε , density threshold μ , scale threshold δ_s , continuous threshold δ_t , the candidate set of traveling partners C

Output: traveling partners

Step:

(1) for each new incoming RFID trajectory data
(2) initialize the temporary candidate set C'
(3) using algorithm 1 to get new sub-trajectories clusters
(4) select randomly a candidate r_i from C
(5) select a objects set from set of objects sets that generates sub-trajectories cluster
(6) if the scale of r_i is less than the scale threshold δ_s , then delete it from C
(7) if C is null , then break
(8) else goto (4)
(9) intersect r_i with the select object set to get the new candidate r_i'
(10) the continuous time of r_i plus this time is the continuous time of r_i'
(11) delete objects with the same one of r_i' from r_i
(12) if the scale of r_i' is more than the scale threshold δ_s , then add r_i' to C'
(13) if the continuous time of r_i' is more than the continuous threshold δ_t , then output r_i' as traveling partners
(14) repeat (5)-(13) until all sub-trajectories clusters are selected
(15) checking the closed property of every sub-trajectory
(16) if it is closed , then add it to C'
(17) copy C' to C

4. Experiment and Results Analysis

As collecting trajectories in scenario of RFID real life is not easy, we process these related experiments to test the performance of our algorithm on synthetic dataset. By comparing it with TCMM algorithm, we can find the outstanding feature and advantages of our algorithm. TCMM [7] is an incremental clustering algorithm to cluster those trajectory data, it is fit for comparing with our algorithm to prove the better performance of our algorithm.

Dataset: The synthetic dataset contains 150 moving objects, they generate 30000 records.

Experiment Environments: The experiment was performed on a computer with Intel i3 Dual-Core CPU 1.8GHz, and 2GB RAM. The operating system is Windows 7, all algorithms are implemented in C# on VS2008.

We firstly test and evaluate the effectiveness of our algorithm, and we use the accuracy rate to measure it. The accuracy rate is the percentage of the mining result of MTP over true traveling partners.

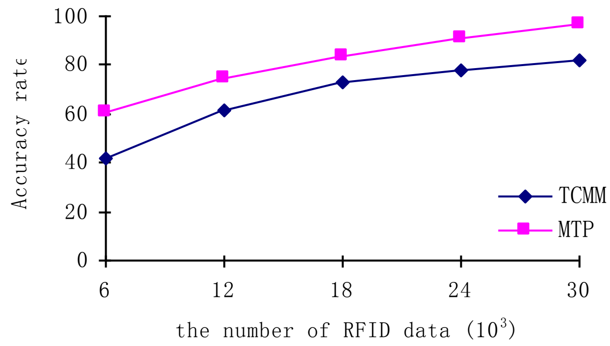

Figure 2. Effectiveness Comparison

In Figure 2, we can find that the performance of our algorithm is obviously better than TCMM, the accuracy rate of MTP is more nearly 15% than TCMM . Because TCMM uses a representative line segment to measure the distance between sub-trajectory and the cluster of sub-trajectories, it makes some sub-trajectories not add in clusters.

Secondly, we test the efficiency of our algorithm, and we use the parameter of execution time to evaluate it. The result of efficiency experiment is shown in Figure 3.

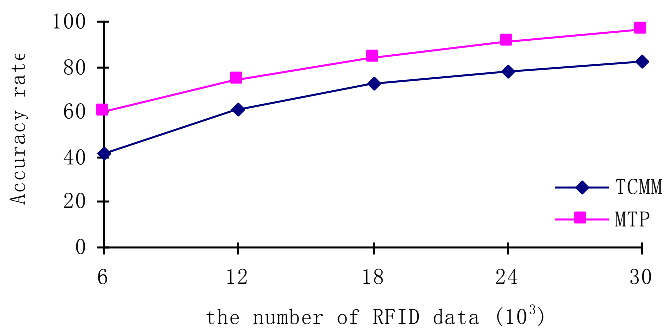

Figure 3. Efficiency Comparison

Figure 3 shows that MTP costs less time than TCMM, it illustrates MTP algorithm has a higher efficiency. This is due to TCMM has more complicated calculation when seeking the candidate of traveling partners, it goes against quickly finding the result of experiment. In addition, we use the closed property to fast deal with those RFID data.

5. Conclusion

Discovering the moving objects with moving together from RFID trajectory data in RFID database is much difficult, and many existing methods cannot efficiently deal with these RFID trajectory data. This paper proposed a suited RFID data method to find the traveling partners those groups with moving together for solving the problem of discovering the traveling partners. We proposed the mining the traveling partners algorithm (MTP) can effectively and efficiently mining these partners. First, it is creating sub-trajectories clusters, and then intersecting candidates in the candidate set with moving objects which are corresponded by sub-trajectories clusters, using closed property to reduce running time of our algorithm, the algorithm can achieve the function of quickly obtaining these corresponding results.

MTP algorithm can effectively and efficiently handle the problem of discovery the traveling partners in RFID trajectory data. By comparing TCMM algorithms, we test and verify the better performance of our algorithm. In the future work, we are ready to study anomaly detection about the traveling partners, and apply to intrusion detection, anomaly identification management and other practical scenario.

Acknowledgments

The research in the paper is supported by the National Natural Science Foundation of China under Grant No. 61003180; the Cooperative Innovation Fund of Jiangsu Province of China under Grant No. BY2013063-08; the “Six Talent Peaks Program” of Jiangsu Province of China; the “333 Project” of Jiangsu Province of China.

References

[1] K. Buchin, M. Buchin, M. van Kreveld, B. Speckmann and F. Staals, “Trajectory Grouping Structures”, Proceedings of the 13th International Symposium, WADS, London, ON, Canada, **(2013)** August 12-14.
[2] K. Hu, Z. Li, L. Chen and A. Song, “Mining Frequent Path of Moving Objects in Radio Frequency IDentification Database”, Advanced Science Letters, vol. 6, no. 1, **(2012)**, pp. 563-566.
[3] Z. Li, J. Han, B. Ding and R. Kays, “Mining Periodic Behaviors of Object Movements for Animal and Biological Sustainability Studies”, Data Mining and Knowledge Discovery, vol. 24, no. 2, **(2012)**, pp. 355-386.
[4] J. Shi, Y. Li, W. He and D. Sim, “SecTTS: A secure track & trace system for RFID-enabled supply chains”, Computers in Industry, vol. 63, no. 6, **(2012)**, pp. 574-585.
[5] M. Benkert, J. Gudmundsson, F. Hübner and T. Wolle, “Reporting flock patterns”, Journal Computational Geometry: Theory and Applications, vol. 41, no. 3, **(2008)**, pp. 111-125.
[6] Z. Li, B. Ding, J. Han and R. Kays, “Swarm: Mining relaxed temporal moving object clusters”, Proceedings of 2010 International Conference on Very Large Data Bases, Singapore, **(2010)** September 13-17.
[7] Z. Li, J. Lee, X. Li and J. Han, “Incremental clustering for trajectories”, Proceedings of the 15th International Conference on Database Systems for Advanced Applications, Tsukuba, Japan, **(2010)** April 1-4.

Authors

Kongfa Hu, he received the master degree in computer application from the Anhui University of Science and Technology in 1997. He received the Ph.D. degree in computer application technology from the Southeast University in 2004. Currently, he is a professor at Nanjing University of Chinese Medicine and the dean of the College of Information Technology of Nanjing University of Chinese Medicine. His research interests include RFID Data Management and Data Mining, Traditional Chinese Medicine Informatics and Internet of Things.

Long Li, he is postgraduate in Yangzhou University. His research interests include RFID Data Management and Data Mining.

