ISSN 0866 708X

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM
VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

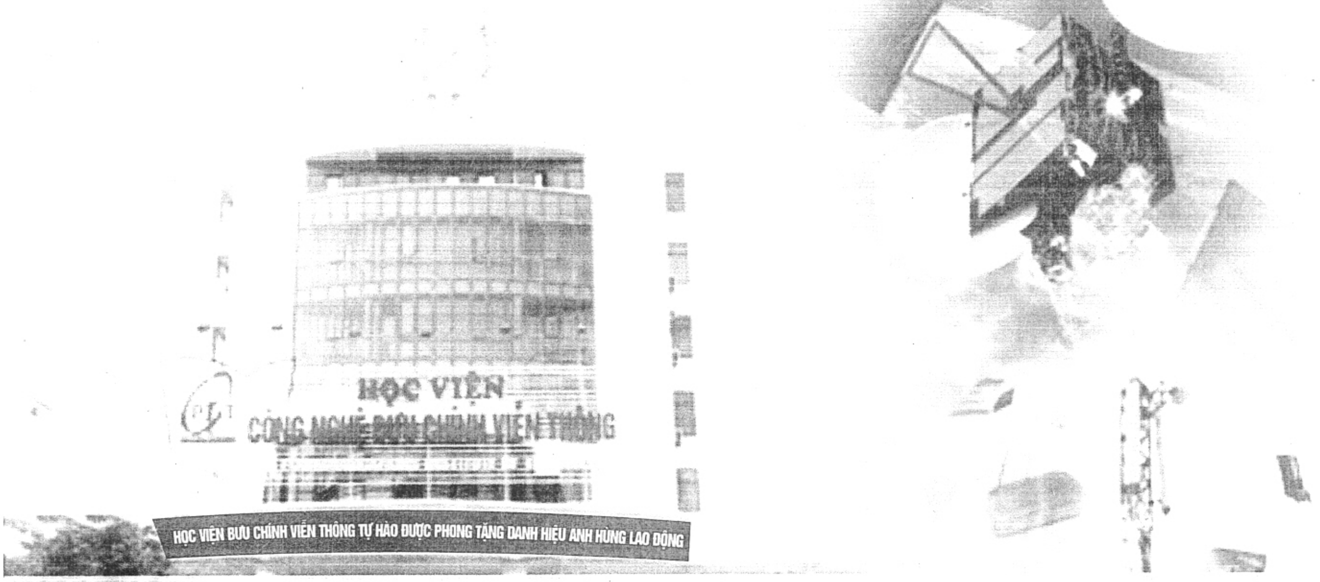
TẠP CHÍ
KHOA HỌC VÀ CÔNG NGHỆ

JOURNAL OF SCIENCE AND TECHNOLOGY

TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Tập 51 - Số 4A
2013

Chuyên san năm thứ 4
CÁC CÔNG TRÌNH NGHIÊN CỨU VỀ ĐIỆN TỬ, VIỄN THÔNG VÀ CÔNG NGHỆ THÔNG TIN



KỈ NIỆM 60 NĂM *(07/9/1953 - 07/9/2013)*
THÀNH LẬP TRƯỜNG BƯU ĐIỆN

TẠP CHÍ
KHOA HỌC VÀ CÔNG NGHỆ

Thư kí Toà soạn: **Lê Thị Thiên Hương:** tc_khcn@yahoo.com
Địa chỉ: Tòa soạn và Trị sự: Tầng 1, Nhà A16, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội;
Điện thoại/Fax: (84)-4-9422825; Email: **v-jst@vjs.ac.vn**, Website: **http://vjs.ac.vn/index.php/jst.**

Tạp chí Khoa học và Công nghệ **51** (4A) (2013) 30-43

DISCOVERING MOVEMENT SIMILARITY OF MOBILE USERS IN WIRELESS NETWORKS

Thuy Van T. Duong[1,*], Dinh Que Tran[2], Cong Hung Tran[3]

[1]*Department of Information Technology, University of Ton Duc Thang*

[2]*Faculty of Information Technology, Posts and Telecommunication Institute of Technology, km10, Nguyen Trai, Ha Dong, Hanoi, Vietnam*

[3]*Faculty of Information Technology, Posts and Telecommunication Institute of Technology, Ho Chi Minh City, Vietnam*

[*]Email: *vanduongthuy@yahoo.com*

Received: 25 June 2013; Accepted for publication: 23 August 2013

ABSTRACT

Discovering movement behaviours of groups of similar mobile objects has played a crucial role in allocating resources and guaranteeing Quality of Service (QoS) in wireless network. In our previous work, we have proposed a weighted combination similarity measure which integrates simultaneously the spatial and temporal characteristics of mobility patterns in wireless networks. In this paper, we demonstrate the efficiency of the proposed similarity model by means of compared evaluation with two representative clustering techniques and then draw two novel issues. Experimental results demonstrate that: (*i*) it is reasonable and necessary to combine both spatial similarity and temporal similarity; (*ii*) temporal property is as significant to clustering quality as spatial property; (*iii*) the quality of clusters is improved significantly when applying the proposed combined measures. In addition, the experiment results also suggest to: (*i*) construct an effective clustering algorithm to deal with drawbacks of the both representative clustering techniques; (*ii*) better develop a mobility prediction technique using mobility behaviours of the same users group rather than only investigating mobility behaviours of a single individual user.

Keywords: wireless network, clustering, mobility patterns, mobile object, similarity measure, trajectory.

1. INTRODUCTION

In the wireless network, whenever a mobile user moves from one cell to another during using services, the network resources must be reallocated at future destination. In order to still guaranteeing Quality of Service (QoS), sufficient network resources must be allocated before the mobile user move to. This requires mobility information of the mobile user for reserving resources in such cells where the user is likely to be in. Using the user's movement history to predict his next locations is one of the most effective approach for advance network resources reservation and has attracted several research interests ([2, 16, 17, 18]). It is easy to see that successful prediction for the user is possible if his own movement history is complete, otherwise resulting in prediction accuracy may be affected due to the lack of extracted rules. Hence, discovering knowledge about the movement of various groups of mobile users in wireless networks has become critical in mobility prediction in wireless network. Although different mobile users express differences in their movement behaviours and the nature of their movement, they typically share similarities [1]. Many researches indicate that mobility prediction schemes should be based on modelling behaviours in group rather than the individual mobility model ([1, 2, 3, 4, 5, 8, 9, 12]). And in turn, mobility behaviours in group are determined in term of the similarity between movement patterns. Thus, the problem of modelling mobility behaviours in group is reduced to determining the similarity of mobility patterns among mobile users.

In our previous work [12], we have proposed a weighted combination similarity measure which is concerned with characteristics on both time and space of mobility patterns in wireless networks. This paper is a continuation of our study to demonstrate the efficiency of the proposed similarity measure by means of experimental evaluation. The contribution of this work is three-fold:

- Reviewing the model of our similarity measure to show the reasonability of the combination of time and space into mobility patterns in wireless network.
- Conducting experiments to evaluate effectiveness of the proposed measure as well as usefulness of the measure in discovering group similar behaviors of mobile users. Our evaluation is based on implementation and comparison with the same measure of two well known clustering techniques on a synthetic dataset. The experimental results show that it is necessary to integrate simultaneously spatial and temporal properties of mobility patterns. In addition, the best clustering quality at $\alpha = 0.5$ implies that the temporal component, which has been not fully taken into account in most of previous studies, is as significant to clustering quality as the spatial component.
- Drawing two novel issues: First, developing a better clustering algorithm based on the proposed combination similarity measure to overcome the drawbacks of both clustering techniques in [14] and [15]. Second, discovering that mobile users in wireless networks may be classified into groups of similar mobility behaviors. Thus, it is possible for using mobility behaviors in the same user group to deal with the incompleteness of information in individual movement history. It would be better if constructing a mobility prediction technique is based on mobility behaviors in group instead of only based on individual mobility behaviors.

The remainder of this paper is constructed as follows. Section 2 describes the similarity model for mobility patterns in wireless networks. In section 3, some clustering techniques are implemented with our combination similarity measure to evaluate its effectiveness. Section 4 is a discussion and related works. Finally, Section 5 draws conclusion and further work.

2. SIMILARITY MODEL FOR MOBILITY PATTERNS

2.1. Mobility patterns in wireless networks

This section is devoted to a brief description of our model of mobility patterns [2, 13] that is the basis for our proposed similarity measures. In this paper, it is assumed that the radio coverage region is represented by a hexagonal shaped network (figure 1). Each hexagon is a cell which is served by a Base Station (BS) in the communication space. The mobile nodes can travel around the coverage region. The bi-directed graph is utilized to illustrate the mobility model of nodes in wireless network. Suppose that $G = (V, E)$ is an un-weighted directed graph, where V is the set of cells in the coverage region and the set E of edges represents the adjacency between pairs of cells. Each cell in V is determined with the ID number c and then a location of a mobile user at the timestamp t is defined as a couple (c, t). The bi-directed edges illustrate the fact that mobile users may move from one cell to another directly and vice versa. The movement trajectories of mobile users around the coverage region are collected. This concept has been introduced by Yavas et all [18] and in this paper it is formalize as follows:

Definition 1. *Let C and T be two sets of ID cells and timestamps, respectively. The ordered pairs* $p = (c, t)$, *in which* $c \in C$ *and* $t \in T$, *is called a point. Denote P to be the set of all points* $P = C \times T = \{(c, t) \mid c \in C \text{ and } t \in T\}$.

Two point $p_i = (c_i, t_i)$ and $p_j = (c_j, t_j)$ are said to be equivalent if and only if $c_i = c_j$ and $t_i = t_j$. Point $p_i = (c_i, t_i)$ is defined to be earlier than point $p_j = (c_j, t_j)$ if and only if $t_i < t_j$, and it is denoted as $(c_i, t_i) < (c_j, t_j)$ or $p_i < p_j$.

Definition 2. *The trajectory of the mobile node is defined as a finite sequence of points* $<p_1, p_2, \ldots, p_k>$ *in* $C \times T$ *space, where* $p_j = (c_j, t_j)$ *are points for* $1 \le j \le k$ *and ID cells of two consecutive points must be neighbors in the coverage region. A trajectory composed of k elements is denoted as a k-pattern.*

Note that the ascending order of points of trajectory is sorted by using t as the key. For example $<(c_1, t_1), (c_2, t_2), (c_3, t_3), (c_4, t_4)>$ is a trajectory, where $t_1 < t_2 < t_3 < t_4$.

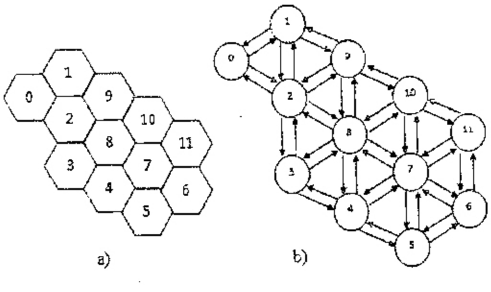


Figure 1. An example coverage region (a) and the corresponding bi-directed graph G (b).

2.2. Similarity measures for mobility patterns in wireless networks

This section presents briefly the computational model of similarity for mobility patterns (For more detail, see [12]) which is experimentally evaluated in the next section. Our similarity model is motivated by the fact that the location of mobility users in wireless networks should be characterized via both spatial and temporal similarities. And then,

- Two mobility patterns are considered to be more similar in space if they share more common cells;
- Two patterns passing through the same cells at the same times must be considered more similar in time than the case they stayed at the different times.

Definition 3. *Let S be a set of mobility patterns. A dissimilarity measure D: S × S → [0, 1] is a function from a pair of patterns to a real number between zero and one and satisfies the following conditions:*

(i) *Reflexivity: for all P ∈ S D(P, P) = 0;*

(ii) *Symmetry: for all P, T ∈ S D(P, T) = D(T, P).*

2.2.1. Spatial Dissimilarity Measure

For simplicity of presentation, this section makes use of the mobility pattern without the time factor. Given such two mobility patterns $P_a = <c_{a1}, c_{a2}, \ldots, c_{an}>$ and $P_b = <c_{b1}, c_{b2}, , \ldots, c_{bm}>$, where $c_{ai} \in V$ and $c_{bj} \in V$, for all i, j. Spatial dissimilarity measure can be defined in terms of spatial dissimilarity between two mobility patterns. The more uncommon cells there are in two patterns, the more spatially dissimilar they are.

Definition 4. *Let f: S × S → R be a function representing the number of uncommon cells in two patterns P_a and P_b. Then, f is determined by the formula:*

$$f(P_a, P_b) = card(\{c_{ai} \mid c_{ai} \notin P_b\}) + card(\{c_{bi} \mid c_{bi} \notin P_a\}) \tag{2.1}$$

A spatial dissimilarity measure between two patterns is then defined as follows:

Definition 5. *The spatial dissimilarity measure $D_{space}(P_a, P_b)$ between two patterns P_a and P_b with length n and m, respectively, is defined as follows:*

$$D_{space}(P_a, P_b) = \frac{f(P_a, P_b)}{n+m} \tag{2.2}$$

It is clear that if two patterns are equal, the result of $f(P_a, P_a)$ is zero. Thus $D_{space}(P_a, P_b) = 0$. Conversely, if the two patterns do not share any cells, the number of uncommon cells in these patterns is $f(P_a, P_b) = (n + m)$, where n and m are the length of patterns P_a and P_b, respectively. Thus $D_{space}(P_a, P_b) = 1$. We have the following proposition:

Proposition 1. *The function $D_{space}(P_a, P_b)$ is a dissimilarity measure.*

Proof. It is clear that

$$0 \le card(\{c_{ai} | c_{ai} \notin P_b\}) \le n \text{ and } 0 \le card(\{c_{bi} | c_{bi} \notin P_a\}) \le m$$

By summation:

$$0 \le f(P_a, P_b) = card(\{c_{ai} | c_{ai} \notin P_b\}) + card(\{c_{bi} | c_{bi} \notin P_a\}) \le (n + m).$$

Thus, $0 \le D_{space}(P_a, P_b) = \frac{f(P_a, P_b)}{n+m} \le 1$

It is easy to check the reflexivity. When $P_a = P_b$, the result of $D_{space}(P_a, P_b)$ is zero due to $f(P_a, P_b) = 0$. Additionally, if the two patterns do not share any cell, then $f(P_a, P_b) = (n + m)$ and $D_{space}(P_a, P_b)$ is equal to 1 when it is divided by $(n + m)$.

The symmetric property is proven as follows. We have:

$$f(P_a, P_b) = card(\{c_{ai} \mid c_{ai} \notin P_b\}) + card(\{c_{bi} \mid c_{bi} \notin P_a\})$$

and

$$f(P_b, P_a) = card(\{c_{bi} \mid c_{bi} \notin P_a\}) + card(\{c_{ai} \mid c_{ai} \notin P_b\})$$

then $f(P_a, P_b) = f(P_b, P_a)$, according to Equation 2.2, $D_{space}(P_a, P_b)$ is equal to $D_{space}(P_b, P_a)$. Thus, symmetry property has been proven. The proposition is proved.

2.2.2. Temporal Dissimilarity Measure

Our temporal dissimilarity measure is defined by means of the temporal dissimilarity between two mobility patterns which is calculated to be the sum of all temporal differences between the timestamps of the common cells in two patterns. The smaller the total temporal difference is, the less temporally dissimilar the two patterns are.

Suppose that T is the set of predefined timestamps. Given such two mobility patterns $P_a = <(c_{a1}, t_{a1}), (c_{a2}, t_{a2}), \ldots, (c_{an}, t_{an})>$ and $P_b = <(c_{b1}, t_{b1}), (c_{b2}, t_{b2}), \ldots, (c_{bm}, t_{bm})>$, where $c_{ai}, c_{bi} \in V$ and $t_{aj}, t_{bj} \in T$ for all i, j. The formalization of the temporal dissimilarity measure between patterns P_a and P_b is given in the following definition.

Definition 6. *The temporal dissimilarity measure $D_{time}(P_a, P_b)$ between two patterns P_a and P_b with length n and m, respectively, is given by*

$$D_{time}(P_a, P_b) = \frac{1}{k} \sum_{i=1, j=1}^{n,m} \frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} \text{ where } c_{ai} = c_{bj} \tag{2.3}$$

where k is the number of common cells of P_a and P_b.

We have the following proposition.

Proposition 2. *The function $D_{time}(P_a, P_b)$ is the dissimilarity measure.*

Proof. Since $0 \le |t_{ai} - t_{bj}| \le max(t_{ai}, t_{bj})$ for all $t_{ai}, t_{bj} \in T$, $0 \le \frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} \le 1 \; \forall t_{ai}, t_{bj} \in T$.

According to Equation 2.3, we just compute the time difference between two timestamps t_{ai} and t_{bj} of common cell ($v_{ai} = v_{bj}$), then $0 \le \sum_{i=1, j=1}^{n,m} \frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} \le k$. Therefore,

$$0 \le D_{time}(P_a, P_b) = \frac{1}{k} \sum_{i=1, j=1}^{n,m} \frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} \le 1.$$

Clearly, when two patterns P_a and P_b pass through the same cells at the same timestamps, $D_{time}(P_a, P_b) = 0$, due to $|t_{ai} - t_{bj}| = 0$. In the case that either P_a or P_b passes through all cells at timestamp t_0, then $\frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} = 1$ and consequently $D_{time}(P_a, P_b) = 1$.

The Symmetry is proven as follows. We have:

$$\frac{|t_{ai} - t_{bj}|}{\max(t_{ai}, t_{bj})} = \frac{|t_{bj} - t_{ai}|}{\max(t_{bj}, t_{ai})} \; \forall t_{ai}, t_{bj} \in T$$

Then according to Equation 2.3, $D_{time}(P_a, P_b)$ is equal to $D_{time}(P_b, P_a)$. Thus, symmetric property has been proven. The proposition is proved.

2.2.3. Combination Dissimilarity Measure

Resulting from the above partial dissimilarity measures, we may construct the weighted combination dissimilarity measure as follows:

Definition 7. *The composition dissimilarity measure between two patterns P_a and P_b is given by*

$$D(P_a, P_b) = W_{space} . D_{space}(P_a, P_b) + W_{time} . D_{time}(P_a, P_b) \quad (2.4)$$

in which W_{space} and W_{time} are respectively weights of spatial and temporal dissimilarity measures, such that $W_{space} + W_{time} = 1$.

It is easy to prove the following proposition.

Proposition 3. *The function $D(P_a, P_b)$ is the dissimilarity measure.*

3. A EXPERIMENTAL EVALUATION OF DISSIMILARITY MEASURE

This section is devoted to investigating the effectiveness of the proposed dissimilarity measure through clustering the mobility patterns of mobile users. Several clustering algorithms have been proposed to group categorical objects. In this study, we make use of two techniques with our dissimilarity measure to cluster mobile users.

3.1. Representative Clustering Techniques

San et al. [14] have proposed a new alternative algorithm that extended the k-means method in clustering categorical data. The problem of clustering is stated in terms of a constrained non-linear optimization problem as follows:

$$\text{Minimize } P(W, Q) = \sum_{l=1}^{k} \sum_{i=1}^{n} w_{i,l} d(X_i, Q_l) \quad (3.1)$$

subject to

$$\sum_{l=1}^{k} w_{i,l} = 1, \quad w_{i,l} \in \{0,1\}, \quad 1 \le i \le n, 1 \le l \le k$$

where $W = \lfloor w_{i,l} \rfloor_{n \times k}$ is a partition matrix, $Q = \{Q_1, Q_2, \ldots, Q_k\}$ is the set of representatives, and $d(X_i, Q_l)$ is the dissimilarity between object X_i and representative Q_l. Instead of $d(X_i, Q_l)$ in Equation 3.1, we use the proposed *composition dissimilarity measure* (Equation 2.4) in order to evaluate its effectiveness.

Seung-Joon Oh et al. [15] have proposed a hierarchical clustering algorithm for clustering sequences. The authors use the following criterion function to choose two most similar clusters and merge them.

$$\text{Maximize } Cf = \sum_{r=1}^{k} \frac{1}{n_r} \sum_{i,j \in C_r} sim(i, j) \quad (3.2)$$

where n_r is the number of sequences in C_r and k is the number of clusters.

In order to understand how this algorithm works, we consider the problem of clustering n sequences. First, each of the $n \times (n - 1) / 2$ pairs of possible merges is evaluated, and two clusters with the maximum value of the criterion function are merged. After performing l

merging steps, each of the $(n - l) \times (n - l - 1) / 2$ pairs of possible merges is evaluated. Repeating this procedure until there are only k clusters. In order to evaluate our measure, instead of $sim(i, j)$ in the criterion function (Equation 3.2), we make use of the proposed *composition dissimilarity measure* defined by Equation 2.4.

3.2. Synthetic Dataset and Basis for Experimental Evaluation

In the scope of this paper, for testing clustering quality, we build a dataset generator which demonstrates the movement behaviors of mobile users around the coverage region (see Figure 1). At first, we manually generate 5 movement behaviors presented by sequences. For each movement behavior, approximately 100 mobility patterns are automatically generated by following procedure. Each mobility pattern is a movement around the graph G in Figure 1. So, mobility patterns are generated by inserting some points $p_i = (c_i, t_i)$ into the movement behavior satisfying two constraints. The former is network constraints, i.e., consecutive cells in patterns are neighbor in graph G and the later is temporal constraints which satisfy ascending order of timestamps in patterns. All of 100 generated mobility patterns are assigned the same label. Repeating the procedure for each movement behavior, we obtain a testing dataset which consists of 500 mobility patterns with 5 clusters, named P500-K5.

In this paper, we also adopt the same criterion to analyze the clustering results [14]. The criterion was called *clustering accuracy* and defined as follows:

$$r = \frac{1}{n}\sum_{l=1}^{k} a_l \tag{4.3}$$

where a_l is the number of data objects that occur in both cluster C_l and its corresponding labeled class, and n is the number of objects in the dataset. The higher the clustering accuracy is, the better clustering quality is.

3.3. Experimental Evaluation

This section is devoted to present the following issues:

- Experimentally evaluating the proposed combination dissimilarity measure;
- Comparing with other similarity/dissimilarity measures

3.3.1. Evaluating the proposed measure

In our experimentation, the measure is of the form $\alpha.D_{space}(P_a, P_b) + (1-\alpha).D_{time}(P_a, P_b)$ as in Definition 7, where $W_{space} = \alpha$ and $W_{time} = (1 - \alpha)$. The value of α is varied in the experimentations to find how significant the spatial and temporal properties of mobility patterns are to clustering quality. The value $\alpha = 0$ means clustering based on purely temporal dissimilarity, while $\alpha = 1$ means clustering based on purely spatial dissimilarity.

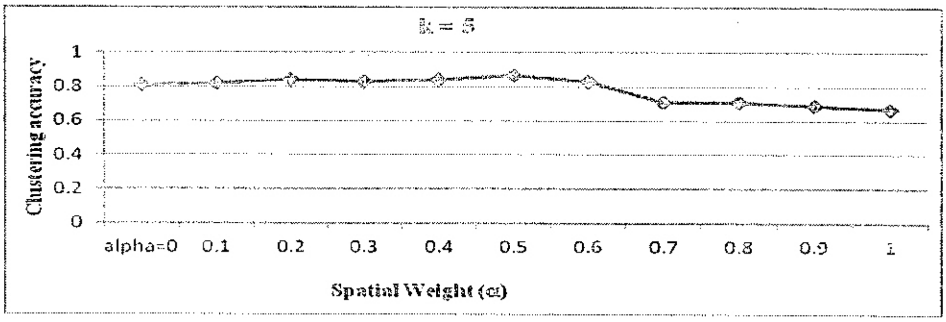


Figure 2. Clustering quality diagram for *k*-Representatives algorithm.

First, *k*-Representatives algorithm [14] is run on the generated P500-K5 dataset with $k = 5$ and α varying from 0 to 1 with 0.1 incremental steps. Figure 2 illustrates the clustering quality of *k*-Representatives algorithm using our measure with respect to the clustering accuracy r. The corresponding data are presented in Table 1.

Table 1. Clustering accuracies for *k*-Representatives algorithm.

Spatial Weight	$\alpha = 0$	0.1	0.2	0.3	0.4	**0.5**	0.6	0.7	0.8	0.9	1.0
Clustering accuracy	r = 0.81	0.82	0.84	0.83	0.84	**0.87**	0.83	0.71	0.71	0.69	0.67

Second, we run the clustering technique in [15] on the generated P500-K5 dataset with $k = 5$ and α varying from 0 to 1 with 0.1 incremental steps. We also obtain good clustering results as in Figure 3, from the obtained data in Table 2.

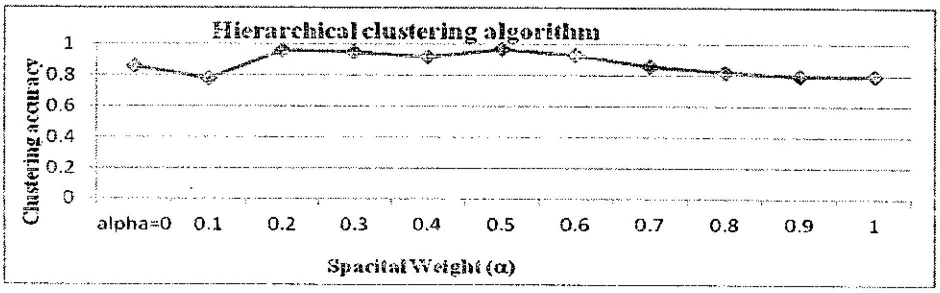


Figure 3. Clustering quality diagram for Hierarchical Clustering.

Table 2. Clustering accuracies for Hierarchical Clustering.

Spatial Weight	$\alpha = 0$	0.1	0.2	0.3	0.4	**0.5**	0.6	0.7	0.8	0.9	1.0
Clustering accuracy	r = 0.86	0.78	0.96	0.95	0.92	**0.97**	0.93	0.86	0.82	0.8	0.8

The experimental results show that the quality of clustering is well done. This improvement in the clustering quality may be contributed by means of our new dissimilarity measure. As expected, the clustering quality is improved with α varying from 0.2 to 0.6. This result indicates that both spatial and temporal properties are important to the clustering quality and therefore our proposed dissimilarity measures were well defined. In addition, another interesting experimental result being needed to mention here is that the best clustering quality is obtained when $\alpha = 0.5$. It implies that the temporal component is as significant to clustering quality as the spatial component.

3.3.2. Comparing with other measures

Intuitively, the measure used to compute the similarity between data objects is one of the key step which affects clustering quality. Several approaches have been proposed to measure the similarity between two sequences of categorical data with simultaneously spatio-temporal consideration. For instance, the work given by Tiakas et. al. [20] determines the similarity between two trajectories $T_a = (v_{a1}, v_{a2}, \ldots, v_{am})$ and $T_b = (v_{b1}, v_{b2}, \ldots, v_{bm})$ as follows:

$$D_{total}(T_a, T_b) = W_{net}.D_{net}(T_a, T_b) + W_{time}.D_{time}(T_a, T_b)$$

where $W_{net} + W_{time} = 1$ and:

$$D_{net}(T_a, T_b) = \frac{1}{m}\sum_{i=1}^{m} d(v_{ai}, v_{bi})$$

$$d(v_{ai}, v_{bi}) = \begin{cases} 0, & \text{if } v_{ai} = v_{bi} \\ \dfrac{c(v_{ai}, v_{bi}) + c(v_{bi}, v_{ai})}{2D_G}, & \text{otherwise} \end{cases}$$

where $D_G = \max\{c(v_i, v_j), \forall v_i, v_j \in V(G)\}$ is the diameter of the graph G of the spatial network.

$$D_{time}(T_a, T_b) = \frac{1}{m-1}\sum_{i=1}^{m} \frac{|(T_a[i+1].t - T_a[i].t) - (T_b[i+1].t - T_b[i].t)|}{\max\{(T_a[i+1].t - T_a[i].t), (T_b[i+1].t - T_b[i].t)\}}$$

The advantage of this approach is that it utilizes network distance and takes into spatiotemporal aspects account simultaneously. However, geographically close trajectories may not necessarily be similar in some real applications.

Another approach was proposed by Kang et. al. [21] which defined spatial similarity $LCSS_{a,b}(n, m)$ and temporal similarity $CVTI(a, b)$ where a and b are two trajectory patterns and n and m are the numbers of cells visited by a and b respectively.

$$LCSS_{a,b}(i,j) = \begin{cases} 0, \text{if } i = 0 \text{ or } j = 0 \\ LCSS_{a,b}(i-1, j-1) + 1, \text{if } i > 0, j > 0, a_i.c = b_j.c \\ max(LCSS_{a,b}(i, j-1), LCSS_{a,b}(i-1, j)), \text{otherwise} \end{cases}$$

$$CVTI(a,b) = \sum_{i=1,j=1}^{n,m} |a_i.I \cap b_j.I|, \text{where } a_i.c = b_j.c$$

The approach focuses on temporal similarity measure with the time interval which is different from the timestamp as in our work.

One of the most suitable methods for our purpose is defined by G´omez-Alonson et al. [22]. Their main purpose is to measure a similarity of sequences of events representing the behaviour of the user in a particular context such as sequences of web pages visited by a certain user or the personal daily schedule. The measure d_{OSS} is based on the comparison of the common elements (i.e. events such as web pages, places, etc.) in two sequences and the positions where they appear. The former computes if the two individuals have done the same things, e.g. if they have visited the same web pages or have gone to the same places. The later takes into account the temporal sequence of the events, e.g. if two tourists have visited the place A after going to the place B or not.

$$d_{OSS}(i,j)=\frac{f(i,j)+g(i,j)}{card(i)+card(j)}$$

where $g(i,j)=card(\{x_{ik}\mid x_{ik}\notin j\})+card(\{x_{jk}\mid x_{jk}\notin i\})$.

and

$$f(i,j)=\frac{\sum_{k=1}^{n}\left(\sum_{p=1}^{\Delta}\left|i_{(l_k)}(p)-j_{(l_k)}(p)\right|\right)}{\max\{card(i),card(j)\}}$$

where $i_{(l_k)}=\{t\mid i(t)=l_k\}$ and $\Delta=\min\{card(i_{(l_k)}),card(j_{(l_k)})\}$.

However, the ordering between the common elements in the two patterns may not fully consider temporal characteristic of patterns and hence our proposed similarity measure is its extension.

In order to demonstrate the efficiency of the proposed dissimilarity measure, this experiment implements the same clustering technique on three similarity measures that are the $d_{OSS}(i,j)$ in [22], the $sim(i,j)$ in [15] and our proposed measure $D(i,j)$ and then comparing their clustering results:

Case 1: The similarity measure in [22]

Case 2: The similarity measure in [15]

Case 3: The proposed dissimilarity measure

Table 3. Comparison of different measures.

Similarity Measure	Clustering accuracy (r)
Other Measure [22]	0.89
Other Measure [15]	0.85
Proposed Measure	0.97

For each case, we run the clustering algorithm on the generated P500-K5 dataset 10 times and calculate the average value of these clustering results. In Case 3, we take the optimal value for combination weight, i.e., α = 0.5. The corresponding data as in Table 3 illustrates the comparison of clustering quality in three cases. The experimental results show that the clustering quality of Case 3 is better than that of Case 1 and Case 2. This considerable improvement in the clustering quality demonstrates the efficiency of the proposed combination dissimilarity measure and indicates that it is suitable for measure the dissimilarity between mobility patterns in wireless network.

4. DISCUSSION AND RELATED WORKS

In reality, the mobility behaviors of mobile objects typically change as a function of time. It means that the temporal property of mobility patterns identifies the importance of the time when the mobile object moves from one location to another [2]. Therefore, it is necessary to concern with temporal aspect in estimating the similarity of two mobility patterns. Some approaches [9, [10, 11] have considered temporal property of mobility patterns but with time interval aspect or ordering type. Our approach is based on intuition that two patterns must be considered temporally similar when they pass through the same cells at the same time. For example, let $S1 = \{(1, t_1), (0, t_3), (5, t_4), (6, t_6), (7, t_9)\}$ and $S2 = \{(0, t_3), (5, t_4), (7, t_9)\}$ be two mobility patterns. In this context, there are 3 common cells 0, 5 and 7, and they have passed through cell 0 at $t3$, cell 5 at $t4$ and cell 7 at $t9$. Therefore, two these patterns are considered temporally similar. The suitability of this approach has been demonstrated by a case study [12] which compared the proposed dissimilarity measure with other ones.

In this paper, we demonstrate the efficiency of the proposed dissimilarity measure by means of experimental evaluation. Experimental results indicate that it is necessary to utilize the characteristics of dependency in space and time of mobile objects in wireless networks to construct a similarity measure for mobility patterns. In general, our weighted combination dissimilarity measure has been proved rationally in mathematics and its efficiency is also demonstrated by experimental evaluation in this paper.

As another contribution, the best clustering quality at $\alpha = 0.5$ implies that both spatial and temporal properties are equally significant to clustering quality while most of previous similarity measures ([9, 10]) have not fully considered.

Moreover, the experimental results also draw the following novel issues:

- The experiments in section 3.3.1 show that the clustering quality of Hierarchical Clustering technique [15] is better than that of *k*-Representatives Clustering algorithm [14]. However, the average run time of *k*-Representatives Clustering algorithm is significantly smaller than the one of Hierarchical Clustering technique. Thus, it is necessary to better develop a Clustering algorithm to take advantage of both these clustering techniques.
- The similarity between mobility behaviors of mobile users in wireless networks may enable to better develop a mobility prediction technique using mobility behaviors of the same users group to deal with the incompleteness of information in individual movement history as in most previous studies ([2, 18, 19]).

5. CONCLUSIONS AND FUTURE WORK

This paper has demonstrated the efficiency of the proposed dissimilarity measure by means of experimental evaluation and drawn two novel issues. We first review briefly two temporal and spatial dissimilarity measures and then made use of a weighted combination to integrate these partial dissimilarity measures. Second, we have conducted experiments with two representative clustering techniques to demonstrate the suitability of the weighted combination dissimilarity measure. The experimental results show that both spatial and temporal properties are equally important for constructing dissimilarity measure of mobility patterns. In addition, the quality of clusters is improved significantly when applying the proposed combined measures. The results may enable to develop an effective mobility prediction technique to deal with the incompleteness of information in individual movement history. We are currently utilizing the proposed combination dissimilarity measure to develop a clustering scheme for classifying mobility patterns into different mobility groups and then utilize the proposed clustering approach to construct a prediction technique based on mobility behaviours in group.

REFERENCES

1. Somayeh Dodge - Exploring Movement Similarity Analysis of Moving Objects, Doctor thesis, University of Zurich, Switzerland, 2011.
2. Thuy Van T. Duong, and Dinh Que Tran - An Effective Approach for Mobility Prediction in Wireless Network based on Temporal Weighted Mobility Rule, International Journal of Computer Science and Telecommunications 3 (2) (2012).
3. Hwang J., Kang H., and Li K. - Spatio-Temporal Similarity Analysis between Trajectories on Road Networks, Second International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS), Springer-Verlag, Austria, pp. 280-289.
4. Lin B. and Su J. - Shapes Based Trajectory Queries for Moving Objects, Proc. of the 13th annual ACM international workshop on Geographic information systems (GIS'05), 2005, pp. 21–30.
5. Zeinalipour-Yazti D., Song Lin S., and Gunopulos D. - Distributed Spatio-Temporal Similarity Search, Proc. of the 15th ACM International Conference on Information and Knowledge Management (CIKM), USA, 2006, pp. 14-23.
6. Vlachos M., Kollios G., and Gunopulos D. - Discovering Similar Multidimensional Trajectories, Proc. of the 18th ICDE, IEEE Computer Society Press, Los Alamitos, 2002, pp. 673–684.
7. Sakurai Y., Yoshikawa M., and Faloutsos C. - FTW: Fast Similarity Search Under the Time Warping Distance, Proc. of the 24th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS), USA, 2005, pp. 326–337.
8. Chen L., Ozsu M. T., and Oria V. - Robust and Fast Similarity Search for Moving Object Trajectories, Proc. of the ACM SIGMOD International Conference on Management of Data (ACM SIGMOD), USA, 2005, pp. 491–502.
9. Ying J. Jia, Ching Lu E. Hsueh-Chan, and Lee Wang-Chien - Mining user similarity from semantic trajectories, Proc. of the ACM International Workshop on Location Based Social Networks, USA, 2010, pp. 19-26.
10. Mohammad H. Pandi, Omid Kashefi, and BehrouzMinaei - A Novel Similarity Measure for Sequence Data, Journal of Information Processing Systems 7 (3) (2011) 413-424.
11. Chang Jae-Woo, Bista Rabindra, Kim Young-Chang, and Kim Yong-Ki - Spatio-temporal Similarity Measure Algorithm for Moving Objects on Spatial Networks, Proc. of the International Conference on Computational Science and Its Applications (ICCSA), Springer-Verlag Berlin Heidelberg, Malaysia, 2007, pp. 1165-1178.

12. Thuy Van T. Duong, Dinh Que Tran, and Cong Hung Tran - A Weighted Combination Similarity Measure for Mobility Patterns in Wireless Networks, International Journal of Computer Networks and Communication (IJCNC) **4** (3) (2012) 21-35.
13. Thuy Van T. Duong and Dinh Que Tran - Modeling Mobility in Wireless Netwok with Spatiotemporal State, International Conference in Mathematics and Applications (ICMA), Thailand, 2011, pp. 147-155.
14. Ohn Mar San, Van Nam Huynh, and Yoshiteru Nakamori - An Alternative Extension of the k-Means Algorithm for Clustering Categorical Data, International Journal Applied Mathematics Computer Science **14** (2) (2004) 241-247.
15. Seung-Joon Oh and Jea-Yearn Kim - A Sequence-Element-based Hierarchical Clustering Algorithm for Categorical Sequence Data, International Journal of Information Technology & Decision Making **4** (1) (2005) 81-96.
16. Prasad P. S. and Agrawal P. - Mobility prediction for wireless network resource management, Proc. of the IEEE 41st Southeastern Symposium on System Theory, USA, 2009, pp. 98-102, USA.
17. Al-Hanabani M., Subramaniam S., Othman M., and Zukarnain Z. - Mobility Prediction Based Resource Reservation for Handoff in Multimedia Wireless Cellular Networks, International Arab Journal of Information Technology **5** (2) (2008).
18. Yavas G., Katsaros D., Ulusoy O., and Manolopoulos Y. - A data mining approach for location prediction in mobile environments, Journal of Data and Knowledge Engineering **54** (2005) 121-146.
19. Vu T. H. N., Lee J. W., and Ryu K. H. - Spatiotemporal pattern mining technique for location-based service system, ETRI Journal **30** (3) (2008) 421-431.
20. Tiakas E., Papadopoulos A. N., Nanopoulos A., Manolopoulos Y., Stojanovic D., and Djordjevic-Kajan A. - Searching for similar trajectories in spatial networks, the Journal of Systems and Software, 2009, pp. 772-788
21. Kang HyeYoung, Kim Joon Seok, and Li KiJoune - Similarity measures for trajectory of moving objects in cellular space, Proc. of the 2009 ACM symposium on Applied Computing, USA, 2009, pp. 1325-1330.
22. Gomez-Alonso C., and Valls A. - A similarity measure for sequences of categorical data based on the ordering of common elements, Proc. of the 5th International Conference on Modeling Decisions for Artificial Intelligence, 2008, pp.134-145.

TÓM TẮT

KHÁM PHÁ ĐỘ TƯƠNG TỰ DI CHUYỂN CỦA NGƯỜI DÙNG DI ĐỘNG TRONG MẠNG KHÔNG DÂY

Dương Thị Thùy Vân[1,*], Trần Đình Quế[2], Trần Công Hùng[3]

[1]*Khoa Công Nghệ Thông Tin, Trường Đại học Tôn Đức Thắng, Thành phố Hồ Chí Minh.*

[2]*Khoa Công Nghệ Thông Tin, Học Viện Công Nghệ Bưu Chính Viễn Thông, km10, Nguyễn Trãi, Hà Đông, Hà Nội*

[3]*Khoa Công Nghệ Thông Tin, Học Viện Công Nghệ Bưu Chính Viễn Thông, Thành phố Hồ Chí Minh*

[*]Email: *vanduongthuy@yahoo.com*

TÓM TẮT

Khám phá hành vi di chuyển của các nhóm đối tượng di chuyển tương tự đóng vai trò rất quan trọng trong việc cấp phát tài nguyên và đảm bảo chất lượng dịch vụ trong mạng không dây. Trong công trình nghiên cứu trước, chúng tôi đã đề xuất một mô hình đo độ tương tự kết hợp theo trọng số nhằm tích hợp đồng thời đặc trưng không gian và thời gian của mẫu di chuyển trong mạng không dây. Trong bài báo này, chúng tôi chứng minh tính hiệu quả của mô hình tương tự đã đề xuất bằng cách đánh giá thực nghiệm độ đo dựa trên hai thuật toán phân cụm điển hình và từ đó đề xuất hai ý tưởng mới. Kết quả thực nghiệm chứng minh rằng: (*i*) việc kết hợp cả hai độ đo tương tự theo không gian và thời gian như đã đề xuất thật hợp lí và cần thiết; (*ii*) thuộc tính thời gian của mẫu di chuyển có ý nghĩa tương tự thuộc tính không gian trong việc gom nhóm mẫu di chuyển; (*iii*) chất lượng gom nhóm được cải thiện đáng kể khi áp dụng độ đo tương tự kết hợp được đề xuất. Ngoài ra, kết quả thực nghiệm cũng kiến nghị rằng: (*i*) cần xây dựng một thuật toán gom nhóm mẫu di chuyển hiệu quả nhằm khắc phục những nhược điểm của hai thuật toán gom nhóm điển hình trên; (*ii*) phát triển một kỹ thuật dự đoán sự di chuyển tốt hơn bằng cách sử dụng hành vi di chuyển của những người sử dụng trong cùng nhóm để dự báo thay vì chỉ xem xét hành vi di chuyển của từng cá nhân riêng lẻ.

Từ khóa: mạng không dây, gom nhóm, mẫu di chuyển, đối tượng di chuyển, độ tương tự, đường đi.

TẠP CHÍ KHOA HỌC VÀ CÔNG NGHỆ
Journal of Science and Technology
Chuyên san Điện tử, Viễn thông và Công nghệ thông tin

Tập 51, số 3A, năm 2013

MỤC LỤC
contents

In tại Xưởng in II, Nhà in Khoa học và Công nghệ, Viện Khoa học và Công nghệ Việt Nam.
Chỉ số: 12874 In xong và nộp lưu chiểu tháng 9 năm 2013.

TẠP CHÍ

KHOA HỌC VÀ CÔNG NGHỆ

Tạp chí Khoa học và Công nghệ đăng các bài tổng quan, các kết quả công trình nghiên cứu và các thông báo ngắn về các lĩnh vực Khoa học và Công nghệ. Bài gửi đăng được viết bằng tiếng Việt hoặc tiếng Anh. Tác giả đảm bảo nội dung của bài gửi là kết quả riêng, chưa đăng và chưa gửi đăng ở bất kì tạp chí nào khác