Mining Mobile Group Patterns: A Trajectory-Based Approach

San-Yih Hwang[1], Ying-Han Liu[1], Jeng-Kuen Chiu[1], and Ee-Peng Lim[2]

[1] Department of Information Management National Sun Yat-Sen University, Kaohsiung, Taiwan 80424

[2] School of Computer Engineering Nanyang Technological University, Singapore 639798, Singapore

Abstract. In this paper, we present a group pattern mining approach to derive the grouping information of mobile device users based on a trajectory model. Group patterns of users are determined by distance threshold and minimum time duration. A trajectory model of user movement is adopted to save storage space and to cope with untracked or disconnected location data. To discover group patterns, we propose ATGP algorithm and TVG-growth that are derived from the Apriori and VG-growth algorithms respectively.

1 Introduction

Behavior research on sociology show that peer pressure and group conformity can affect the buying behaviors of individuals [1]. With a good knowledge of groups a customer belongs to, one can derive common buying interests among customers, and develop group-specific pricing models or marketing strategies for personalized services. There are many ways one can determine the groups a person belongs to, for example, by the set of product items s/he purchased, his/her occupation or income, and/or the places s/he visited. As implied by the loads of research in spatial-temporal databases [7], the information about users' locations over time can play a crucial role in determining the groups. As mobile phones and other similar devices become widely used, users' locations of errors usually less than 1km can be gathered by mobile communication operators using the existing communication infrastructure. With more accurate positioning technologies, the errors can be reduced even further.

In this research, we are interested in discovering groups of users such that users in the same group are geographically close to one another for significant amounts of time. Finding such grouping information of mobile users, based on the spatio-temporal distances among them, is known as "Group patterns mining", originally proposed by Wang et al. [9, 10]. Previous research represents the movement data of an object as a synchronous time series of locations. This representation, however, has the following three pitfalls:

1. To maintain accurate location tracking, the frequency of sampling users' locations must be high. As a result, the movement database can be become huge.
2. Moving objects may be disconnected from time to time voluntarily or involun arily. It is therefore not realistic to assume that the location information is present for each time point.

T.B. Ho, D. Cheung, and H. Liu (Eds.): PAKDD 2005, LNAI 3518, pp. 713 – 718, 2005.
© Springer-Verlag Berlin Heidelberg 2005

3. Lastly, it is almost impossible to have perfectly synchronized sampling of users' locations in reality due to clock differences of base stations conducting the sampling and the locations of moving objects.

To deal with the first and the third problems, a trajectory-based model to represent object movements can be adopted instead [4, 5, 6, 8]. A trajectory is a function that maps time to locations. To represent object movement, a trajectory can be decomposed into a set of linear functions, one for each disjoint time interval. The derivative of each linear function yields the direction and speed in the associated time interval. Various approaches have been proposed to accurately induce the trajectory of an object from its location update data while saving storage space using dead-reckoning policies [11] or regression techniques [2]. In this paper, we use trajectories for modeling moving objects and develop efficient algorithms for discovering mobile group patterns from trajectory data. Furthermore, we address the second problem by allowing the trajectory of each object not to cover the entire location tracking period.

The rest of the paper is organized as follows. In Section 2, we formally define the mobile group pattern discovery problem in the context of using trajectories to represent moving objects. In Section 3, we describe the algorithms for discovering mobile group patterns. Finally, we conclude in Section 4.

2 Problem Definition

A trajectory is a set of piecewise linear functions, each of which maps from a disjoint time interval to an n-dimensional space. That is, one can perceive a piece of a trajectory as a set of n linear functions of the time variable t, one for each dimension, and the trajectory may change speed and direction at finitely many time instants. Each linear piece can be represented as a conjunction of linear constraints using the time variable and coordinate variables. A trajectory is a disjunction of all its linear pieces. For example, a trajectory of the user moving on a 2-D space may consist of 3 linear pieces as shown below:

$$
\begin{aligned}
&[(x = 2t - 40) \wedge (y = -t + 23) \wedge (0 \le t < 21)] \\
&\vee[(x = 2) \wedge (y = -t + 23) \wedge (21 \le t < 22)] \\
&\vee[(x = 0.5t - 9) \wedge (y = 1) \wedge (22 \le t \le 30)]
\end{aligned}
$$

An *object movement database D* consists of a set of trajectories, one for each object. That is, $D = \bigcup_{i=1}^{M} T_i$, where M is the number of moving objects. Each linear piece in a trajectory T_i is a set of 4-tuples: (*reference_point*, *velocity*, *start_time*, *end_time*), denoting the location function $f(t) = velocity \times t + reference\_point$ during time interval [*start_time*, *end_time*). Table 1 shows the trajectories of three example objects in a 2-dimensional space. Note that the trajectory of each object may become untraceable at some time points, resulting in a sequence of non-continuous linear pieces. As shown in Table 1, moving object o_1 is disconnected in time [5, 6) and [9, 10), object o_2 is disconnected in time [5, 6), and object o_3 is untraceable during time interval [8, 10).

Table 1. An example object movement database

	reference_point	velocity	start_time	end_time
o_1	(1,1)	(3,1)	0	3
	(7,-11)	(1,5)	3	5
	(10,-3)	(4,3)	6	9
o_2	(2,2)	(2,1)	0	3
	(2,-13)	(2,6)	3	5
	(-4,5)	(3,2)	6	10
o_3	(2,4)	(3,1)	0	3
	(17.-5)	(-2,4)	3	5
	(12,35)	(-1,-4)	5	8

Definition 1. Given a group of objects G and a maximum distance threshold *max_dis*, we say objects in G are *geographically close* at a time point t if every pair of objects in G are no farther than *max_dis* apart, and *geographically far* at t if there exists one pair of objects in G whose distance is larger than *max_dis*. We also say objects in G is *geographically decided* if they are either close or far, and *geographically undecided* otherwise.

Definition 2. Given a group of objects G and a minimal time duration threshold *min_dur*, a time interval $[t,t+k]$ is called a close interval of G if

1. objects in G are geographically close at any time point in $[t, t+k]$,
2. objects in G are not geographically close at time $t-\varepsilon$, where ε is an arbitrarily small positive number,
3. objects in G are not geographically close at time $t+k+\varepsilon$, where ε is an arbitrarily small positive number, and
4. $k \geq min\_dur$.

A far interval can be similarly defined.

A group of objects G, *max_dis*, and *min_dur* are said to form to a group pattern, denoted $P=(G, max\_dis, min\_dur)$ [9]. Given an object movement database, a group pattern may have multiple close intervals and multiple far intervals, within which the geographical property associated with G can be decided. For the time points not covered by any close or far intervals, aggregated as the undecided intervals, the geographical property associated with G is not clear. We quantify the significance of a group pattern by the proportion of the total length of close intervals and *estimated* close subintervals of the undecided intervals.

Definition 3. Let P be a group pattern with n close intervals $c_1, c_2, \ldots, c_n$, m far intervals $f_1, f_2, \ldots, f_m$, and k undecided intervals $u_1, u_2, \ldots u_k$. The weight of P is defined as:

$$weight(P) = \frac{L_{close} + L_{undecided} \cdot \dfrac{L_{close}}{L_{close} + L_{far}}}{L_{close} + L_{far} + L_{undecided}} = \frac{L_{close}}{L_{close} + L_{far}}, \tag{1}$$

where $L_{close} = \sum_{i=1}^{n} c_i$, $L_{far} = \sum_{i=1}^{m} f_i$, and $L_{undecided} = \sum_{i=1}^{k} u_i$.

The weight represents the proportion of the time when users of P (are expected to) stay close. Thus, the larger is the weight, the more significant is the group pattern.

Definition 4. Given a threshold *min_wei*, a group pattern P= <G, *max_dis*, *min_dur* >, is valid if the weight of P exceeds the threshold *min_wei*.

The problem is how to identify all valid group patterns given a trajectory-based object movement database and the thresholds *max_dis*, *min_dur*, and *min_wei*.

3 The Algorithms

The geographical property of a mobile group (i.e., close or far) is determined by the distances between all pairs of objects at any time point. Here, the distance function of o_1 and o_2 for each corresponding linear piece (i.e., with the same time interval) can be easily computed. For example, suppose the location functions of objects o_1 and o_2 between time 0 and 3 are as follows:

Location of object o_1 at time t: $(1 + 3t, 1 + t)$
Location of object o_2 at time t: $(2 + 2t, 2 + t)$

The Euclidean distance between o_1 and o_2 when $0 \le t < 3$ is $\sqrt{(1-t)^2 + (1)^2}$

A complete distance function between o_1 and o_2 whose location data listed in Table 1 is the following:

$$distance_{o_1,o_2}(t) = \begin{cases} \sqrt{(1-t)^2 + 1^2}, 0 \le t < 3 \\ \sqrt{(-5+t)^2 + (-2+t)^2}, 3 \le t < 5 \\ undecided, 5 \le t < 6 \\ \sqrt{(-14-t)^2 + (8-t)^2}, 6 \le t < 9 \\ undecided, 9 \le t \le 10 \end{cases}$$

Given a distance function $dist(t)$ of two objects o_1 and o_2 within an interval I, we would like to identify the subintervals I' in I such that $dist(t) \le max\_dis$, $t \in I'$. This can be done by computing the roots of the equation $dist(t) = max\_dis$. In case of Euclidean distance, there will be two roots, denoted t_a and t_b, where $t_a \le t_b$. When both t_a and t_b are real numbers, we have $dist(t) \le \max\_dis, \forall t_a \le t \le t_b$. Obviously, within the time interval $I \cap [t_a, t_b]$, the distance between o_1 and o_2 is no more than *max_dis*, and at any time in $I - [t_a, t_b]$, the distance between o_1 and o_2 is greater than *max_dis*. The weight of a mobile group can be subsequently decided by looking at each pair of objects in the group, and the function for computing the weight of a group c_k is named Group-Weight(c_k). Group-Weight(c_k) starts with synchronizing the linear pieces of the trajectory in each trajectory of c_k such that each trajectory has the same set of time segments, followed by computing the close and far intervals in each time segment. The pseudo-code of Group-Weight(c_k) is omitted due to space limitation.

Given two group patterns, P = <G, *max_dis*, *min_dur* > and P' = <G', *max_dis*, *min_dur* >, P' is called a sub-group pattern of P if $G' \subseteq G$. The Apriori property within the context of mobile group pattern mining states that any sub-group pattern of a valid group patterns must also be valid. In [9], both AGP and VG-growth algorithms are based on this Apriori property. To re-use both algorithms, we must ensure that Apriori property still holds for trajectory-based movement data.

Theorem 1. [Apriori property for group patterns] Given a database D and thresholds *max_dis*, *min_dur*, and *min_wei*, if a group pattern is valid, all of its subgroup patterns will also be valid.

Based on the AGP algorithm [9], we develop an algorithm called Apriori Trajectory-based Group Pattern Mining, abbreviated ATGP, whose pseudo-code is shown in Fig. 1. In the algorithm, we use C_k to denote the set of candidate k-groups, and G_k to denote the set of valid k-groups. From G_1, the set of all distinct objects, the algorithm first computes C_2, the pair set of objects in G_1. This algorithm performs join operation to generate candidate k groups C_k from G_{k-1} (C_k=Generate_Candidate_Groups(G_{k-1})), and the generated candidates are verified by computing their weights.

Input: D, *max_dis*, *min_dur*, and *min_wei*
Output: all valid groups G
$G=\varnothing$; G_1=all distinct objects;
for ($k = 2$; $G_{k-1} \neq \varnothing$; k++)
  C_k = **Generate_Candidate_Groups**(G_{k-1}); $G_k=\varnothing$;
  for each candidate k-group $c_k \in C_k$
    c_k.*weight* = **Group-Weight**(c_k, *max_dis*, *min_dur*);
    if (c_k.*weight* >= *min_wei* } $G_k = G_k \cup c_k$;
  $G = G \cup G_k$;
return G;

Fig. 1. Algorithm ATGP

In [WL03], Wang et al. proposed a data structure VG-graph whose edges represent all valid 2-groups and an algorithm VG-growth that traverses VG-graph to identify all valid groups. We adapt VG-graph to work for trajectory-based object movement database and call the resultant data structure TVG-graph, and the traversal algorithm TVG-growth. Similar to VG-graph, the edges of TVG-graph are determined by the set of valid 2-groups. However, since each object may have some untraceable periods, every edge in TVG-graph is associated with a set of close intervals as well as another set of far intervals. The group mining procedure of TVG-growth remains the same as that of VG-growth, however, the set of close and far intervals associated with each edge has to be updated as the recursive mining procedure proceeds [3].

4 Conclusion

In this paper, we reported a novel approach that discovers moving object group patterns from a database comprising trajectories of moving objects. Furthermore, our research allows non-continuous trajectories which model the disconnected behavior of moving objects.

In this work, the location of an object at a time point was assumed to be either accurately determined or completely unknown. In some applications, location data of an object may incur different degrees of uncertainties over time. Our future work includes mining mobile group patterns by considering the inherent uncertainty of location data.

References

1. D.R. Forsyth. *Group Dynamics*. Wadsworth, Belmont, CA, 1999.
2. V. Guralnik, J. Srivastava. "Event Detection from Time Series Data." Proceedings of ACM International Conference on Knowledge Discovery and Data Mining (KDD2000), 2000.
3. Y.-H. Liu "Mining Mobile Group Patterns: A Trajectory-based Approach," *master thesis, Department of Information Management, National Sun Yat-sen U.*, available at http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0730104-102312.
4. H. Mokhtar, J. Su, and O.H. Ibarra, "On moving object queries", *Proceedings of the ACM Symposium on Principles of Database Systems (PODS)*, 2002.
5. H.K. Park, J.H. Son, M.-H. Kim, "An Efficient Spatiotemporal Indexing Method for Moving Objects in Mobile Communication Environments." *Proc. Of Int'l. Conf. on Mobile Data Management (MDM*2003), 2003.
6. S. Saltenis, C.S. Jensen, S.T. Leutenegger, M.A. Lopez, "Indexing the positions of continuously moving objects," *Proc. of 2000 ACM SIGMOD Conference*, 2000.
7. S. Shekhar and S. Chawla, "Introduction to Spatial Data Mining," Chapter 7, *Spatial Databases: A Tour*, Prentice Hall, New Jersey, 2003.
8. M. Vazirgiannis and O. Wolfson, "A spatiotemporal model and language for moving objects on road networks", *Proc. of Symposium on Spatial and Temporal Databases* (SSTD), 2001.
9. Y. Wang, E.-P. Lim, and S.-Y. Hwang, "On Mining Group Patterns of Mobile Users." *Proc. Of the 14th International Conference on Database and Expert Systems Application* (*DEXA 2003*), 2003.
10. Y. Wang, E.-P. Lim, and S.-Y. Hwang, "Effective Group Pattern Mining Using Data Summarization," 9*th International Conference on Database Systems for Advanced Application* (DASFAA2004), 2004.
11. O. Wolfson, A. P. Sistla, S. Chamberlain, Y. Yesha, "Updating and querying databases that track mobile units," *Distributed and Parallel Databases*, 1999.