MoveMine: Mining Moving Object Databases *

Zhenhui Li[‡] Ming Ji[‡] Jae-Gil Lee[§] Lu-An Tang[‡] Yintao Yu[‡] Jiawei Han[‡] Roland Kays[†]

[‡] University of Illinois at Urbana-Champaign

[§] IBM Almaden Research Center

[†] New York State Museum

ABSTRACT

With the maturity of GPS, wireless, and Web technologies, increasing amounts of movement data collected from various moving objects, such as animals, vehicles, mobile devices, and climate radars, have become widely available. Analyzing such data has broad applications, *e.g.*, in ecological study, vehicle control, mobile communication management, and climatological forecast. However, few data mining tools are available for flexible and scalable analysis of massive-scale moving object data. Our system, MoveMine, is designed for sophisticated moving object data mining by integrating several attractive functions including moving object pattern mining and trajectory mining. We explore the state-of-the-art and novel techniques at implementation of the selected functions. A user-friendly interface is provided to facilitate interactive exploration of mining results and flexible tuning of the underlying methods. Since MoveMine is tested on multiple kinds of real data sets, it will benefit users to carry out versatile analysis on these kinds of data. At the same time, it will benefit researchers to realize the importance and limitations of current techniques as well as the potential future studies in moving object data mining.

Categories and Subject Descriptors

H.4.0 [**Information Systems**]: General

General Terms

Algorithms, Experimentation

*The work was supported in part by the Boeing company, NSF BDI-07-Movebank, NSF CCF-0905014, U.S. Air Force Office of Scientific Research MURI award FA9550-08-1-0265, and by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF- 09-2-0053 (NS-CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. E-mails: {zli28, mingji1, tang18, yintao, hanj}@uiuc.edu, leegj@us.ibm.com, rkays@mail.nysed.gov

SIGMOD'10, June 6–11, 2010, Indianapolis, Indiana, USA.

Keywords

Moving objects, pattern/trajectory mining, visualization

1. INTRODUCTION

The capabilities of collecting moving object data have been increasing rapidly with the fast development of satellite, RFID, sensor, wireless, and video technologies. The popular adoption of GPS technology promotes many new applications. Animal scientists attach telemetry equipment on the wildlife to analyze ecological behavior; mobility managers embed GPS in cars to better monitor and guide vehicles; and meteorologists use weather satellites and radars to observe hurricanes. Massive-scale moving object data are becoming rich, complex, and ubiquitous.

Despite the growing demands for diverse applications, there have been few scalable tools available for mining massive and sophisticated moving object data. Our system, MoveMine, integrates many data mining functions including moving object pattern mining and trajectory mining based on state-of-the-art methods. MoveMine has many application scenarios. For example, it can automatically detect an approximate period in movements; it can reveal collective movement patterns like flocks, followers, and swarms; and it can perform trajectory clustering, classification and outlier detection for geometric analysis of trajectories.

The development of MoveMine is motivated by several real application needs, including (i) the Movebank[1] project with a data repository containing hundreds of different animal movements; (ii) analysis of traffic data including the highway traffic data in the Bay Area and Chicago as well as taxi positioning in a city; and (iii) analysis of climate data such as hurricane tracks[2]. All the functions supported by MoveMine have been or will be tested on these and other available moving object data sets. At the same time, MoveMine provides a platform for users to flexibly tune parameters and supports visualization of the results in different formats. The output can be written in Google Map[3] and Google Earth[4] format to help users better explore the results.

2. GENERAL SYSTEM ARCHITECTURE

Figure 1 depicts the system architecture of MoveMine that consists of three layers: (i) collection and cleaning, (ii) mining, and (iii) visualization. The lower layer is responsible for

[1]http://www.movebank.org/

[2]http://weather.unisys.com/hurricane/atlantic

[3]http://code.google.com/apis/maps/

[4]http://earth.google.com/

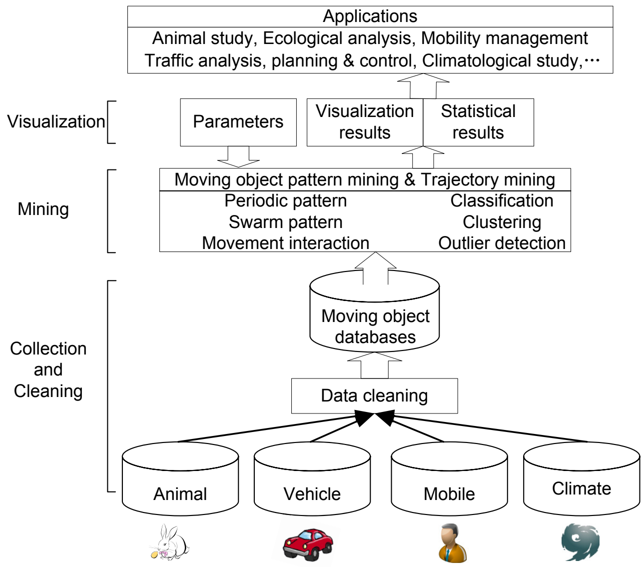


Figure 1: System Architecture

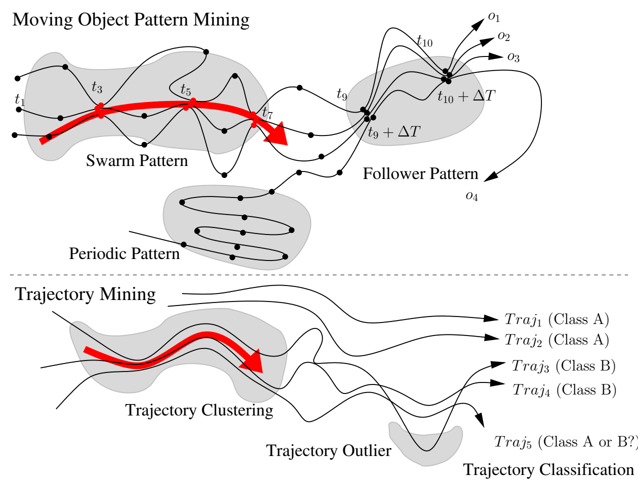


Figure 2: Major Functional Modules Overview

collection and cleaning of moving object data. Various moving object data sets are collected from different resources like animals, vehicles, mobile devices, and climate observations. Due to the limitations of technology, data could be inaccurate, inconsistent, and noisy. So preprocessing is needed to integrate and clean the raw data and to interpolate missing points. Mining is then performed on the preprocessed data sets stored in the moving object databases.

A rich set of data mining modules operate on top of the databases, enabling users to analyze data from different angles. The major functional modules we developed include periodic pattern mining, swarm pattern mining, movement interaction discovery, trajectory clustering, outlier detection, and classification. The details of these functions are described in Section 3.

The top layer shows the visualized results with some statistics. The visualized results can be plotted on 2D plane or embedded into other visualization tools (*e.g.*, Google Map and Google Earth). Along with the visualized results, some statistics, if possible, are presented to provide users with more insights into these results.

3. MAJOR FUNCTIONAL MODULES

Figure 2 illustrates an overview of the major functional modules implemented in the MoveMine system. We divide the functions into two categories based on the nature of the methods. The first category, *moving object pattern mining*, emphasizes the analysis of discrete locations with temporal information. For example, the swarm pattern, as shown in Figure 2, finds a group of objects that travel together in a sporadic way, meeting at certain timestamps, although their concrete trajectories could be rather different. The second, *trajectory mining*, focuses more on the mining of trajectories, associated with geometric shapes, such as clustering and finding outliers from hurricane paths across years. This section briefly introduces each functional module.

3.1 Moving Object Pattern Mining

We introduce several interesting moving object patterns in this section.

3.1.1 Periodic Pattern

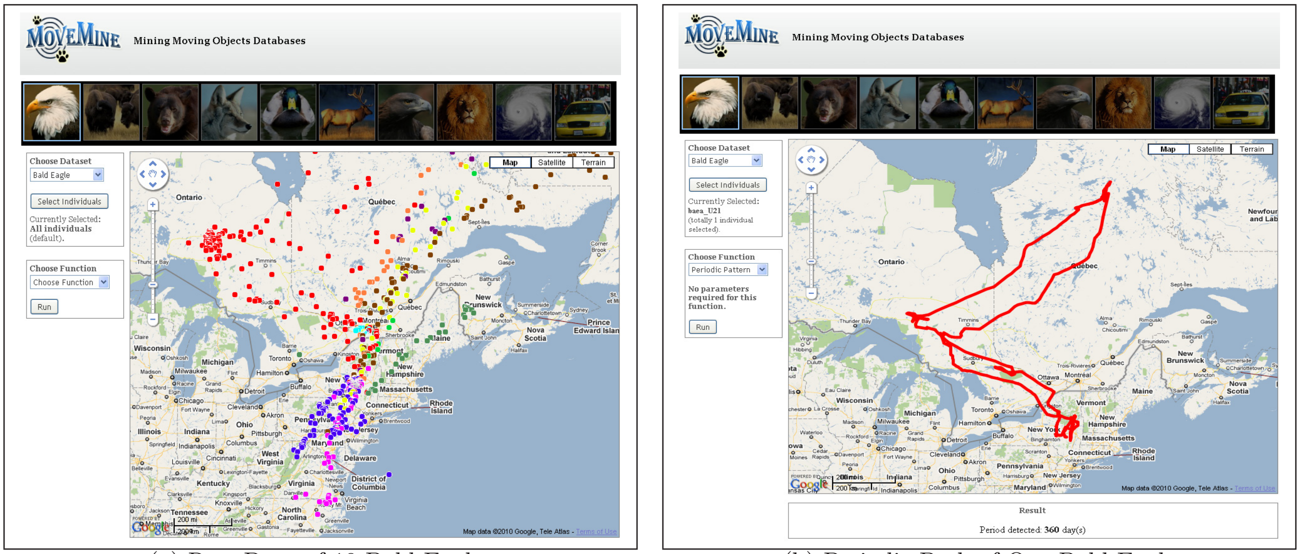
Periodicity naturally exists in moving object behavior. For example, birds have yearly migration patterns and people have daily routines between office and home. Periodic pattern mining can be used to discover intrinsic behavior of moving objects, and it is useful to predict future movements.

Detecting periodicity is a challenging task. Even though a moving object may follow a periodic pattern, it does not repeat the movement at the *exactly* same points (in terms of (x, y)) and on the *exactly* same time instance of a period. Approximation on both spatial and temporal dimensions is essential to mine the hidden periodic patterns. In previous work [1] on the discovery of periodic patterns for moving objects, the algorithm requires a period to be given in advance, which is usually not clear for users. We design a method [8] that can *automatically* detect the period by discovering and selecting several important locations to observe the movement and then applying Fourier Transform on the simplified movement sequences in those locations.

Let us examine a real example of bald eagles' movement. Figure 3(a) shows the raw trajectories of 10 bald eagles in 3 years. We further select a single eagle to analyze. Using the method in [8], we automatically discover that the period with the highest confidence is 360 days (as shown at the bottom in Figure 3(b)). And the plotted trajectory shown in Figure 3(b) is the interpolated path obtained by averaging the positions of the same day over 3 years.

3.1.2 Swarm Pattern

With a set of moving objects, people might be curious to know whether there exist groups of objects that move together. We call such groups *moving clusters*. There have been a lot of studies on moving clusters in terms of flock [3] and convoy [4]. However, they require the objects in a moving cluster to be together for at least k *consecutive* times. This might not be applicable in many real cases. We propose a new kind of moving clusters, called *swarm* [9], which relaxes the *consecutive time* constraint and allows an indi-



(a) Raw Data of 10 Bald Eagles

(b) Periodic Path of One Bald Eagle

Figure 3: Screen shots of the Bald Eagle Example

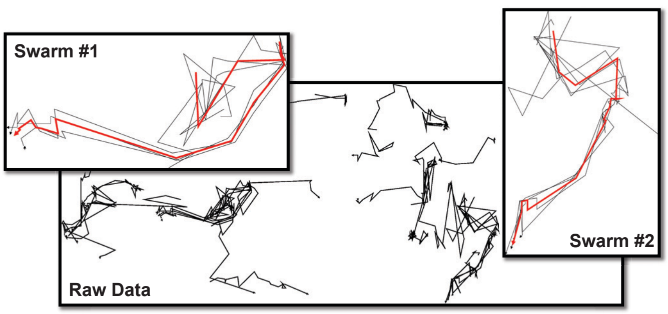


Figure 4: Swarm Pattern of Elks

vidual moving object to temporarily leave its group as long as it is close to other group members for *many* of the time.

Figure 4 shows an example of a real elk data set. Raw data, as plotted in black lines in background, contain 31 elks' movements in one day. Two swarm patterns are discovered, and the red lines in each swarm pattern connect the positions that the elks in a swarm are close to each other. One can see that the relaxation of the time constraint plays an important role in swarm discovery because the elks in a group do not strictly stick together all the time.

3.1.3 Movement Interactions

There are various moving object patterns that reflect interactions among movements. An interesting relationship among movements is the *follower* pattern. As depicted in Figure 2, a moving object set $S = \{o_1, o_2, o_3\}$ is a swarm pattern. Interestingly, a moving object o_4 arrives at similar locations as S, but there is a time lag ΔT. In real scenarios, it could be a group of wolves following a herd of sheep or a group of criminals following their target persons. Given a time lag ΔT or a time lag range $[\Delta T_1, \Delta T_2]$ provided by users or domain experts, MoveMine adapts the technique for mining swarms to discover the follower patterns. After shifting the movement of o_4 ahead by ΔT, o_4 actually joins swarm S.

Besides, Gudmundsson *et al.* [2] propose several moving object patterns based on the movement directions and locations, *e.g.* *flock*, *leadership*, and *convergence*. There are several subsequent articles studying the discovery of these patterns. MoveMine has also implemented and integrated those functions for mining such interesting patterns.

3.2 Trajectory Mining

In previous frameworks of trajectory mining, the whole trajectory is considered as an atomic unit. In contrast, in our framework [7, 5, 6], a trajectory is processed after it is partitioned into a set of line segments. Thus, our framework can take advantage of *partial* trajectories.

3.2.1 Trajectory Clustering

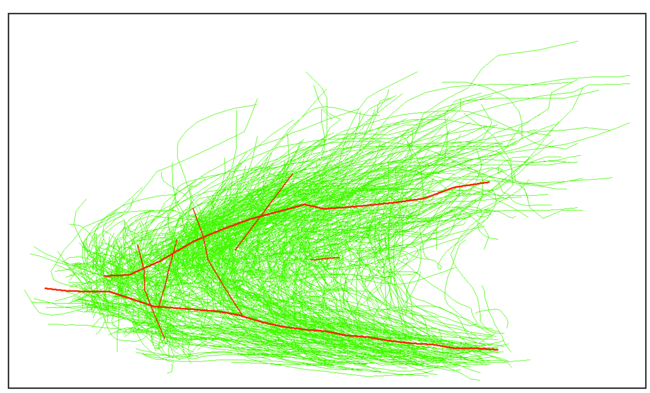


Figure 5: Trajectory Clustering

The goal of trajectory clustering is to find similar movement traces. Many clustering methods haven been proposed using different distance measures between trajectories. While most of those studies cluster trajectories *as a* *whole*, our method [7] can discover similar *portions* of sub-trajectories. Note that a trajectory may have a long and complicated path. Hence, even though two trajectories are similar in some sub-trajectories, they may not be similar as a whole. Discovery of common sub-trajectories is useful, especially if one considers regions of special interest in analysis.

This observation leads to the development of a sub-trajectory clustering algorithm: *TRACLUS* [7]. The method discovers clusters by grouping *sub*-trajectories based on density. Each trajectory is first partitioned into line segments using Minimum Description Length (MDL). Then, density-based clustering is applied on the segments. Finally, a representative sub-trajectory is summarized over all the line segments in the cluster. Figure 5 shows a clustering of hurricane data.

3.2.2 Trajectory Outlier Detection

Figure 6: Trajectory Outlier Detection

Trajectory outliers are the movements that do not comply with the general behavior. The general behavior is usually obvious, however, users are often interested in abnormal ones. Trajectory outliers could be informative signals of environment changes or unexpected accidents.

In general, there are two kinds of trajectory outliers. For the first kind, a moving object could be an outlier in comparison with its peers since it may not follow the similar paths or not move together with other moving objects. The second kind is called an *outlying sub-trajectory*, which is a *portion* of a trajectory that does not follow the general trend of other (sub-)trajectories. The first one can be detected using various methods once the distance measure between two moving objects is fixed. The second case is more challenging, and a partition-and-detect framework is proposed [5] and implemented in our system. The red lines in Figure 6 show anomalous *portions* of hurricane trajectories in 10 years (1995 $\sim$ 2004).

3.2.3 Trajectory Classification

Trajectory classification is the process of predicting the class labels of moving objects based on their trajectory-related features. For example, by labeling a set of trajectories, one may depict the normal vehicle trajectories and predict suspicious vehicle trajectories that could be caused by illegal immigration or drug smuggling.

MoveMine implements the classification function using the solution provided in our research [6]. Two types of features, *region-based* and *trajectory-based* features, are used in our method. Region-based feature extraction cover the regions having trajectories mostly of one class regardless of their movements. Trajectory-based feature extraction discovers the sub-trajectories that indicate common movement patterns of each class.

4. ABOUT THE DEMONSTRATION

The MoveMine system integrates data mining functions presented in Section 3. The methods embedded in the system are novel, practical, and derived from recent research. To demonstrate its effectiveness, a large collection of various real data sets from different resources are used. Moreover, we also communicate with some data providers to ensure the system meets many real-case requirements.

MoveMine provides a user-friendly interface. Figure 3 is a preliminary screen shot of our system. Users can select a data set and the corresponding raw data is plotted on the Google Map (as shown in Figure 3(a)). Since some functions may focus on individual moving object (such as periodic pattern) whereas others may use all or a subset of moving objects in the data set (such as clustering), a user can further select particular sets of moving objects of interest. After the data set and moving objects in this data set are selected, a user can choose the function to look into the data. Parameters for the selected function will be shown correspondingly. The default parameter values are set to the "optimal" ones derived by our heuristics. To better browse the results, outputs returned are visually displayed. Similarly, the results will embedded in Google Map (as shown in Figure 3(b)) and a user can zoom in/out or drag the map. Furthermore, a user can plot the results in Google Earth for 3-D visualization of the results.

By experimenting with real data sets, one can observe interesting functions and applications, and at the same time, find the limitations of the current methods. This will promote further research into the new challenge issues in moving object mining.

5. REFERENCES

[1] H. Cao, N. Mamoulis, and D. W. Cheung. Discovery of periodic patterns in spatiotemporal sequences. In *TKDE'07*.

[2] J. Gudmundsson, P. Laube, and T. Wolle. Movement patterns in spatio-temporal data. In *Encyclopedia of GIS 2008*.

[3] J. Gudmundsson and M. van Kreveld. Computing longest duration flocks in trajectory data. In *GIS'06*.

[4] H. Jeung, H. T. Shen, and X. Zhou. Convoy queries in spatio-temporal databases. In *ICDE'08*.

[5] J.-G. Lee, J. Han, and X. Li. Trajectory outlier detection: A partition-and-detect framework. In *ICDE'08*.

[6] J.-G. Lee, J. Han, X. Li, and H. Gonzalez. Traclass: Trajectory classification using hierarchical region-based and trajectory-based clustering. In *VLDB'08*.

[7] J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: A partition-and-group framework. In *SIGMOD'07*.

[8] Z. Li, B. Ding, J. Han, and R. Kays. Mining hidden periodic behaviors for moving objects. In *Submission*.

[9] Z. Li, B. Ding, J. Han, and R. Kays. Swarm: Mining relaxed temporal moving object clusters. In *Submission*.