MoveMine 2.0: Mining Object Relationships from Movement Data

Fei Wu† Tobias Kin Hou Lei‡ Zhenhui Li† Jiawei Han‡
† Pennsylvania State University ‡ University of Illinois at Urbana-Champaign
†{fxw133,jessieli}@ist.psu.edu ‡{klei2,hanj}@illinois.edu

ABSTRACT

The development in positioning technology has enabled us to collect a huge amount of movement data from moving objects, such as human, animals, and vehicles. The data embed rich information about the relationships among moving objects and have applications in many fields, *e.g.*, in ecological study and human behavioral study. Previously, we have proposed a system MoveMine that integrates several start-of-art movement mining methods. However, it does not include recent methods on relationship pattern mining. Thus, we propose to extend MoveMine to MoveMine 2.0 by adding substantial new methods in mining dynamic relationship patterns. Newly added methods focus on two types of pairwise relationship patterns: (i) attraction/avoidance relationship, and (ii) following pattern. A user-friendly interface is designed to support interactive exploration of the result and provides flexibility in tuning parameters. MoveMine 2.0 is tested on multiple types of real datasets to ensure its practical use. Our system provides useful tools for domain experts to gain insights on real dataset. Meanwhile, it will promote further research in relationship mining from moving objects.

1. INTRODUCTION

The rapid advancement in modern positioning technology enables easy collection of large-scale movement data from a variety of objects. For instance, animal scientists attach sensor tags on animals to track their movement and mobile users share their location information via smartphones. Many methods have been proposed to utilize the movement data [11]. Such movement data capture the dynamic relationships among objects and have applications in many fields, such as in ecological study and human behavioral study.

In previous work, we have proposed a system, MoveMine [7], which integrates several state-of-art moving object pattern mining and trajectory pattern mining methods. However, it does not include *recent* relationship mining methods.

Proceedings of the VLDB Endowment, Vol. 7, No. 13

To accommodate the needs in mining relationship patterns from movement data, we extend our system to MoveMine 2.0 by incorporating several recent methods in *mining relationship patterns*. In particular, we focus on two types of pairwise relationship patterns from most *recent* work: (i) attraction and avoidance relationship [6], (ii) following relationship [8]. Our system also supports group-level relationship analysis by constructing relationship network and matrix.

MoveMine 2.0 provides users the flexibility in parameter tuning and supports visualization of the result in different formats (*e.g.*, map, network, and matrix). The result can also be exported to Google Map[1] and Google Earth[2] formats for interactive exploration.

The development of MoveMine 2.0 is motivated by the Movebank[3] project with a data repository containing hundreds of different animal movements. We tested all the functions of MoveMine 2.0 on animal movements (*e.g.*, Movebank) as well as human movements (*e.g.*, Reality Mining dataset[4] and Geolife dataset[5]).

The MoveMine 2.0 is publicly downloadable at `http://faculty.ist.psu.edu/jessieli/MoveMine/`. We will maintain the software and provide updates at the same link.

2. GENERAL SYSTEM ARCHITECTURE

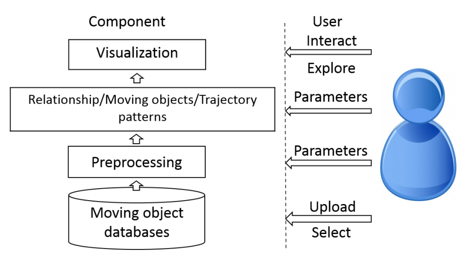


Figure 1: MoveMine 2.0 system architecture

Figure 1 demonstrates the system architecture of MoveMine 2.0 that has four components as its essences: (i) real world data repository, (ii) data pre-processing, (iii) relationship mining, and (iv) visualization. MoveMine 2.0 is closely

[1]http://www.google.com/maps/
[2]http://www.google.com/earth/
[3]http://www.movebank.org/
[4]http://reality.media.mit.edu/
[5]http://research.microsoft.com/en-us/projects/geolife/

integrated with Movebank repository. In addition, we can also gather datasets from various sources, such as geo-tagged social media (*e.g.*, Twitter), vehicles data from GPS, and human movement recorded by mobile phones (*e.g.*, Geolife and Reality Mining datasets). As the raw movement data are often unevenly sampled and may contain long periods of missing data, having a preprocessing step to interpolate missing data is essential. A user can either use the default settings or specify own parameters.

Pairwise relationship mining modules then operate on top of the processed data. In particular, MoveMine 2.0 focuses on two types of pairwise relationship patterns that are attraction/avoidance relationship and following pattern. We design effective and scalable algorithms to address the challenges in detecting both relationship patterns. We will present the details in Section 3. The result can be shown in different formats (*e.g.*, network and matrix) and exported to other visualization tools (*e.g.*, Google Map and Google Earth).

3. MAJOR MODULES

3.1 MoveMine functions

The previous MoveMine system integrates several state-of-art movement mining methods. The methods can be divided into two categories. The first category, *moving objects pattern mining*, includes moving objects clustering method (*i.e.*, swarm pattern) and periodic pattern detection method. The second, *trajectory pattern*, focuses more on the geometric shapes of trajectories. Accordingly, trajectory clustering, trajectory outlier detection, and trajectory classification methods are included. For details, readers can refer to our previous paper [7].

In MoveMine 2.0, we add substantial new functions that focus on two types of relationship patterns. It also incorporates several baseline methods for comparison.

3.2 Attraction and avoidance relationship

The attraction relationship is commonly seen in animal herds or human groups (*e.g.*, colleague or family). Meanwhile, the avoidance relationship naturally exists among moving objects. For example, in animal movements, preyers avoid predators, different animal groups of the same species tend to avoid each other, and subordinate members tend to avoid more dominant group-mates within a group.

In the literature, study of moving object relationship has been largely restricted to attraction relationship only. Particularly, various similarity measures [2, 3, 10, 9] have been utilized to quantify the strength of attraction. Meanwhile, moving object patterns, included in MoveMine, are detected by counting the frequency of objects being spatially close, *i.e.*, *meeting frequency.* Those studies are based on the assumption that the smaller the distance is or the higher the meeting frequency is, the stronger the attraction relationship is.

Though the similarity measures and the meeting frequency provide indications of the closeness among moving objects, we cannot simply conclude whether there is an attraction or avoidance relationship between objects. For example, two animals may be observed to be spatially close for 10 out of 100 timestamps. But is this significant enough to conclude attraction relationship between them? Furthermore, another pair of animals are being spatially close for 20 out of 100 timestamps. Can we say the latter pair is more significantly attracted to each other? Finally, supposedly two animals are never being spatially close, do they necessarily have an avoidance relationship?

To determine the type as well as the degree of relationship, we design a method in [6] that considers the *background territories* of moving objects. Given two movement sequences R and S, the method determines the relationship by comparing the *expected* meeting frequency $E[freq(R,S)]$ with the *actual* meeting frequency $freq(R,S)$. Intuitively, if they meet less (or more) then expected, the relationship is likely to be avoidance (or attraction).

However, one cannot determine a universal *degree of relationship* from the simple comparison. We propose to determine the degree of relationship via *permutation test* under the null hypothesis that R and S are independent. In permutation test, we randomly permutate orders in the movement sequence. The intuition is that the meeting frequency between permutated trajectories of R and S should be similar to the *actual* meeting frequency, if R and S are independent. If the *actual* meeting frequency between R and S is higher or lower than certain percentage (e.g., 95%) of the permutated results, we reject the hypothesis and claim that R and S have significantly non-independent relationship (i.e., attraction or avoidance). More specifically, let σ and σ' denote two random permutations. We define the significance value of the relationships as:

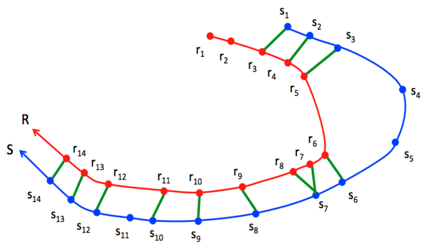
$$\begin{aligned}
\mathrm{sig}_{\mathrm{attract}}(R,S) &= \Pr\left[freq(R,S) > freq(\sigma(R),\sigma'(S))\right] \\
&\quad + \frac{1}{2}\Pr\left[freq(R,S) = freq(\sigma(R),\sigma'(S))\right], \\
\mathrm{sig}_{\mathrm{avoid}}(R,S) &= \Pr\left[freq(R,S) < freq(\sigma(R),\sigma'(S))\right] \\
&\quad + \frac{1}{2}\Pr\left[freq(R,S) = freq(\sigma(R),\sigma'(S))\right].
\end{aligned}$$

As the number of permutations is factorial, Monte Carlo sampling is used to approximate the significance value. The method is scalable with proposed pruning techniques. MoveMine 2.0 incorporates both standard measures and the newly proposed attraction/avoidance detection method.

3.3 Following pattern

Another interesting and dynamic relationship is the following pattern. For example, a predator may follow its prey, and group members may follow their leader during migration. Intuitively, a follower has similar trajectory as its leader but always arrives at a location that the leader visited with some time lag. However, detecting such behavior is a non-trivial task. The challenges lay in three aspects: (i) the follower may not have the exact same trajectory as the leader; (ii) the following time lag is usually unknown and varying; (iii) the following relationship may happen in a short period of time.

A limited number of methods have been proposed to detect following patterns in movement data. In the REMO framework [1, 5], a leader should appear in the front region of the follower(s), which is a fan-shaped area in front of a follower. However, this assumption is often violated in real scenarios, *e.g.*, the follower might take a detour. Furthermore, in such a dynamic relationship, the speed of objects can change rapidly (resulting in varying time lag). The leader can easily go out of the front region of its follower(s). Thus, the definition of the front region is overly restricted.



r_i	r_1	r_2	r_3	r_4	r_5	r_6	r_7
nearest point	s_1	s_1	s_1	s_2	s_3	s_6	s_7
$f(i)$	×	×	1	1	1	0	0

r_i	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}
nearest point	s_7	s_8	s_9	s_{10}	s_{12}	s_{13}	s_{14}
$f(i)$	1	1	1	1	0	0	0

Figure 2: A toy example of following pattern. R (blue) follows S (red) in the time interval $[3:11]$.

We design a method in [8] to capture such dynamic relationship between moving objects. Supposedly we have two moving objects $R = r_1 r_2 ... r_n$ and $S = s_1 s_2 ... s_n$ with synchronized timestamps. Given one point r_i, the method first defines the *local minimizer* as a point s_j that is the closest point to r_i. Furthermore, s_j should be both spatially and temporally close to r_i to indicate interaction between R and S. Figure 2 shows a toy example, where R (blue) follows S (red) in the time interval $[3:11]$. The green line connects each r_i with its local minimizer. If we further have $j < i$ for r_i and its local minimizer s_j, we call such location pair a *following pattern*.

To find significant following time intervals, we first define an indicator function $f(\cdot)$ for each timestamp i that indicates whether r_i and its local minimizer s_j form a following pattern. Figure 2 shows the $f(\cdot)$ for the example. A significant following time interval should have *substantially more following patterns compared with the expectation*. Thus, we define the *following score* for an interval I as the *difference* between *actual* and *expected* number of following patterns.

If R and S are moving independently in an interval I, there is a 50% chance for a following pattern to occur at one timestamp. Therefore, the expected number of following patterns is half of the interval length $|I|$. Thus, we can define the following score $g(I) = f(I) - 0.5 * |I|$ for the interval I. In the running example (shown in Figure 2), the following score for interval $[3:11]$ is $g([3:11]) = 7 - 0.5 * 9 = 2.5$.

A significant following time interval should maximize the *following score*. The problem of finding following time interval is equivalent to the Maximum Sum Segment problem and all the following intervals can be found in linear time. In MoveMine 2.0, we implement both REMO method and our method.

4. ABOUT THE DEMONSTRATION

The MoveMine 2.0 adds recent relationship mining methods presented in Section 3. Our system connects to a large collection of real datasets from different resources. Figure 3 shows a screenshot of the prototype system. Each component is presented to users interactively and sequentially. Based on users' selection, the interface only displays relevant options in the subsequential steps.

Load dataset and preprocessing. A user can first download the dataset from data repositories or load the data

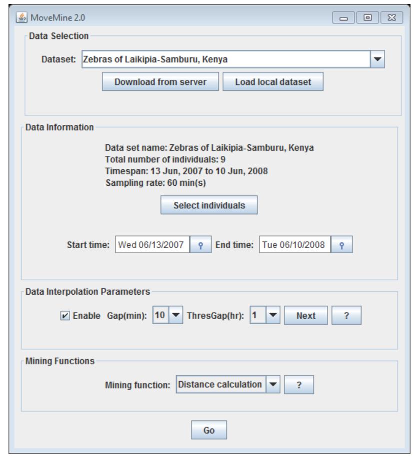


Figure 3: MoveMine 2.0 screenshot.

from local files. A description of the dataset will be shown upon selection. A user can also select specific individuals and a time period. Linear interpolation will then be performed to fill in the missing data with user-specified parameters. Before applying a method, a user can view the raw trajectory in Google Map and Google Earth. Figure 4(a) shows the density map of one trajectory on Google Map.

Detect attraction and avoidance relationship. We show a case of using MoveMine 2.0 to detect attraction and avoidance relationship on a capuchin monkey (cebus capucinus) dataset. The dataset contains trajectory of 12 capuchin monkeys with tracking time from 11/10/2004 to 04/18/2005. The average sampling rate for this dataset is about 15 minutes. The monkeys form six different groups.

To detect attraction and avoidance relationship pattern, a user can select corresponding method in the dropdown menu and specify parameters. The default parameter values are set to be "optimal" based on heuristics. The result can be visualized as a pairwise relationship matrix as shown in Figure 4(b). The columns and rows of the matrix are object IDs and each cell corresponds to the degree of attraction/avoidance relationship. In addition, the matrix can also show pairwise distances using the measures mentioned in Section 3.2. A user can also choose to visualize the relationship network. Figure 4(c) shows the pairwise attraction relationship network for the group of capuchin monkeys. The group information is shown for reference. In the network, a green line indicates significant attraction relationship and a red line indicates significant avoidance relationship. It is clear that monkeys in the same group all have significant attraction relationships. A user can further explore the data by plotting the trajectories in Google Earth. The co-locating places are also marked by pins on the Map. Figure 4(d) shows the home range of #83 (marked in yellow) and #52 (marked in blue) in Google Earth. Almost half of the home range for #83 overlaps with the home range of #52. However, they only met 8 times. According to a report from animal scientists [4], there have been 13 fights reported between FC and BLT group. This well explains the avoidance relationship between #52 and #83 detected by MoveMine 2.0.

Detecting following relationship. Next, we show a following example from two Baboons named A and B de-

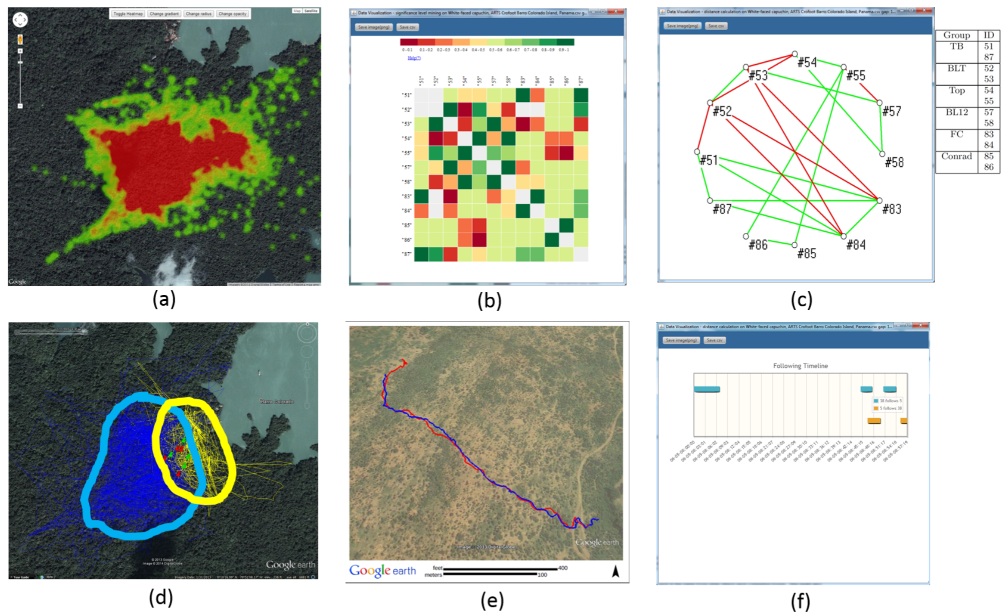


Figure 4: (a) Density map for one trajectory. (b) Pairwise relationship matrix. (c) Attraction/avoidance relationship network (capuchin monkeys). (d) Trajectories of #83 and #52 (capuchin monkeys). (e) A following pattern between Baboon A and B. (f) All following intervals between Baboon A and B.

tected by MoveMine 2.0. The Baboon movement datset comes from Movebank repository. The dataset contains GPS locations (longitude and latitude) of a group of 26 baboons tracked from August 1 to August 27, 2012 in Laikipia, Kenya. The sampling rate is about 1 second.

To detect the following relationship, a user can select corresponding method in the dropdown menu. Besides the text output, the result can be viewed in both pairwise relationship matrices and relationship networks. In addition, a user can view the animation in Google Earth. Figure 4(e) shows a following case between Baboon A (red) and B (blue), which happens between 8:00AM-9:00AM on August 3. The entire following relationship lasts about 9 minutes. Since the following patterns happen in multiple time intervals, it is useful to visualize the time information. Figure 4(f) shows the time line output, where a user can view all the following intervals during the selected time period.

5. SUMMARY

MoveMine 2.0 adds substantial new functions that focus on attraction and avoidance relationship and following pattern to our previous work. We design effective algorithms to address the challenges in detecting both relationship patterns [6, 8]. Our system enables relationship pattern analysis on real datasets from various resources. It also incorporates several visualization formats to facilitate interactive exploration of the result. MoveMine 2.0 provides useful tools for domain experts to gain insights for their problem. At the same time, one can evaluate newly developed methods by examining the performances on real datasets.

6. REFERENCES

[1] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica*, 2008.

[2] L. Chen and R. T. Ng. On the marriage of lp-norms and edit distance. In *Proc. 2004 Int. Conf. VLDB*.

[3] L. Chen, M. T. Özsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *Proc. 2005 ACM-SIGMOD ICDM*.

[4] M. Crofoot, I. Gilby, M. Wikelski, and R. Kays. Interaction location outweighs the competitive advantage of numerical superiority in cebus capucinus intergroup contests. *PNAS'08*.

[5] P. Laube, M. van Kreveld, and S. Imfeld. Finding REMO—detecting relative motion patterns in geospatial lifelines. In *Proc. 11th Int. Symp. SDH'04*.

[6] Z. Li, B. Ding, F. Wu, T. K. H. Lei, R. Kays, and M. C. Crofoot. Attraction and avoidance detection from movements. *Proc. 2014 VLDB Endowment*, 2013.

[7] Z. Li, M. Ji, J.-G. Lee, L.-A. Tang, Y. Yu, J. Han, and R. Kays. Movemine: mining moving object databases. In *Proc. 2010 ACM SIGMOD*.

[8] Z. Li, F. Wu, and M. C. Crofoot. Mining following relationships in movement data. In *Proc. 2013 ICDM*.

[9] M. Vlachos, D. Gunopulos, and G. Kollios. Discovering similar multidimensional trajectories. In *Proc. 2002 ICDE*.

[10] B.-K. Yi, H. V. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In *Proc. 1998 ICDE*.

[11] Y. Zheng and X. Zhou. *Computing with Spatial Trajectories*. Springer, 2011.