ISSN 1232-3829

tts 10/2013

Technika Transportu Szynowego koleje ▪ tramwaje ▪ metro

CZEJDO Bogdan, BHATTACHARYA Sambit, MALHOTRA Rakesh

INTERACTIONS OF DYNAMIC GEOSPATIAL OBJECTS WITH STATIC LANDMARKS

Abstract

Analyzing the behavior of moving objects has multitude of applications e.g. in the area of transportation. Each application might require identification of different behavior patterns and their relationships to different landmarks. Machine learning algorithms can help in automatic recognition of spatiotemporal patterns. However this is still a largely unsolved problem, especially identification of the relationships of moving point objects with stationary objects or landmarks on a map. In our project we considered dynamic objects such as cars and humans on a terrain with static elements such as road networks and buildings e.g. airports, bus stops etc. We created application specific ontologies of patterns of moving objects in relation to static landmarks. Based on ontologies we built machine learning models to classify trajectories of moving objects.

INTRODUCTION

Analyzing the behavior of moving objects has multitude of applications in the area of transportation, safety and security, retail marketing, sports, natural disaster, mobile computing, and sensor networks. Each application might require identification of different behavior patterns and their relationships to different landmarks. These behavior patterns carry special meaning for the target application. Tracking the movement of dynamic objects in different areas is important to understand the higher level patterns of movement and their relationship with patterns for the specific application.

In general moving objects can be grouped into two categories. The first group moves in geographic space such as humans, animals, or vehicles. The second group moves in non-geographic space such as mouse movement, eye movement or particles in a bubble chamber [1]. A good example of a complex system to support the first group is Maritime Safety and Security (MSS) system to monitor vessel traffic in [4]. The system includes abstraction and simulation of trajectory sensor data, fusion of multiple heterogeneous data sources, reasoning, and visual analysis of the combined data sources.

Recent advances in sensor technology and computer software have made it easier to capitalize on automated, real time tracking of moving objects. This can help improve human performance, provide continuous authentication, and monitor specific areas. Current developments point to a future where an Internet of Things will enable mobility data collection of various devices [3]. This research develops solutions that are applicable to the analysis of data collected in traditional geospatial environments.

Machine learning algorithms for automatic recognition of spatiotemporal patterns are needed in geospatial science due to high volumes of data that are impossible to process only through standard statistical and visualization approaches. However this is still a largely unsolved problem, especially identification of the relationships of moving point objects with stationary objects or landmarks on a map.

In geospatial science spatiotemporal data query and updates are implemented in databases storing moving objects. The collection, visualization, and analysis of movement data are active research areas [1], however the problem of machine learning and automatic recognition of spatiotemporal patterns is largely an unsolved problem. Data on a moving object is collected by recording its spatial location at discrete time intervals in the form of a sequence of coordinate values. While object shape can be important consideration for certain applications, taking shape into account creates problems that are harder to solve. Also the combination of space and time attributes, representation of time and designing statistical tests on spatiotemporal data are still challenging problems. Seven classes of methods have been identified by Long and Nelson [1]: (1) time geography, (2) path descriptors, (3) similarity indices, (4) pattern and cluster methods, (5) individual and group dynamics, (6) spatial field methods, and (7) spatial range methods. Some areas of future research identified by [1] include measuring interactions between moving objects, developing predictive frameworks for movement data, integrating movement data with existing geographic layers, and incorporating theory from time geography into movement models.

This project addresses some of these important open problems using model of moving point objects (MPOs). The goal of this project was to create a system for (a) storing ontology for patterns of geospatial objects movement and landmarks (b) visualizing and cleaning data for moving objects, (c) generating simulated data for moving objects where available data are insufficient for pattern discovery, (d) segmenting MPOs trajectories, (e) recognizing the type of interaction of MPOs with stationary landmarks on a map, and (f) discovering new types of interactions of MPOs with stationary landmarks on a map. Different approaches were developed to address these specific tasks.

1. ONTOLOGY FOR PATTERNS OF GEOSPATIAL OBJECTS MOVEMENT AND LANDMARKS

In the standard taxonomies of geospatial movement patterns [2], the patterns are broadly classified as either generic or behavioral. Generic patterns such as repetitive movements, movements that lead to encounter/break-up are used to describe a wide range of dynamic object types. Behavioral patterns are those that typically occur over larger space-time scales, being created with generic behavior patterns as building blocks. In this project we concentrate on generic patterns.

Consider Figure 1 which illustrates early and late stages of the *meet* and the varying *meet* patterns. Many interesting movements' aspects can be observed through these examples. First, in the *meet* pattern, many objects are meeting roughly at the same time. However in *varying meet* pattern the objects may not intersect in time but they do intersect in space.

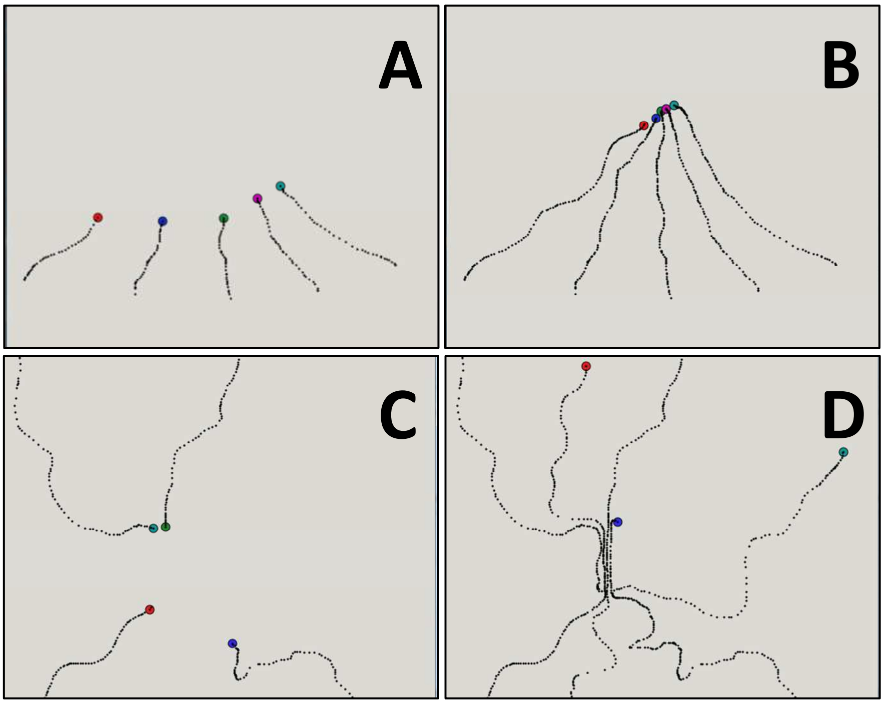

Fig. 1. (A) Moving point objects engaged in the *meet* pattern (early stage) (B) Same moving point objects at a later time (C) Moving point objects engaged in the *varying meet* pattern (early stage) (D) Same moving point objects at a later time.

The study of moving point objects (MPOs) is a continuously evolving research area but there is a persistent need to properly specify taxonomy or ontology of MPOs [1]. A formal model for representing point trajectories in two-dimensional spaces was used in [7].

In our project we consider movement of dynamic objects such as cars and humans on a terrain with static elements such as road networks and buildings e.g. airports and bus stops. These requirements lead to creation of an ontology that consists of *Objects* and *Landmarks* classes, *Movement Patterns*, and relationships between them as shown in Figure 2.

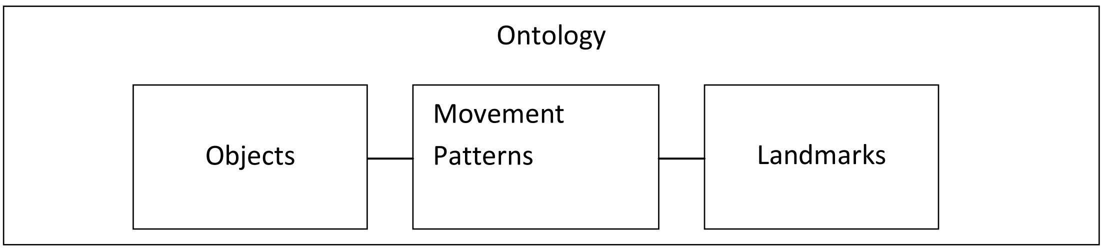

Fig. 2. Ontology for patterns of geospatial objects movement and landmarks.

Even though in our project we did not consider the shape of moving objects, we still allowed for MPOs properties. The *Objects* class was therefore used to model object properties. The *Landmarks* class is used to model various landmarks' properties including their static location. MPO behavior patterns are based on objects space and time data, which are used to derive meaningful movement information. There could be different meaningful movement patterns for different objects and for different landmarks.

For example, stored in our ontology “meet” pattern [1] could be used to check if a meeting has taken place between two suspects in a security application domain. More specificaly we can check if the movements to be within a certain threshold of spatial coordinates with corresponding time stamps that do not show any changes in the coordinates. Similarly, we can check spatial coordinates for the “lagged co-incidence in space and time” pattern [1] which may be interpreted as two MPOs where one is following another.

In our ontology system the movement patterns can be related to landmarks. For example “meet” patterns occurring in the neighborhood of an open gas station are likely to be unintentional whereas similar pattern near a closed and infrequently visited location maybe considered security concerns. There are numerous situations involving unintentional versus intentional patterns for other patterns and most of them can be reflected by our ontology.

2. VISUALIZATION AND CLEANING DATA FOR MOVING POINT OBJECTS AND LANDMARKS

A generalized Moving Object Model can specify three main data types of a moving object; moving points, moving lines and moving regions and a set of operations over them [5]. In this project we concentrated on data processing of the moving point data type.

In our experiments, however, we also use data from shape files to compute and visualize the road network from the underlying data. We used Python Script Tool within an open instance of ArcMap to accomplish this task [14]. Figure 3a shows the visualization of road network in Cumberland County in North Carolina.

Fig. 3a. Road Network in Cumberland County in North Carolina

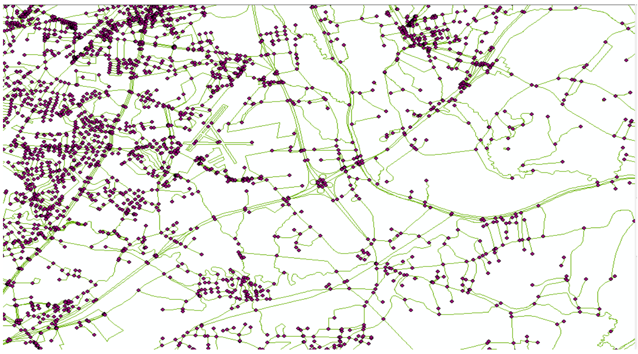

Fig. 3b. Zoomed Road Network in Cumberland County in North Carolina

One of the problems that needed to be solved was the fact that the geometric intersections of the polylines may not correspond to the physical intersections. That is shown better on the zoomed version of the visualization in Figure 3b.

Once we got underlying road network, we could visualize MPO data as an extra layer. Let us consider real GPS data collected from taxi drivers in San Francisco area. We visualized this data by displaying all taxi routes for the selected taxi and for the specified time as shown in Figure 4a. This initial visualization is already showing some interesting patterns for taxi routes e.g. most frequent places for taxi to drive.

Fig. 4a and 4b. Visualization of taxi drivers routes in the San Francisco area

For more precise analysis, however, data needs to be cleaned and more carefully interpreted. Figure 4b and 4c show problems with such simple visualization of data. One of the problems is that the taxi routs seem to be crossing the water. The main reason was that simple visualization does not take into account that the presented GPS data are discrete and with varying time interval. Another reason, however, was that data contained some errors.

Fig. 4c and 4d. Visualization of taxi drivers routes in the San Francisco area

When we apply selection operation limiting number of the routs to those with relatively short time interval as shown in Figure 4c we can clearly see the erroneous data identifying one of the taxi positions in the middle of the bay. We needed, therefore to apply data cleaning process. Figure 4d shows additional problems with such simple visualization of data. Some behavioral pattern can be discovered only on a proper aggregation level.

The typical format of GPS trace data we use is arranged in x, y (location), t (time), and o (occupancy) columns. The visualization above shows the trajectories without incorporating occupancy information. With occupancy data we can visually analyze different characteristics of landmarks such popularity and time of use. For example with occupancy information we can extract sub-trajectories which carried a passenger on route and further use the end GPS coordinate (drop-off location) to do a frequency count of landmarks which were most visited. Also we can use the time information to find out which locations get visited most frequently at specific times of the day.

3. GENERATION OF SPATIAL DATA FOR MOVING POINT OBJECTS AND LANDMARKS

Real data are very important, but it is almost impossible to have data covering all movement patterns and their relationships with landmarks. Therefore, in addition to real data there is a need to generate data covering as much as possible the missing aspects. As a consequence we generated various spatial movements of point objects that can be classified into different known patterns. Spatial movements were generated based on an ontology that includes the landmarks and descriptive MPOs behavior attributes. The generated set of agent movements can be modified semi-automatically based on changes in ontologies. One challenging aspect of this goal is to define an expandable ontology that can be easily adjusted not only to changes in landmark configurations but also for changes in behavior attributes of MPO. Another aspect of this goal is to automatically generate the beginning and ending time-points of agent behavior segments, where each segment may have motions most relevant to the ontologies.

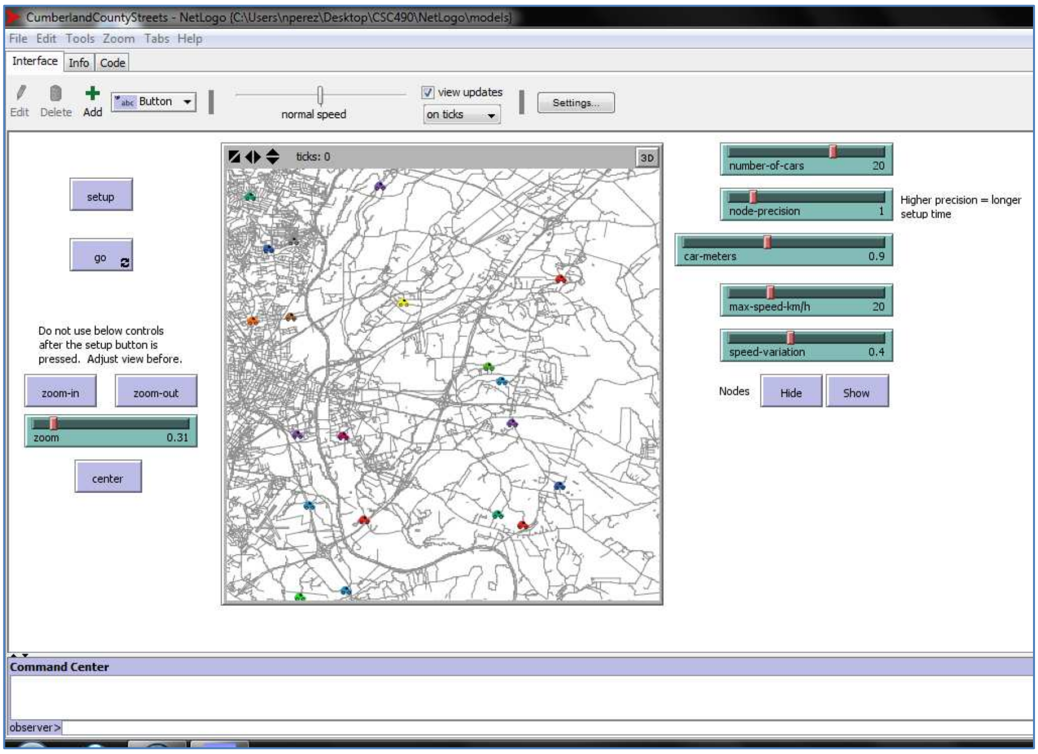

Fig. 5. Simulated MPO movement on a road network in Cumberland County, North Carolina.

Figure 5 shows an example of simulated test data using NetLogo [13] with GIS extension. Geospatial agents are coded with NetLogo to simulate MPO patterns on a road network. Each artificially intelligent NetLogo agent has detailed instructions on patterns of movement and usage of static landmarks on a real road network.

4. TRAJECTORY SEGMENTATION AND CLASSIFICATION

We are using a probabilistic framework based on machine learning algorithms to automatically identify and discover pattern movements of MPOs. More precisely we are creating algorithms to compute principal sub-trajectory patterns from training data. In order to accomplish that, an input trajectory has to be split into sub-trajectories. The problem of detecting where a sub-trajectory begins and ends in time is solved through probability estimates of classifications produced by available classification algorithms.

We considered many machine learning tools like Artificial Neural Networks (ANN), K Nearest Neighbors (kNN), Decision Tree (DT) learning, and Support Vector Machines (SVM) [9, 10, 11]. SVM is our primary machine learning tool.

Let us discuss the application of SVM in our experiments. Given a simple case of training data from 2 movement patterns the question we want to answer is what is the best way to draw a line that leaves the maximum gap on either side from the training data. This best line gives the best prediction accuracies on previously “unseen” data which were not part of the training set, namely test data. The straight line does not always provide the best separation therefore a non-linear seperation is used more often as shown in Figure 6.

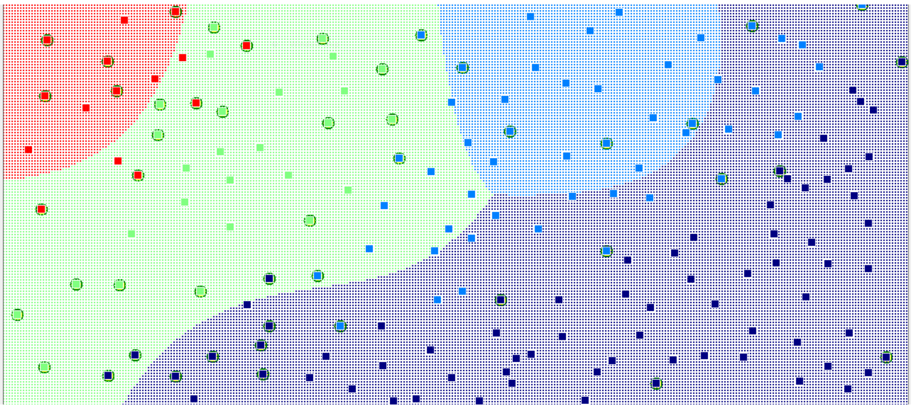

Fig. 6. Separation of multi-class data using a non-linear approach.

We can be confident about our trained machine learning algorithm if it achieves high recognition accuracy on “unseen” data, that is data that have not been used for training. Cross-validation is the standard way to address the problem of confidence of machine learning method. In cross validation we usually randomly split the available data for training & testing (e.g. 80%-20%) and then repeat this many times (k-times) all the while recording the accuracy of prediction/recognition.

Data segmentation is a significant challenge when we want automatically to classify patterns that are sub-sequences of larger sequences. The problem can be restated as automatic detection of change points in trajectories where a sub-trajectory of a certain type begins and also the point where it ends in time. In our research we addressed this problem of change-point detection by using probability estimations done by machine learning algorithms [8, 12] as shown in Figure 7a and b.

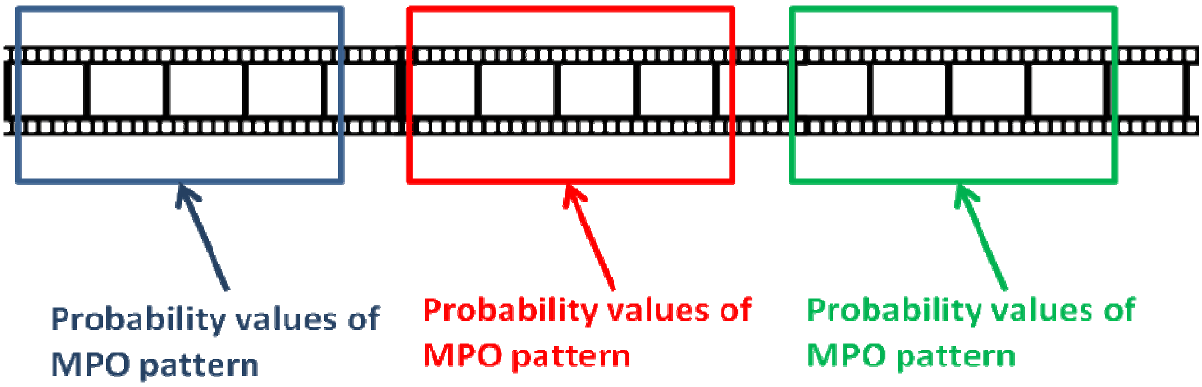

Fig. 7a. Conceptual diagram of MPO movement patterns segmented from a continuous data stream using probability estimates.

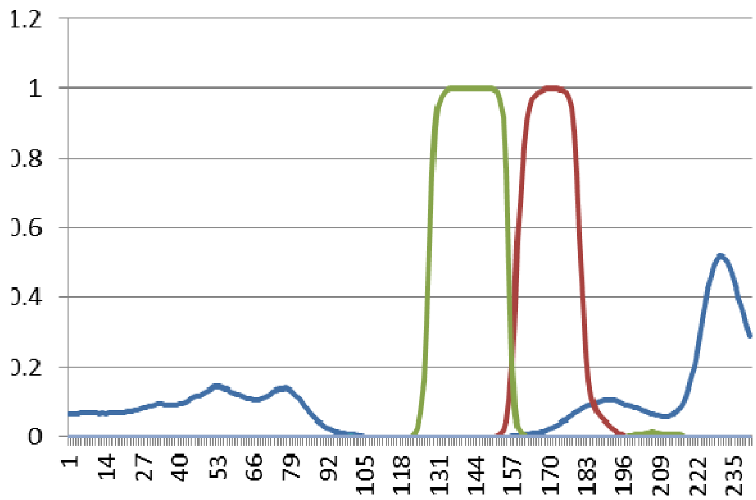

Fig. 7b. Probability estimates for a sequence of patterns taking place over time.

We trained classification algorithms to recognize sub-trajectory patterns in whole trajectories using a sliding window approach as shown in Figure 7a. Probabilistic variants of these algorithms were used to compute probability estimates of class-membership so that change points can be detected with high accuracy [15]. Probability estimation [4] along with a hard coded rule based approach is highly successful at the data segmentation tasks in practical applications for which high amounts of training data are available.

Another aspect of our work is anomaly detection on moving objects using the one-class classification approach. We apply classification algorithms to predict the states that moving objects go through in time when approaching landmarks or other moving objects. Geospatial anomalies can arise when a moving object behaves in unexpected ways in the context of nearby landmarks. We apply this probabilistic framework for anomaly detection using a one-class SVM classification algorithm.

5. SYSTEM ARCHITECTURE

In our project we created a system to process or generate trajectory data, and classify it according to defined ontology as shown in Figure 8. General and application specific entries were made in our ontology for MPOs interacting with stationary objects on maps or landmarks. We designed parameters for quantitative description of behaviors of objects in relation to landmarks. Methods of measuring indicators of spatiotemporal behaviors from raw data were created by encoding semantic notions of an object visiting a certain landmark and using it in a certain way.

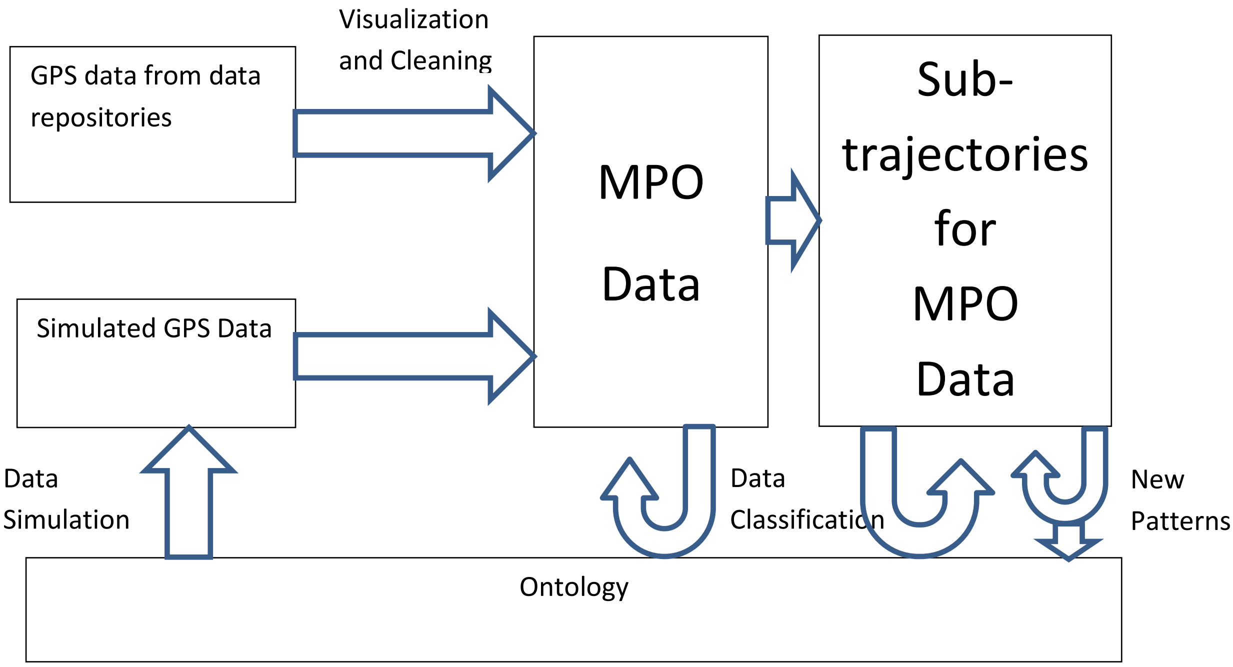

Fig. 8. System architecture.

For trajectory classifications we used machine learning algorithms, mainly SVM. For more complex classifications and new pattern discoveries trajectory data were decomposed into sub-trajectories which were homogenous under certain criteria. Simple examples of sub-trajectories are subsequence of the data which show no motion, constant velocity etc. These sub-trajectories were contextualized with respect to common patterns occurring near landmarks over historical data.

SUMMARY AND FUTURE WORK

We are engaged in a two-phase approach. In the first phase, described in this presentation, we concentrated on addressing geospatial dynamic data processing in environments where GPS data is available or can be simulated. The first phase efforts we provided some solutions for creating a consistent geospatial data ontology and use this ontology for automatic trajectory classification. In the second phase, we will address the issues related to the collection of data from social networks on the Internet. That will involve appropriate extraction from the textual files the movement data and landmark data. Such an approach will allow integrating the GPS data and geospatial data from the social media to provide timely solutions for geospatial data interpretation.

The cleaning and integrating data extracted from social networks will be supported by several computational methods. Part of the process of data correction is to determine if an identifier refers to the same geo-located entities or relationships. In integrating new data with the existing data sets it is often necessary to solve this problem by exploiting distance in a multidimensional space between the textual names used for objects and relationships and names existing in the structured data set. We will study techniques for clustering and matching identifier names for both entities and relationships that are used in geospatial context.

ACKNOWLEDGEMENTS

We acknowlege programming work done by Nicolas Perez. We acknowledge ISAS grant from National Science Foundation for supporting this project. We also acknowledge Belk Foundation for supporting this project.

BIBLIOGRAPHY

1. Long, J. A. and T. A. Nelson (2012). "A review of quantitative methods for movement data." International Journal of Geographical Information Science: 1-27.
2. Dodge, S., R. Weibel, et al. (2008). "Towards a taxonomy of movement patterns." Information visualization 7(3-4): 240-252.
3. Percivall, G. (2012). Connecting islands in the internet of things. Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications. Washington, D.C., ACM: 1-1.
4. Platt, J. (1999). "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods." Advances in large margin classifiers 10(3): 61-74.
5. Vapnik, V. (1999). The nature of statistical learning theory, Springer.
6. Joachims, T. (1999). "Making large scale SVM learning practical."
7. Johnson, N. and D. Hogg (1996). "Learning the distribution of object trajectories for event recognition." Image and Vision Computing 14(8): 609-615.
8. Weinland, D., R. Ronfard, et al. (2011). "A survey of vision-based methods for action representation, segmentation and recognition." Computer Vision and Image Understanding 115(2): 224-241.
9. Chih-Wei Hsu, C.-C. C., and Chih-Jen Lin. "A Practical Guide to Support Vector Classification." Retrieved June 29, 2012, from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.
10. C.-C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.
11. Ben-Hur, A. and J. Weston (2010). "A User’s Guide to Support Vector Machines Data Mining Techniques for the Life Sciences." 609: 223-239.
12. T. F. Wu, C. J. L., and R. C. Weng. (2004). "Probability estimates for multi-class classification by pairwise coupling." Journal of Machine Learning Research 5: 975-1005.
13. NetLogo URL. http://ccl.northwestern.edu/netlogo/
14. ArcGIS URL. http://www.esri.com/software/arcgis
15. Murphy, K. P. (2012). "Machine Learning: a Probabilistic Perspective."

Authors:

CZEJDO Bogdan - Department of Mathematics and Computer Science, Fayetteville State University, Fayetteville, NC, USA

BHATTACHARYA Sambit - Department of Mathematics and Computer Science, Fayetteville State University, NC, USA

MALHOTRA Rakesh - Department of Government and History, Fayetteville State University, NC, USA