International Journal of Hybrid Information Technology
Vol. 9, No.6 (2016), pp. 317-326
http://dx.doi.org/10.14257/ijhit.2016.9.6.28

Study of Traveling Partners' Discovery Algorithm Based Closed Clustering and Intersecting

Kongfa Hu*, Jiadong Xie, Chengjun Hu, Tao Yang, Yuqing Mao, Yun Hu and Long Li

School of Information Technology of Nanjing University of Chinese Medicine, Nanjing 210023, China
kfhu@njutcm.edu.cn

Abstract

As the rapid development of IOT (the Internet of Things), RFID technology has been widely applied, and it generates a large of RFID trajectory stream data with the spatial-temporal characteristic. Because RFID has many characteristics, it leads to become very difficult that extracting moving objects groups that together moving (ie. traveling partners) in a period of time from RFID trajectory stream data. Existing methods are difficult to efficiently find this model. This paper presents a closed clustering and intersecting algorithm (CCI) for RFID data to detect movement along traveling partners, which is mainly constituted by two steps: first step is clustering sub-trajectory, it generates sub-trajectory clusters; second step is intersecting sub-trajectories with the traveling partners' candidate set to improve the candidate set, and find out traveling partners. In this process, we use the principle of Closure to accelerate our processing. Through experiments on the RFID synthetic dataset, we demonstrate the effectiveness and efficiency of our algorithm, thus show that our algorithm is suitable for discovering traveling partners in RFID applications.

***Keywords**: Internet of Things; RFID trajectory streams; Sub-trajectory clusters; Closed clustering and intersecting algorithm*

1. Introduction

In recent years, with the continuous development of sensor technology and wireless communication network technology , the state has focused on Internet of things included in the cultivation and development of emerging industry directory[1].The Internet of things is refers through the device of various information sensing device on the object, according to the communication protocol, and through the corresponding interface, the item is connected to the Internet, the exchange of information and communication, so as to realize intelligent identification, location, tracking and monitoring and management of a large network[2]. Due to the rapid development of Internet of things , all kinds of wireless sensor devices and positioning equipments have been widely used, such as scientific research, logistics management, traffic management, tracking management and security monitoring[3],*etc*., including the radio frequency identification (RFID) technology because of the low cost and simple deployment more favored by people[4].As RFID gradually mature and widely deployed RFID devices daily produce large mass RFID data, how to extract useful information from the RFID data becomes more important.

In this kind of information, the moving objects exist in a large amount of people living together in the movement, which is very concerned about the travel companion. For example, the movement of goods in bulk and its moving tendency in the modern logistics system, *etc*. The existing methods such as flock method, because of its shape

*Corresponding author: Kongfa HU, Email: kfhu@njutcm.edu.cn will be moving objects set limit to round, the radius parameter sensitivity ;The Swarm model is based on the concept of frequent item sets, detecting the frequent item sets of large size and applying to large data set is very difficult[6]; the general clustering intersection method [7-8] in terms of time and space overhead is too large and cannot be directly used for the RFID trajectory data stream such incremental data .This paper studies the group movement of the moving objects in the RFID application, puts forward the problem of discovery travel companion, the trajectory data clustering by using a sub trajectory distance measurement method to produce a sub trajectory cluster, and then quickly found travel companions through closed clustering intersection. The traveling companions can found frequent path for further research and play a significant role in future trend of the movement.

2. Sub Trajectory Clustering Algorithm

In different contexts, travel companions share some common principles based on RFID applications .

•**Cluster**：Travel companion is the movement of the mobile object, namely in the same cluster .As people, cars and animals often move and organize in any way, travel companion shape is not fixed, can be round, oval, square and so on.

•**Consistency**：Travel companion should be consistent, enough to last for a continuous period of time. This feature makes the different time periods on the cluster intersection to find travel companion.

•**Scale**：Most of the users are only interested in objects large enough group. They may be on the scale required travel companion. For example, the user set scale threshold to 4 and required travel companion to last for at least 4 consecutive times.

The original RFID record was composed by shape such as <EPC, location, time> set of triples, where EPC is the electronic product code that has the global uniqueness, location is the reader position that read label, and time is the time point of the label read by the reader. RFID trajectory data flow generated by the M mobile objects is defined as following:

$$S = (L_1^1 L_2^1 \cdots L_i^1 \cdots L_n^1 L_1^2 L_2^2 \cdots L_i^2 \cdots L_n^2 \cdots L_1^j L_2^j \cdots L_i^j \cdots L_n^j \cdots L_1^m L_2^m \cdots L_i^m \cdots L_n^m)$$

Where L_i^j represents the first *i* mobile object is located in the *j* point in time reader position.

2.1. Sub Trajectory Distance Metric

Definition 1(Sub Trajectory). To have the same EPC value of the mobile object, a point in time the location of t_i written L^i, point in time the location of the t_{i+1} written L^{i+1}, $\overrightarrow{L^i L^{i+1}}$ is called the sub trajectory. As shown in Figure 1, the ls_i and ls_i is two sub track.

Definition 2(Sub Trajectory Distance). As shown in Figure 1, for the two moving objects in the same time interval generated sub trajectory for ls_i and ls_i, the distance between them as shown in equation 1, is from the initial distance, center distance and the distance of the end points of the composition.

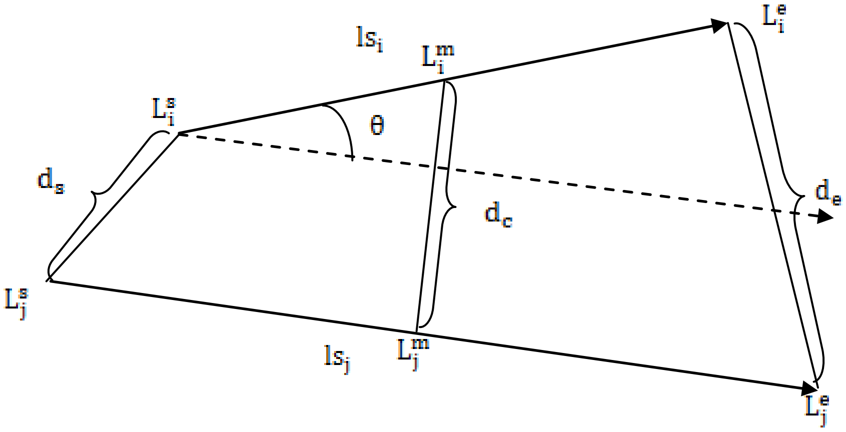


Figure 1. Sub Trajectory Distance

$$dis\tan ce(ls_i, ls_j) = \alpha * d_s + \beta * d_c + \gamma * d_e \qquad (1)$$

Where $d_s = \left|L_j^s - L_i^s\right|$, is the Euclidean distance between two sub trajectory starting point; $d_c = \left|L_j^m - L_i^m\right|$ is the Euclidean distance of two sub path center point; $d_e = \left|L_j^e - L_i^e\right|$ is the Euclidean distance between two sub trajectory ending point.

α、β、γ respectively represent three distance weights, its definition as shown in equation 2-4.

$$\alpha = \begin{cases} \frac{1}{3} - \frac{\sin\theta}{6}, 0 \le \theta \le \frac{\pi}{2} \\ \frac{\sin\theta}{6}, \frac{\pi}{2} \langle \theta \le \pi \end{cases} \qquad (2)$$

$$\beta = \frac{1}{3}, 0 \le \theta \le \pi \qquad (3)$$

$$\gamma = \begin{cases} \frac{1}{3} + \frac{\sin\theta}{6}, 0 \le \theta \le \frac{\pi}{2} \\ \frac{2}{3} - \frac{\sin\theta}{6}, \frac{\pi}{2} \langle \theta \le \pi \end{cases} \qquad (4)$$

Where θ is the internal angle between two sub trajectory.

2.2. Sub Trajectory Clustering Based on Density

Definition 3(The Neighbor Sub Trajectories). Let *D* be a sub track set in the time interval and ε be the distance threshold of two note track. For sub locus *ls_i*, any sub tracks $ls \in D$ meet $dis\tan ce(ls, ls_j) \le \varepsilon$, sub tracks *ls* is called the neighbor of sub trajectory ls_i.

Definition 4(The Core Sub Trajectories). Let *D* be a sub track set in the time interval, ε be the distance threshold of two note track and μ is the density threshold. For sub trajectory ls_i, the neighbor sub trajectory set is denoted as $N_\varepsilon(ls_i) = \{ls \mid ls \in D \, and \, dis\tan ce(ls, ls_j) \le \varepsilon\}$, $\mid N_\varepsilon(ls_i) \mid$ said the neighbor sub trajectory set scale. If $\mid N_\varepsilon(ls_i) \mid \ge \mu$, then we call ls_i as the core sub trajectories.

Definition 5(Direct Density Reachability). Let *D* be a sub track set in the time interval, ε be the distance threshold of two note track and μ is the density threshold, $N_\varepsilon(ls_i) = \{ls \mid ls \in D \, and \, dis\tan ce(ls, ls_j) \le \varepsilon\}$.If $ls_i \in N_\varepsilon(ls_i)$ and $\mid N_\varepsilon(ls_i) \mid \ge \mu$, then ls_j is directly density reachable from ls_i.

Definition 6(Density Reachable). Let *D* be a sub track set in the time interval, ls_j is density reachable from $ls_{i,}$ if there is a sub trajectory chain $ls_1, ls_2, \ldots ls_{n,}$ which makes $ls_1=ls_i$, $ls_n=ls_j$, at the same time for $ls_k \in D(1 \le k \le n)$, from ls_k to ls_{k+1} is directly density reachable on ε and μ .

Definition 7(Density Connected). Let *D* be a sub track set in the time interval, ls_j and $ls_i \in D$ are about and density of connections, if there exists a cop sub trajectory $ls \in D$, which makes the ls_j and ls_i be density reachable from *ls* on ε and μ .

Definition 8(Sub Trajectory Clusters).A sub trajectory cluster is a collection of sub trajectory of density connectivity. Let *C* be a sub trajectory cluster, $sc \subseteq D$, $\forall ls_i$、$ls_j \in c$, meet the requirements of ls_i and ls_j are density connectivity and $\forall ls_i$、$ls_j \in D$, if $ls_i \in c$, the ls_j is density reachable from ls_i, then $ls_j \in c$.

2.3 The Algorithm Description

Algorithm 1: Sub trajectory clustering algorithm based on RFID
Input: RFID trajectory flow *S, the* distance threshold ε and the density threshold μ
Output: Sub trajectory clusters *c*
Steps

```
（1） The new trajectory flow stream data with the last arrival trajectory flow data together produced sub trajectory set D;
（2） Make all sub trajectories in D marked "unvisited";
（3） Randomly select a sub track ls marked as "unvisited";
（4） Label ls as "visited;
（5） If the $\varepsilon$-neighborhood in ls has at least $\mu$ sub trajectories;
（6） Create a new cluster c, add ls to c;
（7） Let N is the sub trajectories collection of the $\varepsilon$-neighborhood in ls;
（8） For each sub trajectory $ls'$ in N;
（9） If $ls'$ is “unvisited”;
（10） Label $ls'$ visited;
（11） If the $\varepsilon$-neighborhood $ls'$ has at least $\mu$ sub trajectories, add these sub tracks to the N;
（12） If $ls'$ is not any cluster members, then add $ls'$ to the c;
（13） Repeat (9) - (12) until there is not sub trajectory marked as "unvisited" in N;
（14） Output c;
（15） Repeat (3) - (14) until all of the sub tracks are marked as "visited".
```

3. Travel Companion Generating Algorithm

3.1. The Related Definition and Concepts

With the proposal of the above-mentioned concept, we formally define travel companion and its candidate based on RFID application are as follows:

Definition 9 (Travel Companion). For the RFID trajectory flow *S*, let δ_s be a size threshold and δ_t be a persistent threshold, object set *O* called a travel companion must satisfy:

（1） When $t \ge \delta_t$, within the scope of the time *t*, the sub trajectories generating by member in object set *O* is density connected each other;

（2） The scale of object set *O* $size(O) \ge \delta_s$ 。

Figure 2 shows the RFID path flow, we set the δ_s=4, δ_t=4, then obviously we can find {*O1*, *O2*, *O3*, *O4*} be a travel companion that satisfies the conditions, other objects trajectories corresponding do not satisfy the conditions of travel companion.

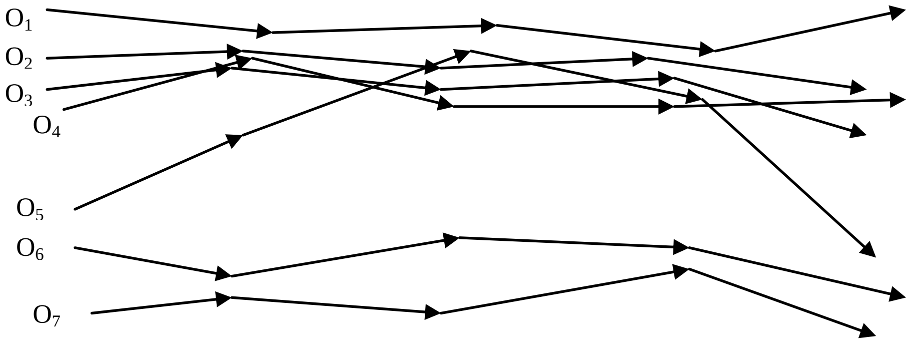


Figure 2. RFID Path Flow Definition 10 (Travel Companion Candidate)

For the RFID trajectory flow *S*, let δ_s be a size threshold and δ_t be a persistent threshold, travel companion candidate C must meet the following conditions:

（1） When $t\langle\delta_t$, the cycle of the t range, the sub trajectories generating by member in object set *C* is density connected each other;

（2） The scale of object set $C\,size(C) \geq \delta_s$ 。

Because the RFID trajectory flow is incremental, in order to efficiently obtain RFID moving objects over time in moving, we adopt a closed intersection method to improve the discovery efficiency of travel companion. The system always clusters the sub trajectories from the upcoming track flows, intersects the object and the stored corresponding candidate set objects, thus gradually get results.

Definition 11 (Closed Candidate Item).A travel companion candidate r_i, if there is not another candidate r_j that meet $r_i \subseteq r_j$ and to duration of r_i is less than duration of r_j, then r_i is a closed candidate Item.

3.2. Detailed Description of the Algorithm

Algorithm 2 closed clustering intersection algorithm（CCI）

Input: RFID trajectory data stream *S*, the distance threshold ε, the density threshold μ, the scale threshold δ_s, the time threshold δ_t, travel companion candidate item set *C*

Output: travel companion

Steps:

（1） For every new arrival trajectory data stream;

（2） Initialize the temporary candidate set $C^{'}$;

（3） Use the algorithm 1 to obtain the new sub trajectory cluster;

（4） Randomly select *C* from a candidate r_i;

（5） Select an object set from the collection object sub trajectories corresponding to the clusters produced in;

（6） If the r_i is smaller than the size threshold of δ_s, then remove it from C;

（7） If the C is empty, ends;

(8) Otherwise it returns (4);
(9) Let r_i and the object set intersect to obtain the new $r_i^{'}$;
(10) The duration will be continuous time of r_i plus duration of this time, resulting in a $r_i^{'}$;
(11) Delete $r_i^{'}$ with the same object from r_i;
(12) If $r_i^{'}$ is greater than the scale threshold δ_s, then insert the $r_i^{'}$ into C';
(13) If the duration is greater than the sustainable threshold δ_t, then output $r_i^{'}$ as a travel companion;
(14) Repeat（5）-（13）;
(15) For each sub trajectory cluster to check its sealing ability;
(16) if it is closed, its corresponding object set to join C ';
(17) The $C^{'}$ data is copied to the C.

The algorithm 2 first processes the arrival of the trajectory data stream to generate sub trajectory, and initializes a temporary candidate item sets, using the algorithm 1 to obtain the sub trajectory clusters (1-3 line).Then the system checks the candidate set size of surplus, after meeting the conditions, through candidate set and the corresponding object set of the new generation of cluster sub trajectory intersect to improve the candidate set(line 4-10),then the object that has been intersected is removed from the candidate set intersection (line 11).With sufficient scale results are stored as the new candidate set (line 12).Those enough duration was reported as a travel companion (line 13).When adding a new clusters to the candidate set, the algorithm always check whether the longer duration of candidate objects are the same, only those candidates through the enclosed check will be added as a new candidate (15-16 line).Finally, a set of candidate *C* is updated to wait for the processing of subsequent trajectory flow (line 17).

4. Experiment and Result Analysis

We evaluate the performance of our algorithm through the experiments in the synthetic data set . We protrude the characteristics and advantages of this algorithm by comparing with the Swarm mode (SW) and CMC algorithm,. SW is used to capture the moving objects of arbitrary shape of clusters in non continuous time, CMC is an incremental method for similar trajectories from a trajectory database cluster discovery.

4.1. Experimental Equipment

Data set : We evaluate the related algorithms through the RFID synthetic trajectory data sets of experiments .The synthetic data set contains 20,000 records 200 moving objects produced .

Experimental Environment : The experiment is run on a 2.53GHz frequency for the Intel I5 dual core processor,4G of memory on the computer. Operating system is the Windows7 operating system and all the algorithms use Visual C++ language.

4.2. Analysis of Effectiveness

We first evaluate the quality detection based on the results of the RFID travel companion. In the experimental data set, we test the validity of the results by means of accuracy and regression.

Accuracy: True companion accounts for the proportion by the searching algorithm. It represents that the algorithm finds meaningful peer selection rate.

Regression: The true peer discovery accounts for the proportion of the actual situation. This standard shows sensitivity for detecting the companion algorithm.

We use different δ_s to experiment. Figure 3 and Figure 4 show the results of effectiveness evaluation. We can obviously see that the CCI in the accuracy and the regression of both is better than CMC and SW.CCI increased nearly 20% than SW in accuracy, about 30% higher than that of CMC or so. SW generates frequent encounter group model objects, which have high sensitivity to help find all companions, but also easy to produce more false misinformation, reduce the algorithm sensitivity. CMC has the same problem with lower accuracy and regression. We also found that, with the increase of δ_s, the accuracy of CCI, CMC and SW all increased .However, if δ_s is set too high (more than 20), regression of algorithm rapidly descends, this is because some real partner will be filtered out.

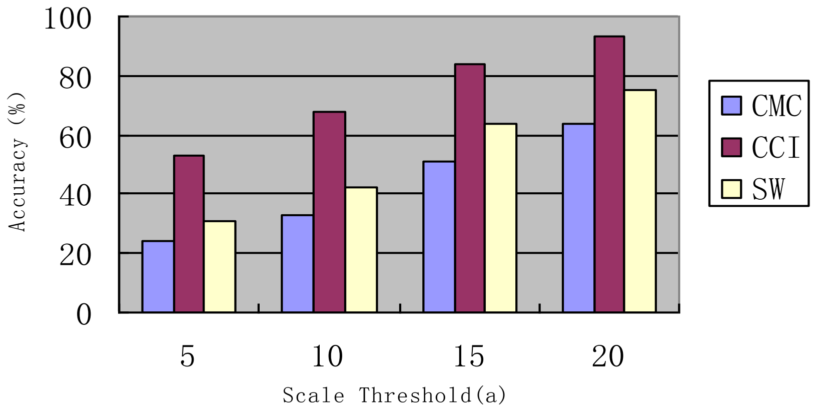


Figure 3. Comparison of Three Kinds the Algorithm Accuracy

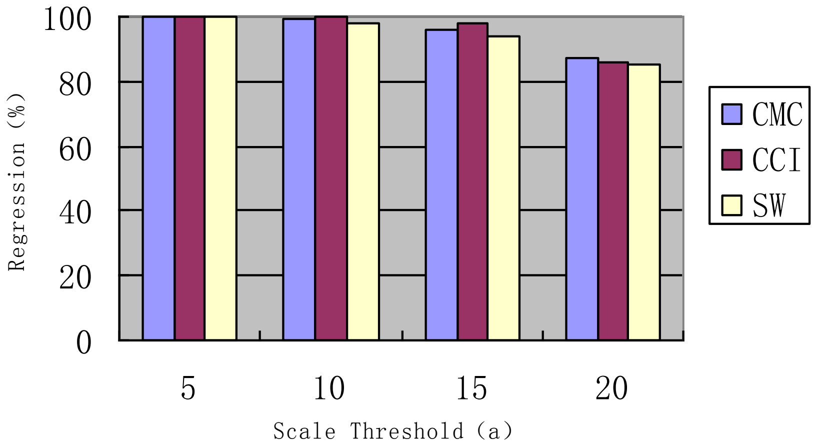


Figure 4. Comparison of Three Algorithms oof Regression

4.3. Efficiency Analysis

After the analysis of validity, we research the efficiency of algorithm by Experiment. Because SW can not output results incrementally, we will measure the time spent on the entire data set.

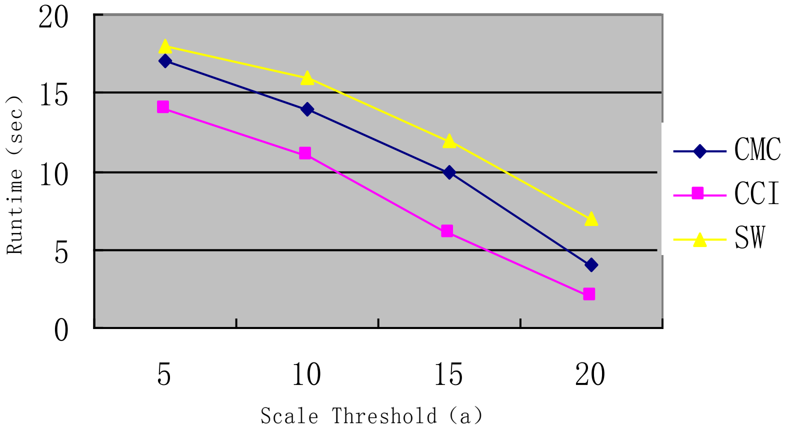


Figure 5. Analysis of Three Kinds of Algorithm Efficiency

Figure 5 illustrates the influence of size threshold of δ_s on experimental results. Generally, with the scale of the threshold becomes larger, travel companion candidate pruning efficiency is higher, the drop of time spent is more obvious.

5. Conclusion

Group objects motion together from the discovery of RFID locus on data streams is more difficult, many existing methods can not efficiently process RFID data of trajectory data stream. Aiming at the problems of the travel companion discovery based on the RIFD trajectory data stream applications, this paper presents a method for RFID data to find travel companion object group movement. We propose the closed clustering intersection (CCI) algorithm in order to mining this kind of peer quickly and effectively. First of all, by clustering the sub trajectory based on the method of density, get the sub trajectory clusters, then the candidate with the corresponding object candidate in the intersection set in turn, reduce the running time of the algorithm through the closed principle, achieve fast get the corresponding results of the function.

The CCI algorithm can effectively and efficiently handle travel companion discovery problems of RFID trajectory flow data. By comparing with SW and CMC algorithm, we verify the efficiency of the algorithm. In the next step, we are ready to aim at the discoverable travel companion, through its representative frequent path to study its mobile mode and apply to the actual situations of logistics management and traffic management.

Acknowledgements

The research in the paper is supported by the National Natural Science Foundation of China under Grant No. 81503499 and 61403156; Qing Lan Project of Jiangsu Province of China; the Natural Science Foundation of Jiangsu Province, China (BK20140958).

References

[1] J. Chen, J.Q. Ji and B.G. Chen, “Radio frequency identification (RFID) technology development strategy research in China. Scientific decision-making”, vol. 1, **(2010)**, pp. 8-20

[2] Y.F. Zhang, X.B. Zhao and S.D. Sun, “Implementing Method and Key Technologies for IoT-based Manufacturing Execution System”, Computer Integrated Manufacturing Systems, vol. 18, no. 12, **(2012)** , pp. 2634-2642.(in Chinese)

[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 and W. He, “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”, In: Proceedings of 36th International Conference on Very Large Data Bases, Singapore, **(2010)**, pp. 723-734.

[7] J. Gudmundsson and M. Kreveld, “Computing longest duration flocks in trajectory data”, In: Proceedings of ACM-GIS 2006, Virginia, USA, **(2006)**, pp. 35-42

[8] H. Jeung, M. Yiu, X. Zhou, C. Jensen and H. Shen, “Discovery of convoys in trajectory databases”, In: Proceedings of the 34th International Conference on Very Large Data Bases, Auckland, New Zealand, **(2008)**, p. 1068-1080.

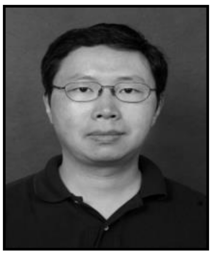
Authors

Kongfa Hu received the M. E. degree in computer application from Anhui University of Science & Technology, Huainan, China, in 1997, and the Ph.D. degree computer application technology from Southeast University, Nanjing, China, 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 Internet of Things and cloud computing , Traditional Chinese Medicine Informatics and big data analysis.

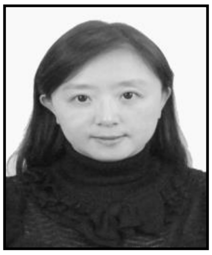
Jiadong Xie received the B. E. degree in computer science and technology from Nanjing University of Chinese Medicine, Jiangsu, China, in 2014. Currently, he is a graduate student at Nanjing University of Chinese Medicine. His research interests include cloud computing and machine learning.

Chenjun Hu received the B. E. degree in Educational Technology from Nanjing Normal University and the M. E. degree in Computer Science from Southeast University's School of Computer Science and Engineering, China, in 2000 and 2008 respectively. He is working toward the PH.D. Degree. He is a teacher at Nanjing University of Chinese Medicine. He is currently researching on Internet of Things and Cloud Computing, Bioinformatics and Machine Learning.

Tao Yang received the bachelor’s degree in computer science and technology from Nanjing University of Chinese Medicine, Nanjing, China, in 2009, and the Ph.D. degree in Diagnosis of Chinese Medicine from Nanjing University of Chinese Medicine, Nanjing, China, in 2014. Currently, he is a lecturer at Nanjing University of Chinese Medicine. His research interests are in TCM informatics and big data analysis.



Yuqing Mao received the B. S. and M. S. degrees in computer science from Nanjing University, Nanjing, China, in 1997 and 2000 respectively, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2012. Currently, he is a professor at Nanjing University of Chinese Medicine. His research interests include biomedical text mining, health informatics, machine learning and information retrieval.



Yun Hu received the M. E. degree in computer application technology from Southeast University, Nanjing, China, in 2007, and the Ph.D. degree in computer application technology from Nanjing University, Nanjing, China, in 2014. Currently, she is an associate professor of the College of Information Technology of Nanjing University of Chinese Medicine. Her research interests include Complex Networks, and data mining.