Geographic Community-based Mobility Model

Wendong Zhang[1,2,3], Xiaolin Gui[1,2], Jian An[1,2], Jinghua Jiang[1,2]

[1] *Department of Computer Science and Technology, Xi'an Jiaotong University,*
Xi'an 710049, P.R. China
[2] *Shaanxi Key Laboratory of Computer Networks,*
Xi'an 710049, P.R. China
[3] *School of Information Science and Engineering, Xinjiang University,*
Urumqi 830046, P.R. China, phone: 86-29-82668653-575,
xlgui@mail.xjtu.edu.cn

Abstract—**Mobility model plays a crucial role in performance evaluation of mobile wireless networks. However, the majority of existing mobility models either does not exhibit realistic movement characteristics or modeling methods are too complex. In this paper, inspired from the trend that users usually move between several popular areas in daily movement, a new mobility model based on Geographic Community (GCMM) is proposed. We concern with the topology construct of geographic environment and the destination selection scheme for user moving in GCMM. Simulation result shown that GCMM better depict the mobility patterns of human.**

Index Terms—**geographic community, hot region, mobility model, mobile wireless networks**

I. INTRODUCTION

With the advances of computation and communications technology, various mobile wireless devices (e.g. mobile phones, PDAs, and laptops) are becoming more popularity in people's lives. A new networking paradigm, which composed with mobile devices carried by humans, has been attracting increasing interest [1], [2]. In such networking environments, data forwarding between two nodes depend on opportunistic encounters of mobile nodes, and it has been demonstrated that the movement pattern of nodes has a significant impact on network performance [3]. Consequently, modeling human mobility patterns in realistic way plays an essential role in accurately evaluating and analyzing the performance of protocols and applications.

Accurately modeling realistic mobility pattern of human is non-trivial. A large number of empirical studies show that human's movement has a strong regularities [4], [5] and several realistic mobility models have been proposed [6]–[8]. Kyunghan el al. [6] proposed SLAW (self-similar least action walk) mobility model, which capture the features of human movement, namely, truncated power-law distributions of flights, pause-times and inter-contact times. Zheng et al. [7] extracted mobile nodes' social activities characteristic, proposed an agenda driven mobility model (AMM). In this

model, a person's social activities are represented by agenda, which cause the motion of a person. Rhee el al. [4] constructed a simple Levy walk mobility model (LWMM) that mimics human walk patterns through statistic GPS traces of 44 volunteers. The model exploits the statistical features of human walk that observed in mobility studies, including flight length, inter-contact times, etc. Heterogeneous Human Walk (HHW) model, which, from the society networks theory point of view, analyze the real traces and construct human mobility patterns in literature [8].

Undoubtedly, these models mentioned above are more reality than Random Walk Mobility model (RWMM), however, by in-depth analyzing, it can be found that AMM are close to real life scenarios, however the modeling methods are too complex and there exist a large application limitation. SLAW and LWMM comprehensive exhibit human motion patterns, it only focuses on individual motion patterns, and neglect the group characteristics of human clustering. HHW, similar to AMM, construct people motion patterns using changing different society roles. It also cannot exhibit the clustering features of human movement.

In fact, human has a significant clustering trend in real life, that is, the more population stay in a place, the stronger desire that one person wants to go to there. Meanwhile, human always stayed in several frequently visited places. In this paper, we focus on humans clustering features by analyzing GPS trace data, study the clustering region and the node's density distribution characteristic in an observation area, propose Geographic Community-based Mobility Model (GCMM). According to non-uniform distribution of node's density, we imitate the nodes' mobility patterns, and realize node's trajectory automatic generation.

In next section, several definitions were depicted in detail. By analyzing GPS trace data and extracting human movement characteristic, modeling method were proposed in the third section. In the fourth section, a simulation analysis for GCMM is presented. Finally, we discuss and conclude this paper in the fifth section.

II. PRELIMINARY WORK

A. Definition 1 (visited position)

Let *T* be a human trace dataset, *p* is a visited position when

$p \in T$.

The human trace dataset usually are denoted by a series of concrete points, the basic format of the data is (t, x, y, z), where t represents the time, and (x, y, z) represents the coordination of a position.

B. Definition 2 (pause position)

Support $p_1, p_2, ..., p_n \in T$ be one person's visited position sequence, if $\exists\ i,j \in [1,n]$, for $\forall\ k,l \in [i,j]$, satisfy $|p_{i-1},\ p_i| \geqslant \tau$ and $|p_j,\ p_{j+1}| \geqslant \tau$ and $|p_k,\ p_l| \leqslant \tau$, then p_m is a pause position, where $m \in [i,j]$, and τ is a threshold of distance between two visited position. For example, human walk average velocity is 5 kilometers per hours (equivalent to $1.4 m/s$), it can be used to calculated the threshold.

C. Definition 3 (hot region)

Generally refer to a sub-area that be visited frequently by human in an observation area. In fact, a hot region is a collection that composed by a number of pause positions.

Geographic communities are ubiquitous scenes of society, such as restaurant, office, school, and so forth. In other words, geographic communities are a region where human will cluster together, namely, hot region. So, human movement terrain can be represented by several hot region and path (i.e. connect two hot region), that is, a person move among hot regions on predefined path.

III. DATA ANALYSIS AND MODELING METHOD DEPICTION

Generally, the movement process of human in real life can be depicted as fellows: as far as any given area is concerned, there exist several sub-areas (hot regions) that attract nodes move to it and stay a period of time in the region. When a node want to move to a place, it always been attracted by the sub-area and different sub-area has the different attractive degree. For instance, in an amusement park, a person always select the highest attractive degree of areas and move toward there, when the person arrived the place, he/she will stay a period of time, with the decrease of attractiveness of this place, the person will choose another highest attractiveness place, and move to there. The process is repeated in circle.

According to above mentioned idea, GCMM design process consists of three steps:

1) Data Analysis: Given the method for analyzing GPS trace data;
2) Extract Human Clustering Characteristics: Obtain the number and attractive degree distribution regularity of hot regions;
3) Destination selection scheme: Implement dynamic destination selection.

A. Data Analysis

We use GPS trace from the CRAWDAD [9], which were collected from the real world, it include five different scenarios which were chosen for collecting human mobility traces using GPS receivers. Each record represents a visited position of a person.

Due to the GPS traces are recorded at every 30 seconds, we haven't the consecutively GPS data of volunteer's motion. So, according to empirical analysis, we consider that mobility node has a pause state if the distance between two consecutive positions is less than 40 meters (the human walk average velocity is 5 kilometers per hours), and the pause time period is 30 seconds, otherwise, the node is keeping walk and the pause time period is 0 second. After the preprocessing to the original trace, we can get the discrete pause state sets of volunteers. Subsequently, calculate the total pause time period using plus the pause time period of consecutively pause states and obtain volunteer's pause position set, that can describe the volunteer's "move-stay" movement patterns.

In our study, we aim to obtain the hot regions set, so we ignore the shape of the clusters and assume that all hot regions are circular area. Actually, with the increase of observation radius, irregular shapes can be contained in a hot region, namely, the shape of hot region cannot influence selecting the core position. Thus, we can obtain the different hot region sets by means of different radius. Extract hot region is a iterative process. Algorithm 1 depicts our methodology as follow:

Algorithm 1. Hot Region Sets Extract Algorithm
Input: Pause position datasets P
Output: hot region sets H
(1) P : Pause position set.
(2) N_i^r : Neighborhood position set of position p_i within radius r, where $p_i \in P$.
(3) $|U|$: Represent the element number in set U
(4) H_k : An element of H
(5) **do**
(6) $k \leftarrow 1$
(7)   **for** each p_i in P
(8) $N_i^r = \{ p_j \mid dist\ (p_i, p_j) < r, p_i, p_j \in P \}$
(9)   **end for**
(10) $m \leftarrow \max(\mid N_m^r \mid)$
(11) $H_k \leftarrow P_m$
(12) $P \leftarrow P - N_m^r$
(13) $k \leftarrow k + 1$
(14) **until** $\mid N_i^r \mid < s$
(15) $H : \{ H_k \}$

B. Extract Human Clustering Characteristics

Based on the above algorithm, the number of hot regions and pause position within each hot region can be counted. The number of hot regions change with the increase of observation radius from 100m to 800m is given in Table I.

TABLE I. THE NUMBER OF HOT REGIONS WITHIN FIVE SCENARIOS.

Radius(m) \\ Site	100	300	500	700	800
KAIST	31	28	22	16	14
NCSU	24	19	17	15	16
NewYork	32	32	33	29	29
Orlando	53	21	13	13	13
Statefair	11	6	6	6	6

As shown in Table I, with the increase of observation radius, there is a stable trend of the number of hot regions. This means that, towards the confined area, there only exist scant regions, in which, human can present a clustering movement

patterns.

Subsequently, we analyze the density of point of regions. Actually, it is hard to appreciate describe the attraction degree distribution of a location. Empirical observations indicate that the more attraction of a location, the higher frequency of a person visits the location and the longer pause time periods in the location. Hence, we denoted the density of pause point in a region as the attractive degree. Formal describe is as follows

$$AD(i) = \frac{S_i}{M}, \qquad (1)$$

where S_i is the number of pause position in hot region H_i, M is the total pause position in observation area, $AD(i)$ represent the attractive degree of hot region H_i, which is a scalar and its value range is $(0, 1)$.

For simplicity, take Orlando scenarios for example, by analyzing the distribution of density of hot regions, we find that the density of hot regions satisfy closely an exponential decline by means of fitting methods, as shown in Fig. 2. The feature basically accordance with the real situation, that is, there are only few region that be visited frequently by people, and the rest region usually seldom be visited.

C. Destination selection scheme

This sub section discusses the destination selection scheme, and takes the attenuation mechanism of attraction into account. Let us recall the process of human movement in real life, as above mentioned, hot region is a landmark that attract the people move to there and stay. When a person has arrived a hot region, following the attraction of hot region reduce gradually, the person will choose a new hot region with higher attraction, and move toward it.

In order to mimic such dynamic situation, we consider the influence of distance between a hot region and the current position of a node, and use *decision factor* θ as the selection condition

$$\theta_j^i = \frac{AD(i)}{dist(p_c, p_i)}. \qquad (2)$$

(2) represent the decision factor of node j to hot region H_i, where $AD(i)$ is the attractive degree of hot region H_i, p_c is the current position of a person, p_i is a position within region H_i, the $dist(p_c, p_i)$ represent a sum of segments of a path from p_c to p_i. We define $\theta_j^k = 0$ when the node just stays in the hot region H_k. Then, we can obtain the maximum decision factor of node j

$$\max\{ \theta_j^i \mid 1 \leq i \leq |H| \}. \qquad (3)$$

In general, the destination selection process can be described as follow: When a mobility node select a destination, the attractor for all hot regions is calculated by formula (2) and (3), and then the node select the greatest one, and move toward the hot region. Similar to RWMM, the node will pause a random time period in the hos region, and then the node calculates the decision factor again. To avoid a loop between two hot regions, we consider the attraction

attenuation mechanism: the attractive degree of a hot region will reduce to half of previous value after the node has visited the hot region. The whole process construct a circle, and such circle process will be repeated periodically.

IV. MODEL ANALYSIS

In this section, we show the simulation of GCMM, and verify whether the GCMM is consistent with the statistical characteristic that we observed from the GPS trace of human.

To produce the synthetically trace, we initial the simulation area using $1000 \times 1000 m^2$, the number of nodes is 200, hot region's number is 13, $a = 0.089, b = -0.007$, $c = 0.65, r = 0.42$. Meanwhile, we set the pause time period are chosen randomly in range [30, 60] seconds and the speed of nodes are chosen randomly from [5, 20] m/s. The simulation time is 1000 seconds. In order to mimic the law distribution characteristic of walker flight length, we construct the topology of terrain using hierarchical approach in our model [5].
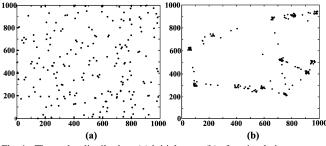

Fig. 1. The nodes distribution: (a) initial state; (b) after simulation.

Fig. 1 (a) and (b) are the nodes distribution of simulation start and end, respectively. Actually, the nodes mobility is a dynamic process, during the simulation, the nodes follow the location attraction change regular, exhibit a clustering characteristic by means of destination selection method. As Fig. 1 (b) shown, the nodes exhibit a better clustering characteristic.
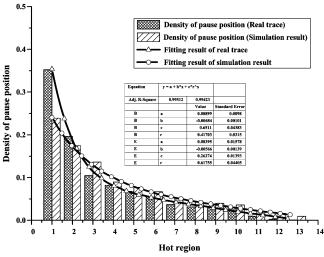

Fig. 2. Density distribution of hot regions.

Furthermore, in order to verify the mobility model whether abide by the statistical characteristic that we have observed from human movement patterns. We run the simulation 5 times, and then calculate the mean density of pause position

for each hot region. Finally, we analyze the statistical characteristic of synthetically trace.

As the Fig. 2 shown, simulation result better exhibit the statistical characteristics of real GPS trace data, and the pause position density of hot regions meets the exponential decline. Due to the nodes are random distribution when the simulation initialization, and move toward the nearest hot region at first. So, there exists some error about density of hot regions. In general, the simulation result is better consistent with the real scenario.

V. CONCLUSIONS

As an essential component of simulation for mobile wireless network, mobility model always received extensive attentions during the last decade. In particular, the research about human movement patterns causes a great attention in recent years. In this paper, motivated by modeling a realistic mobility model, we analyzed the GPS dataset captured from real scenarios and presented a hot region extraction algorithm. Subsequently, based on statistical characteristics of human movement, we proposed a geographic community mobility model, a.k.a. GCMM. By extracting the hot region set that the human frequently visited position in an observation area, we modeled human's movement pattern, meanwhile, in order to mimic human's dynamic decision-making process, we present a destination selection scheme and attraction attenuation mechanism. The simulation results shown that the GCMM better exhibit the human movement pattern in real life.

For the future work, we plan to study the packets forward protocols and performance evaluation for mobile wireless network under realistic mobility model scenarios. Due to influence of different mobility models during simulation, we will take differences between different mobility models into account.

REFERENCES

[1] P. Hui, A. Chaintreau, et al., "Pocket switched networks and human mobility in conference environments", in *Proc. of the ACM SIGCOMM workshop on Delay-tolerant networking*, 2005, pp. 244–251. [Online]. Available: http://dx.doi.org/10.1145/1080139.1080142

[2] L. Pelusi, A. Passarella, M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks", *Communications Magazine*, vol. 44, no. 11, pp. 134–141, 2006. [Online]. Available: http://dx.doi.org/10.1109/MCOM.2006.248176

[3] A. Chaintreau, P. Hui, et al., "Impact of Human Mobility on Opportunistic Forwarding Algorithms", *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007. [Online]. Available: http://dx.doi.org/10.1109/TMC.2007.1060

[4] I. Rhee, M. Shin, S. Hong, et al., "On the Levy-Walk Nature of Human Mobility", *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630–643, 2011. [Online]. Available: http://dx.doi.org/10.1109/TNET.2011.2120618

[5] X. P. Han, Q. A. Hao, B. H. Wang, et al., "Origin of the scaling law in human mobility: Hierarchy of traffic systems", *Physical Review E,* vol. 83, no. 3, pp. 2–6, 2011. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.83.036117

[6] L. Kyunghan, H. Seongik, K. Seong Joon, et al., "SLAW: A New Mobility Model for Human Walks", in *Proc. of the INFOCOM 2009*, 2009, pp. 855–863.

[7] Q. W. Zheng, X. Y. Hong, J. Liu, et al., "Agenda driven mobility modeling", *Journal of Ad Hoc and Ubiquitous Computing*, vol. 5, no. 1. pp. 22–36, 2010. [Online]. Available: http://dx.doi.org/10.1504/IJAHUC.2010.030000

[8] S. S. Yang, X. Y. Yang, C. Zhang, et al., "Using Social Network Theory for Modeling Human Mobility", *IEEE Network*, vol. 24, no. 5, pp. 6–13, 2010. [Online]. Available: http://dx.doi.org/10.1109/MNET.2010.5578912

[9] I. Rhee, M. Shin, S. Hong, et al., "CRAWDAD data set ncsu/mobility models", 2009. [Online]. Available: http://crawdad.cs.dartmouth.edu/ncsu/mobility models