Proc. the 10th International Conference on Intelligent Games and Simulation (GAME-ON 2009), Düsseldorf, Germany, pp. 56-60, November, 2009.

Analysis of User Trajectories Based on Data Distribution and State Transition: a Case Study with a Massively Multiplayer Online Game Angel Love Online

Ruck Thawonmas[1], Junichi Oda[1], and Kuan-Ta Chen[2]

[1]Intelligent Computer Entertainment Laboratory
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577, Japan
ruck@ci.ritsumei.ac.jp

[2]Multimedia Networking and Systems Lab
IIS, Academia Sinica
Taipei 115, Taiwan
swc@iis.sinica.edu.tw

Abstract—**At present, trajectory data can be readily obtained due to the advent of positioning technologies. Clustering of trajectories and giving meanings to the resulting clusters is an active research area. Recently, we proposed an analysis approach that clusters trajectories in two steps: the first step based on data distribution and the second step based on state transition. In this approach, for coping with the distinguished characteristic of each trajectory, a map of interest is dynamically divided into multiple states, according to the trajectory distribution, and a quadtree is generated for each trajectory. The first-step clustering is then performed based on the differences between the quadtrees. For all trajectories in a resulting cluster of interest, the second-step clustering is further performed based on the differences in their state-to-state transition probabilities using a proposed method for comparing a pair of trajectories with different quadtree structures. After presenting a procedure for visualizing a cluster of interest in order to interpret its movement behaviors, we give and discuss a case study where our approach is applied to real trajectory data obtained from Angel Love Online, a massively multiplayer online game, and player behaviors in the target map become manifest.**

I. INTRODUCTION

Main objectives in the study of mobile-object trajectories are to understand the current status of a mobile object of interest and to predict its next action or position. Important technical issues, for achieving these objectives, include (i) determining the distance or similarity between trajectories, (ii) clustering the trajectories based on a given distance or similarity index, and (iii) finding the meaning in each mobile object movement. As for online games, research findings in this particular area are useful in (re)designing a game map of interest as well as the content therein and in providing personalized services to players such as personal route guides.

In related work [1, 2], methods were proposed that express and compare trajectories based on transitions between common states. Therein, the target space is divided into multiple areas, defined as states, and each trajectory is then expressed by the transition probabilities between those common states. Other methods were proposed in [3, 4] that select a smaller number of important states, called landmarks, from the common states and express each trajectory by the landmark-to-landmark transition probabilities. However, use of the common states or the selected landmarks has two problems: over-approximation of trajectory transitions between states and omission of fine movements within a given state. An attempt to solve these problems by increasing the number of states or landmarks will inevitably lead to an increase in computational cost. Rather, the target space should be divided into suitable states for each trajectory, leading to the need for a method for comparing a pair of trajectories with different state sets.

Recently, we proposed an approach for analyzing user trajectories that consists of two steps [5]. In the first step, dynamic map division is performed that generates a quadtree, represented by a bit sequence, for each trajectory, and clustering is performed based on the Hamming distances between all trajectory quadtrees. In the second step, for all trajectories in a cluster of interest, their differences in the state-to-state transition probabilities are derived using a proposed method for comparing a pair of quadtrees, and clustering of those trajectories is further performed. Compared to the other approaches, this two-step approach has less computational costs and enables easier interpretation of clustering results.

In this paper, we present a case study in which the above approach is applied to real trajectory data obtained from Angel Love Online (ALO), a massively multiplayer online game. For this paper to be self contained, the analysis approach in use is described in the next section. After a procedure for visualizing a cluster of interest in order to interpret its movement behaviors, the case study is then given including our discussions as well as interpretations of resulting clusters.

II. Analysis Approach

A. 1st-Step Clustering based on Data Distribution

For each trajectory, a map of interest is divided according to the trajectory data distribution. Let D indicate the level of division. The initial node (or state) where $D = 0$, is divided into four areas. After evaluating its data density, each area will be further divided into another four areas if the density is higher than a given threshold. The concept of dynamic map division is depicted in Fig. 1.

A quadtree is used to conceptually represent the resulting division of a trajectory of interest. Under this representation, the initial state is the root of the tree with node number = 0, and incrementally extending the tree corresponds to iteratively dividing the map. At the end of division, if there are some nodes that have no trajectory, all such nodes will be deleted. Figure 2 shows an example of map division and the resulting quadtree.

For subsequent algorithmic manipulation, a unique identification number based on Z-ordering is assigned to each node in every quadtree. Note that all nodes with the same identification number in their quadtrees represent the same area in the given map. A bit sequence for each quadtree is used in quadtree comparison. In each bit sequence, the n th bit indicates existence or nonexistence of the n th node, i.e., if the n th node exists, the n th bit is set to 1; and otherwise 0.

In the first-step clustering, all N trajectories are clustered with the Ward method [6] whose element in the $(N-1)\times(N-1)$ input distance matrix is the distance between a corresponding trajectory pair. For a pair of trajectories of interest, the Hamming distance between the corresponding bit sequences is used. The number of clusters is decided with the rating index introduced in [6], and any cluster with member trajectories less than ten percent of N is excluded because they are considered outliers.

B. 2nd-Step Clustering based on State Transitions

At the second step, the direction information in each trajectory is approximated by its state-to-state transition probabilities. However, care must be taken here because quadtrees are different from one trajectory to another. As a result, the positions of non-empty elements in their state-to-state matrices are also different. Therefore, for a given pair of trajectories with different quadtree structures, a common quadtree structure is derived by logical multiplication of the corresponding two bit sequences, as shown in Fig. 3.

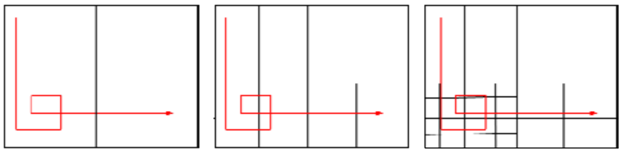

Figure 1 Concept of dynamic map division according to the trajectory data distribution. From left to right D=1, D=2, D=3.

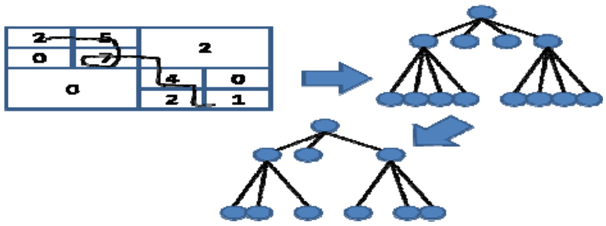

Figure 2 Example of map division and the resulting quadtrees before and after deletion of nodes, where the number of trajectory data is shown in each area.

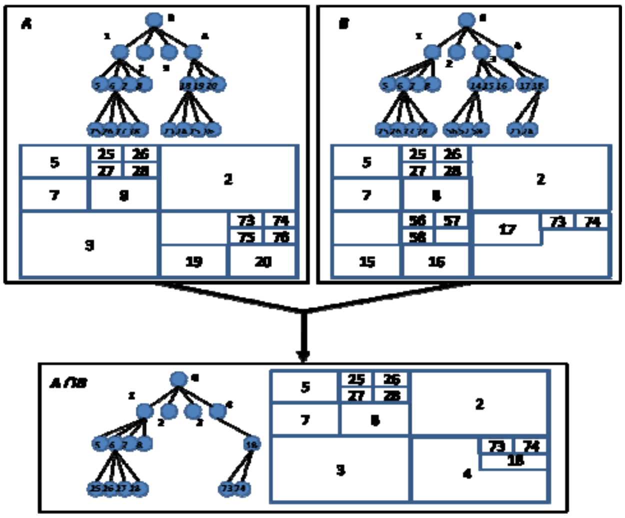

Figure 3 Example of derivation of a common structure for two quadtrees A and B having different structures, where the node number is assigned to each area.

Once a common structure for a pair of quadtrees is derived, an adjustment is performed for each of their state-to-state transition probabilities associated with any state that unifies its former child states in the original structure. Henceforth, such a sate is called a unifying state. For a unifying state of interest, say, state *i*, the transition probability from state *j* to state *i* is the sum of the transition probabilities to the former child states of *i* from *j* while the transition probability from state *i* to state *j* is the average of the transition probabilities from the former child states of *i* to *j* as follows:

i) When only the destination of the transition is a unifying state,

$$a_{ml'} = \sum_{i=1}^{A(l')} a_{ml'(i)} \tag{1}$$

ii) When only the source of the transition is a unifying state,

$$a_{l'm} = \frac{1}{A(l')} \sum_{i=1}^{A(l')} a_{l'(i)m} \tag{2}$$

iii) When both the source and the destination of the transition are a unifying state,

$$a_{l'l'} = \frac{1}{A(l')} \sum_{i=1}^{A(l')} \sum_{j=1}^{A(l')} a_{l'(i)l'(j)} \, , \tag{3}$$

where a_{xy} is the transition probability from state *x* to *y*, *l'* indicates a unifying state of interest, *A*(*l'*) contains the indices to all former child states of *l'*. For example, in Fig. 3, if *l'* is 3, *A*(3) contains 15, 16, 56, 57 and 58. As a result of common-structure derivation, an intermediate node might become a state. For example, in Fig. 3, intermediate nodes 4 and 18 become states.

Next, the distance between two trajectories is defined as the average of the differences in the elements of the corresponding state-to-state transition matrices. This distance is used as an element of the $(N-1)\times(N-1)$ distance matrix in the second-step clustering, where weights are given to the state-to-state transition probabilities, considering the importance of differences in earlier division levels more important than those in later levels, as follows:

$$w_l = \frac{D_{\max} - D_l + 1}{D_{\max}} \, , \tag{4}$$

where D_l is the division level of area *l*, and D_{max} is the maximum of D_l. The aforementioned distance between trajectories *i* and *j* is given as follows:

$$Dist_{ij} = \frac{1}{h_{num}} \sum_{l=1}^{h_{num}} \sum_{m=1}^{h_{num}} w_l w_m \left| a_{lm}(i) - a_{lm}(j) \right| , \tag{5}$$

where $a_{lm}(x)$ is the transition probability from states *l* to *m* in trajectory *x*, and h_{num} indicates the number of common states between *i* and *j*. If D_{max} of the common structure is 0, the distance between these particular two trajectories will be given the maximum possible value of (5), i.e., 1. After obtaining the distance matrix for a cluster of interest, the trajectories in this cluster will be further clustered with the same method described in Ⅱ.A.

III. CASE STUDY WITH ALO

We present here a case study where the analysis approach in II is applied to ALO, developed by UserJoy Tech. Co. Ltd, a leading game designer and publisher in Taiwan. This set of trajectory data was collected from the map in Fig. 4 for about 70 hours. In this map, the blue area is a sea on which players cannot move, and the sea separates the map into two sides connected via a small island in the center. In addition, a town, where players go to receive a service, such as, a quest or assistance, is located near the right of the map center, and monsters, ready to attack nearby players, reside in the left, top-right, and bottom-right regions.

Once clusters are obtained using the analysis approach in II, we use the following procedure for visualizing a cluster of interest in order to interpret its movement behaviors:

i) Perform dynamic map division based on all member trajectories in the cluster

ii) Visualize the division result with grids in order to elucidate the trajectory distribution

iii) Visualize the major state-to-state transitions with arrows in order to provide the information on trajectory directions.

Table 1 shows the clustering results, where the 1st column and the 2nd column indicate the resulting cluster numbers at the first step and the second step, respectively. At the first step, all trajectories were grouped into two clusters. They were further sub-grouped into 21 clusters at the second step. Out of these, all clusters with the number of trajectories higher than 25, i.e., clusters 1, 2, 7, 10, 20, and 21, are discussed below.

Figures 5 and 6 show the visualized results of clusters 1, 2, 7, 10 and those of clusters 20 and 21, respectively; where the direction of each of the highest ten state-to-state transitions in each cluster is depicted by an arrow, and we note that fine grids with no arrows indicate regions players spent long time. In Fig. 5, grid patterns in the map top of clusters 1, 2 and 10 are similar, and those below the left of the map center of clusters 2, 7, and 10 are also similar. In Fig. 6, grid patterns of clusters 20 and 21 are similar in around the town. From these results, a trajectory is assigned to one of the two clusters at the first step clustering according to whether its distribution concentrates in regions outside or inside the town. Our interpretation for the first-step clustering is that the play objective of cluster 1 is to mainly fight monsters outside of the town while that of cluster 2 is to mainly use services inside the town.

Now we discuss clusters 1, 2, 7, and 10 in detail considering both map division (grid) and transition (arrow) patterns as follows:

i) In cluster 1, because an arrow between the two sides of the map and fine grids in the top-left region, these players moved back and forth across the small island to fight monsters in the top-left region and to use provided services in the town.

ii) In cluster 2, having many arrows in the bottom of the map indicates that these players repeated fighting those monsters therein, and, because of fine grids in bottom-left quarter of the map, some events might have occurred and thus attracted players to the region.

iii) In cluster 7, because arrows and fine grids can be seen in the bottom of the map, players in this cluster concentrated on fighting monsters.

iv) In cluster 10, arrows from left to right across the small island and fine grids in the bottom indicate that these players started movement from the top left of the map and then went to the right side in order to fight monsters down there.

Our interpretations for clusters 20 and 21 are as follows:

i) In cluster 20, because of many arrows and fine grids near the map center, players of this cluster repeatedly visit the town in order to use services provided therein.

ii) In cluster 21, because arrows and fine grids below the town and the top right of the town, these players repeatedly visited the town and then went to fight monsters.

IV. Conclusions and Future work

This paper presented a case study where we applied our recently proposed analysis approach to ALO trajectory data. A procedure for visualization of the resulting clusters was given. Interpretation of movement behaviors in each major cluster was conducted based on the visualized results and the a priori knowledge of the map context.

From trajectory data, movement speed can also be derived. The information on speed is useful for finding the detail meaning in mobile object movement. In future, we plan to incorporate the speed information into our approach and to experiment it with other kinds of movement data, such as player trajectories from pervasive games and those in 3D virtual museums.

Acknowledgements

This work was supported in part by Japan Society for Promotion of Science (JSPS) under the Grant-in-Aid for Scientific Research (C) 20500146. The authors are indebted to Tsing-San Cheng at UserJoy Tech. Co. Ltd, who helped us gather the movement logs.

References

[1] Yoshiharu Ishikawa, Yoji Machida, and Hiroyuki Kitagawa, "A Dynamic Mobility Histogram Construction Method Based on Markov Chains", Proc. 18th International Conference on Scientific and Statistical Database Management (SSDBM 2006), pp. 359-368, 2006.

[2] Naohiko Suzuki, Kosuke Hirasawa, Kenichi Tanaka, Yoshinori Kobayashi, Yoichi Sato, and Yozo Fujino, "Learning Motion Patterns and Anomaly Detection by Human Trajectory Analysis", Proc. IEEE Int. Conf. Systems, Man and Cybernetics (SMC2007), pp. 498-503, 2007.

[3] Ruck Thawonmas, Masayoshi Kurashige, Keita Iizuka, and Mehmed Kantardzic, "Clustering Online Game Users Based on Their Trails Using Self-Organizing Map," Lecture Notes in Computer Science, Richard Harper et al. (Eds.), vol. 4161 (ICEC 2006), pp. 366-369, 2006.

[4] Ruck Thawonmas, Masayoshi Kurashige and Kuan-Ta Chen, "Detection of Landmarks for Clustering of Online-Game Players," International Journal of Virtual Reality, vol. 6, no. 3, pp. 11-16， 2007.

[5] Junichi Oda, Ruck Thawonmas, and Kuan-Ta Chen, "Comparison of User Trajectories Based on Coordinate Data and State Transitions," Proc. the Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2009), Kyoto, Japan, pp. 1134-1137, Sep.12-14, 2009.

[6] T. Calinski and J. Harabasz, "A Dendrite Method for Cluster Analysis", Communication in Statistics, Vol. 3, pp. 1-27, 1974.

Figure 4 The ALO map used in our case study.

Table 1 Clustering results.

1st	2nd	Number of trajectories
1	1	56
	2	28
	3	4
	4	6
	5	5
	6	5
	7	37
	8	10
	9	8
	10	27
	11	1
	12	17
	13	8
	14	1
	15	1
	16	3
	17	2
	18	1
	19	1
2	20	56
	21	117

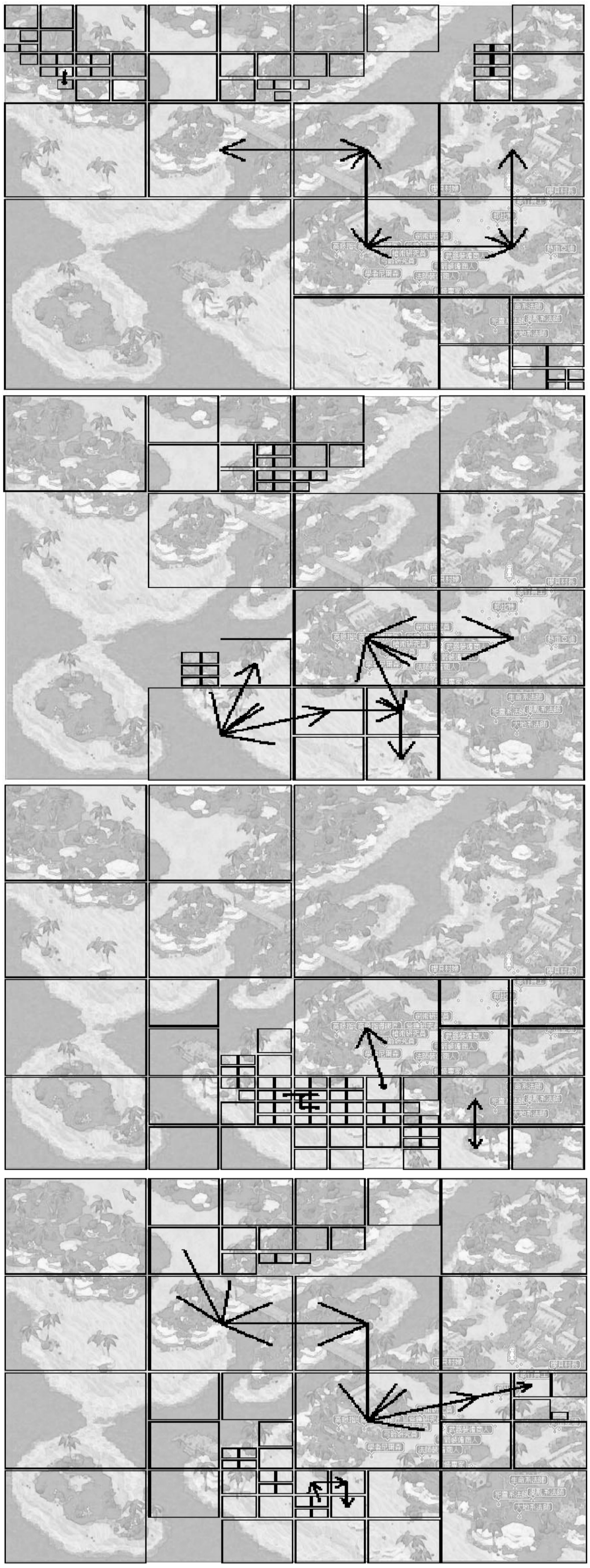

Figure 5 Visualization of divisions and transitions for, from top to bottom, cluster 1, cluster 2, cluster 7, and cluster 10.

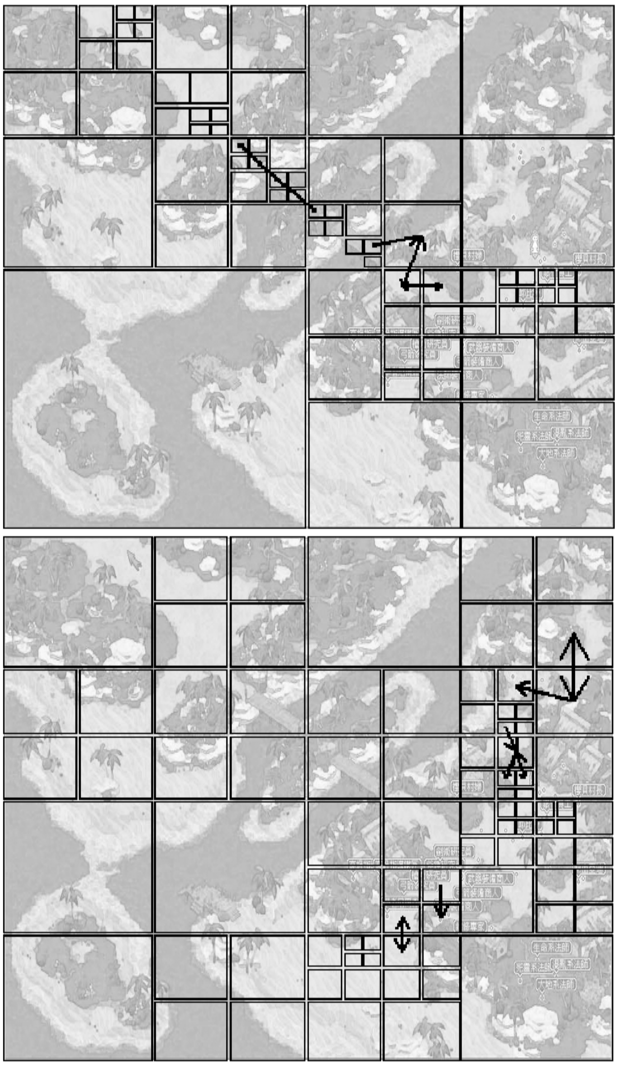

Figure 6 Visualization of divisions and transitions for, from top to bottom, cluster 20 and cluster 21.