Using Ontology Based Knowledge Discovery in Location Based Services

Ali Mousavi[1], Andrew Hunter[1]

[1] Department of Geomatics, Schulich School of Engineering, University of Calgary,
{mousavia|ahunter}@ucalgary.ca

ABSTRACT
Rapid development of information technology for mobile computing, improvement in the accuracy of positioning systems, and ubiquitous use of mobile devices has generated large quantities of raw trajectories that represent the movement of moving objects. Mining such data, which contain not only space and time attributes, but also context attributes, is significant for many applications within the location based services domain. Such systems can provide more effective services to users through an understanding of a moving objects location, its context, and interests. But, in spite of the fact that most service providers offer various services to users, they are unable to identify relevant customers at the right time and right place. Recently different methods of data mining have been used in location based services for extracting patterns and modeling the behavior of users. However, due to the excessive number of extracted patterns it has been very difficult to infer knowledge from these patterns for an application domain. Given that several criteria such as geographic area, the nature of the entity, etc., influence movement behavior, one needs to consider this complexly during the knowledge discovery process. Therefore, this paper proposes a model that integrates an ontology-based approach for efficient interpretation of extracted patterns from an objects movement behavior.

KEYWORDS:
Ontology, data mining, location-based services, Spatio-temporal GIS

Introduction

The development of location technologies in mobile devices and wireless communication have led to deploying variety of Internet–based services such as Location Based Services (LBSs) (Steiniger, Neun, and Edwardes 2006). The application domain for these types of services are typically transportation management, location–aware advertising, tourism and so on (Jensen 2002). The widespread use of these services in our daily activities has led to huge amounts of positioning data that can be represented as trajectories (Ong et al. 2010). Generally, these data take the form of an x, y, t triplet that represent the spatial coordinates and time stamp of a location (Bogorny et al. 2011). In spite of the fact that most service providers offer various services to users, they need to identify relevant customers at the right time and right place. In other words, they need to know where, when, and which services to provide them. Therefore, most of systems designed to support LBSs need users to key in additional relevant information and select their desired services according to needs and location. In addition, current LBSs generally just provide information and services based on a users' context and current location (Chen and Kotz 2000). However, sometimes it could be beneficial to provide proactive services to users that consider their movement patterns and behaviors. Therefore, it seems that suitable modeling of behavior patterns of users is necessary to do this.

To customize these services, an efficient analysis of LBS data from across different application domains is required to identify similar behaviour or discover regularities between users that can be used to predict user's future behaviour (Nanni et al. 2008). Using knowledge discovery methods, LBS can provide a variety of patterns that describe the mobility of people and goods, and could be used to answer questions such as "What will be the next destination of a customer?" or "Given the present location of customer x, what type of information might they want to know?" (Quintas, Costa, and Ribeiro 2003). Hence, it is beneficial to not only understand movement, but also behavior, so that useful knowledge for LBS can be generated.

Over the past few years, research has investigated new analytical techniques and computational methods for the analysis of movement data (Dodge 2011). Different challenges arise when developing new exploratory tools: how should one discover similar trajectories (Lee et al. 2008), periodic movement (Cao, Mamoulis, and Cheung 2007), classify trajectories (Lee et al. 2008), or identify relative motion patterns (Laube, Imfeld, and Weibel 2005). A large number of studies have established approaches to utilize movement data in various aspects of Geographic Knowledge Discovery (GKD), such as trajectory data analysis, movement pattern mining and exploratory visual analytics (Imfeld 2000; Laube 2005; Giannotti, and Pedreschi 2008). However, these algorithms have mainly focused on the geometric properties of trajectories and very little attention has been dedicated to any other parameters. As such, the techniques are good at discovering patterns, but the patterns can be difficult to interpret (Alvares et al. 2007). Furthermore, mined results can be more meaningful when the nature of the movement data is considered as context within the mining process (Ong et al. 2010). According to Dodge, Weibel & Lautenschütz (2008) movement behavior depends on the context of the movement: where movement happens, why movement is occurring, what time of day, what day of the week, etc. Thus, analysis methods for interpreting the semantics of context within the knowledge discovery process can lead to discovery of semantic trajectory patterns (Ong et al. 2010). Semantic technology is an intellectual technology which can improve interaction between system and information (Niaraki and Kim 2009). Using ontologies it is possible to support decision-making through improved communication between user and system, and provide intelligent and flexible services, which are capable of recommending the most suitable services to a user.

Therefore, this research proposes an ontology-based semantic knowledge discovery model, not only to improve preprocessing steps, but also for interpreting discovered patterns during post processing. To this end, we propose a five step process: (i) ontology creation, (ii) data and knowledge acquisition, (iii) data preprocessing, (iv) semantic knowledge discovery, and (v) interpretation. In this research, different ontologies focusing on geometry, geography, themes, and service are considered. Furthermore, a system is proposed as a prototype to evaluate the model that consists of several essential components: a database server; ontology construction; data mining modules; a reasoning engine; LBS server; and user interface.

The reminder of the paper is organized as follows: section 2 presents related work from current research, section 3 describes the proposed methodology and illustrates each component, section 4 introduces a client server architecture, section 5 introduces system components for implementing the architecture, and finally, section 6 concludes the paper.

Related works

Knowledge Discovery for Movement Data

Recently, the database community has focused on the definition of spatiotemporal data types. Guting et al (2005) have paid significant attention to moving object databases. They proposed different data types and operations for modeling and querying moving objects in road networks. Pelekis et al (2006) developed the HERMES prototype to exploit the spatial data types provided by the Oracle database management system. Generally, most research in this field has focused on the geometric and temporal characteristics of trajectories. Several data mining (DM) methods for trajectories have been proposed (Tsoukatos and Gunopulos 2001; Li, Han, and Yang 2004; Nanni and Pedreschi 2006) for discovering similar trajectories or dense regions. Generally, most works adopting classical DM methods have focused on the mining step itself, and ignored the whole knowledge discovery process, which includes data preprocessing, data transformation, DM, and post processing (Bogorny et al. 2011). This may be because the extraction of many patterns can make pattern interpretation very difficult.

Recently, the concept of a semantic trajectory as a sequence of stops and moves (Spaccapietra et al. 2008) was developed. Wherein a stop is defined as an interesting place in which some moving entities have stopped for some temporal duration, while a move is defined as the part of the entities' movement between consecutive stops. Several researchers have adopted this approach as a standard for semantic trajectory data analysis. But only a few such as Vania et al (2011), Baglioni et al (2009) and Trasarti et al (2010) have considered semantic information as a means to help understand trajectory patterns. Little attention has been given to the preprocessing phase that can aid semantic interpretation. Therefore, we posit that extending a typical DM framework by considering movement ontologies allows the knowledge discovery process to be performed more effectively when working with massive geospatial databases that contain highly complex relationships representing real world processes, environments, and phenomena (Usery, Azami, and Kwan 2004).

Ontology and semantic knowledge discovery

Ontology plays an important role in the construction of GIS by establishing correspondences and interrelations among different domains of spatial entities and relations (Smith and Mark 1998). Recently, several works have demonstrated the effectiveness of using ontologies for supporting the knowledge discovery process (Charest and Delisle 2006). Bernstein et al. (2002) proposed an intelligent DM assistant based on the use of an ontology in order to ranking likely DM processes. Phillips and Buchanan (2001) have used ontologies to conduct a feature selection step in their knowledge discovery process. Bauer and Baldes (2005) used an ontology based interface to help non-expert users understand a machine learning system from a semantic perspective. Canataro and Camito (2003) demonstrated the use of a DM ontology in the area of grid computing to simplify distributed knowledge discovery applications. However, none of these works have considered an ontology model that encompasses spatial, temporal, and other dimensions of data concurrently. They mostly focused on data dictionaries, data interoperability, and data presentation. Such ontologies enable the discovery of complex relations among entities and enables meaningful interpretation of multimodal information across different domains as they relate geospatially. Therefore, the main objective of this research is to show the important role that ontologies can play in the knowledge discovery process.

Methodology

Figure 1.1 outlines the overall proposed methodology consisting of five different steps: (i) ontologies; (ii) data and knowledge repository; (iii) data preprocessing; (iv) semantic knowledge discovery; and (v) interpretation. Each of the steps is explained briefly in the following sections.

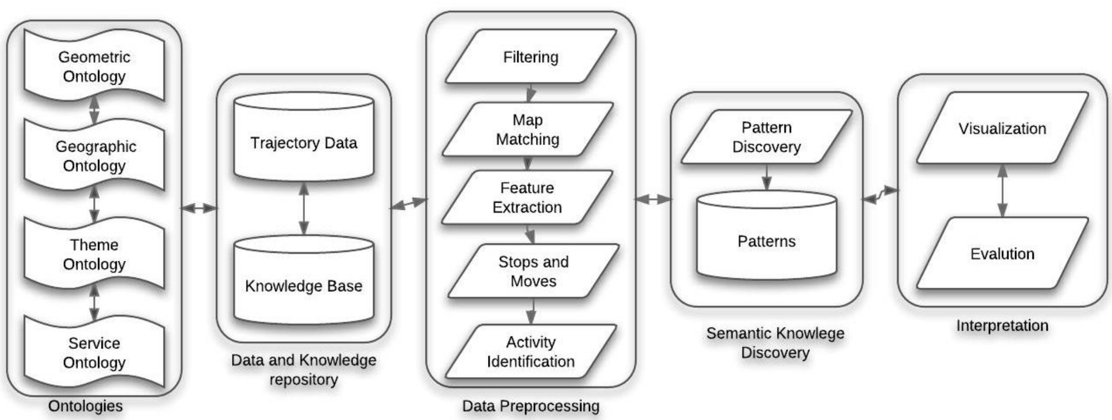

Figure 1.1. The proposed methodology

Ontologies

Uschold and Jasper (1999) defined an ontology as something that “may take a variety of forms, but it will necessarily include a vocabulary of terms and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms.” A vital concern in DM is the aggregation of data at higher abstraction levels (Bogorny et al. 2011). Since considering a variety of parameters representing some relationship or process can be hard, the building of an ontology can provide more flexibility not only for pre-processing data, but also for filtering and interpreting discovered patterns in a post-processing step. This step proposes different ontologies with the goal of supporting creation, management and analysis of semantic trajectory data. As shown in Figure 1.1, ontologies include geometric, geographic, theme and service ontologies.

The geometric ontology is composed of three major components: (i) movement characteristics, (ii) movement path, and (iii) external factors to describe trajectories. The movement characteristics include spatial and temporal information which used to specify spatio-temporal features. This research considers discontinuous movement paths and uses the stops and moves concepts (Spaccapietra et al. 2008) to define segmentation of a trajectory. Identification of stops and associated attributes, such as stop duration and stop frequency, can help to identify different activities (Huang, Li, and Yue 2010). As shown in Figure 1.2, we divided activities into four related types: shopping, recreation, professional, and other. Shopping refers to time spent at malls/retail stores, etc. to acquire ones daily needs. Recreation might include going to the cinema, the pub, sporting and outdoor activities, and other places related to leisure. Professional activities could be categorized as employee, student, etc., and others activity might include cultural and religious activities, staying at home, etc.

Figure 1.2. Activity classification

Additionally, external factors could include local environmental factors, weather, etc., and the context for a movement activity, as well as factors restricting the movement object, such as spatial constraints. The geographic ontology includes a variety of land uses, road networks, landmarks, and points of interest (POI). A POI is a specific location that some user may find useful or interesting. It can hold information about different kind of objects such as restaurants or parks. The theme ontology gathers all application dependent concepts such as traffic management ontologies, bird migration ontologies, and transportation ontologies, etc., and is generally domain specific (Yan et al. 2008). The first three ontologies include the concepts that help in finding out how movement can be identified as a set of structured trajectories. The service ontology focuses on available services. It represents the services that a system offers to users along with their description. The service is linked to the concept of parameters which represent the input required for DM processes.

To create an ontology one needs to follow four steps. The first is specification of activities, which describes why the ontology is being constructed, and who are its intended uses. In this research, the specification includes activity identification as the domain ontology and the preprocessing phase of DM. Secondly, conceptualization of activities converts an informally perceived view of the domain ontology into a conceptual model, typically represented as a graphs and/or tables. Third, formalization, which transforms the conceptual model into a formal machine-readable model, and last, implementation, which codes machine-readable models in a computational ontology language using an ontology editor.

Data and Knowledge repository

The second methodological step is the collation of trajectory data and existing knowledge base for the domain of interest. The knowledge base generally consists of maps/layers, and expert knowledge. Trajectory data consists of the raw [x, y, t] observations, and other relevant attributes such as speed, direction, etc. Trajectory data can be enriched through integration with background knowledge (Ong et al. 2010). The knowledge base composes ontology models and resulting classes and relationships derived from the first step. Therefore, various other data sources can be specified by expert knowledge according to application specific needs.

Data Preprocessing

Data preprocessing converts raw movement data into trajectories, and filters the data to remove outliers and noise to improve consistency. Map matching (Brakatsoulas et al. 2005) can be performed at this point to pair the trajectories with other map data sources. The next step is to extract activity types. Since individual trajectories of daily life contains several activities, it is necessary to detect where potential activities might occur within a trajectory (Andrienko and Andrienko 2007). Extracting activities helps to semantically organize and interpret movements. The ontology represents the concepts, rules, and assumptions present in an application domain.

Semantic Knowledge Discovery

Once the data is semantically annotated, a DM algorithm is being applied to extract patterns through the application of knowledge discovery and DM techniques. The objective is to discover patterns and structure in the movement data that could be used to generate useful knowledge about the behavior of moving objects. The knowledge extraction process can be carried out using the main DM techniques such as pattern discovery, classification, clustering, and similarity analysis (Miller and Han 2009; Giannotti, and Pedreschi 2008).

Interpretation

The task of interpreting movement patterns, typically left to the domain expert, is challenging since different pieces of background information such as specific characteristics of the moving entities must be put together and associating with the discovered patterns (Ong et al. 2010). After the knowledge extraction process, it is essential to reason about the detected patterns, and evaluates the reliability, meaningfulness, and interestingness of the outcomes. Visualization methods are required in order to present suitable interpretation of the results, and also deliver the appropriate knowledge about the movement data (Giannotti, and Pedreschi 2008). As such, patterns should be filtered, visualized, and/or interpreted in the post-processing step based on a service ontology to extract knowledge that is meaningful to the user, and to understand the correlation among moving entities belonging to a given pattern, and correlation among the different discovered patterns themselves. In this research, the evaluation of the conceptual ontology will be performed as follows: (i) logical consistency will be checked for repetitions, and missing relationships, concepts, and instances; (ii) conceptual accuracy will be performed by domain experts to assess the validity of information generated with respect to the domain; (iii) clearness / vagueness need to be assessed against the conceptual ontology to determine reliability of the information generated. Evaluating a model ontology includes assessing the applicability and usefulness of generated information by comparing and evaluating the result from the ontological approach with the reality. We propose that this should be undertaken through the implementation of a recommendation services based on the users' recognized context.

Architecture

Figure 1.3 illustrates a client-server architecture for implementing the system. The architecture consists of three functional layers: a presentation layer; a service layer, and a data layer.

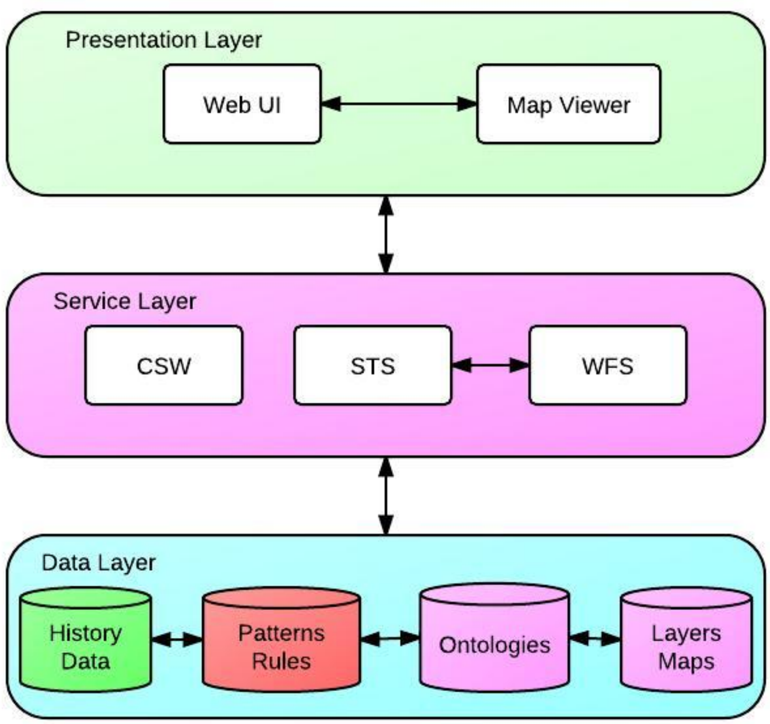

Figure 1.3. The client server architecture

Presentation Layer

The Presentation layer consists of a map viewer and web user interface (Web UI). It manages user interaction and displays results to the user. The Web UI component refers to the service ontology, which provides a listing of all services offered by the system and the input required by the user for accessing that service. The Map Viewer module uses the OpenLayers web mapping API and has capabilities for data visualization. Web client technologies such as JavaScript, HTML5, and CSS can also be used in this case. The Web UI component also provides an interactive interface for service providers to register their services to the system.

Service Layer

The Service Layer comprises a Catalogue Service Web (CSW), a Spatio Temporal Service (STS), and a Web Feature Service (WFS). CSW offers functionality to search and provide all geospatial data and services from a Patterns/Rules database. It can search for service providers and register them as a new service. STS includes services that provide space-time analysis capability, and once a service is matched with a users', based on the service ontology, then the STS will work with the WFS service to send results to the Map Viewer. The STS organizes patterns based on a set of IF-Then rules. The WFS enables users to access geospatial layers from Layers/Maps database.

Data Layer

The Data Layer consists of databases for historical movement data, Patterns and Rules, Ontologies and Layers/Maps. Historical data includes previous trajectory data, and current data received from the presentation layer. The Layers/Maps database includes different layers such as the road network, land use, and other data sets that are relevant to the application. The extracted patterns from historical data, defined rules and a catalogue of available services can be found in the ontology database. Ontologies include model and service ontologies. The model ontology can be used in the preprocessing step for knowledge discovery, and the service ontology can be used to match patterns with available services. For instance, if a recreation activity is identified for a user, the system could send and offer the user services focusing on museums or cinemas, if these were common user specific activities extracted from their historical data.

System Components

In order to develop the system, a number of components are required. As shown in Figure 1.4, the overall system has several essential components such as a database server, ontology construction, DM modules, Reasoning engine, LBS server, and user interface. For this work, the spatial database extension PostGIS 2.0 for PostgreSQL 9.1.4 will be used to manage trajectory data, ontologies, and expert knowledge. The model ontology and service ontology will be built by using expert knowledge in the ontology construction component, and then employed by the DM modules. After that, the results of DM can be parsed as patterns to the service ontology where the reasoning engine interprets them. The reasoning engine is rule based and requires both spatial and non-spatial data in order to deliver potential and relevant services. The aim of the reasoning engine is to match and deduce useful context using the service ontology. The LBS server is composed of a web server such as Apache and a web map such as Open Layers to communicate with the database server and the user interface. Finally, the LBS server sends the results to be displayed via the user interface.

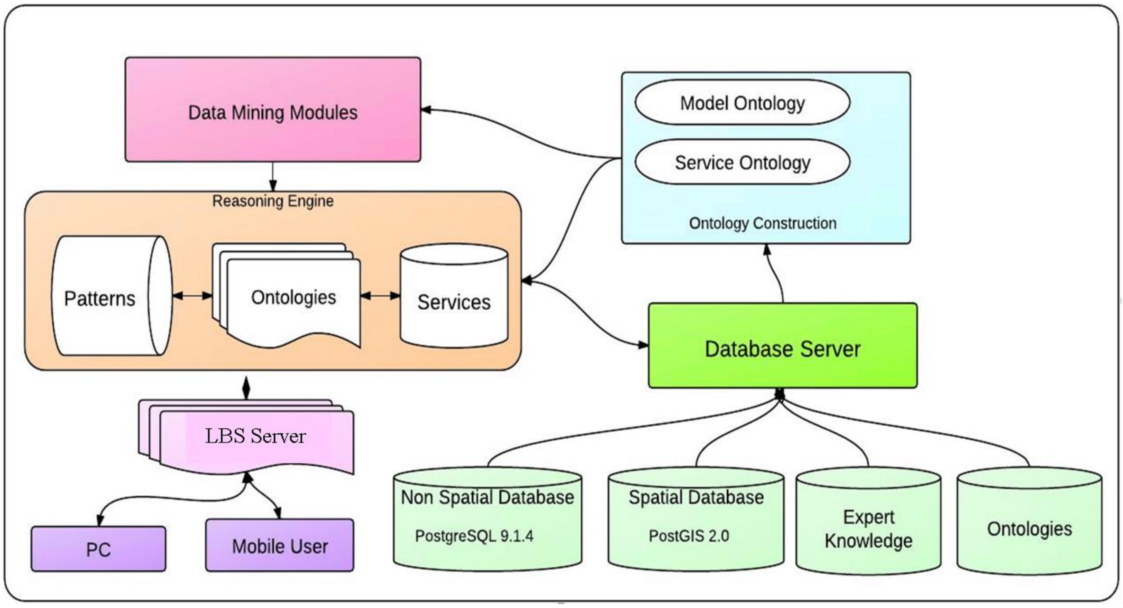

Figure 1.4 System components

Conclusions and future works

In this paper we suggest that knowledge discovery is a process that is application dependent and that there is a need to integrate geographic information into the analysis of trajectories in order to extract clearer and more meaningful patterns. Therefore, we have proposed an ontological approach in order to improve the pre- and post-processing of movement data for semantic knowledge discovery. We have also presented an architecture and system components that we consider necessary for the implementation of an online knowledge discovery LBS.

References

Alvares, L.O., V. Bogorny, J.A.F. de Macedo, B. Moelans, and S. Spaccapietra. 2007. “Dynamic Modeling of Trajectory Patterns Using Data Mining and Reverse Engineering.” In *Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual modeling-Volume 83*, 149–154.

Andrienko, G., and N. Andrienko. 2007. “Extracting Patterns of Individual Movement Behaviour from a Massive Collection of Tracked Positions.” In *Workshop on Behaviour Modelling and Interpretation. Technical Report*, 42:1–16.

Baglioni, M., J.A. Fernandes de Macêdo, C. Renso, R. Trasarti, and M. Wachowicz. 2009. “Towards Semantic Interpretation of Movement Behavior.” *Advances in GIScience*: 271–288.

Bauer, M., and S. Baldes. 2005. “An Ontology-based Interface for Machine Learning.” In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 314–316.

Bernstein, A., S. Hill, and F. Provost. 2002. “Intelligent Assistance for the Data Mining Process: An Ontology-based Approach.” *NYU Working Paper No. 2451/14146*.

Bogorny, V., H. Avancini, B.C. de Paula, C.R. Kuplich, and L.O. Alvares. 2011. “Weka-STPM: a Software Architecture and Prototype for Semantic Trajectory Data Mining and Visualization.” *Transactions in GIS* 15 (2): 227–248.

Brakatsoulas, S., D. Pfoser, R. Salas, and C. Wenk. 2005. “On Map-matching Vehicle Tracking Data.” In *Proceedings of the 31st International Conference on Very Large Data Bases*, 853–864.

Cannataro, M., and C. Comito. 2003. “A Data Mining Ontology for Grid Programming.” *Proceedings of (SemPGrid2003)*: 113–134.

Cao, H., N. Mamoulis, and D.W. Cheung. 2007. “Discovery of Periodic Patterns in Spatiotemporal Sequences.” *Knowledge and Data Engineering, IEEE Transactions On* 19 (4): 453–467.

Charest, M., and S. Delisle. 2006. “Ontology-guided Intelligent Data Mining Assistance: Combining Declarative and Procedural Knowledge.” In *Artificial Intelligence and Soft Computing*, 9–14.

Chen, G., and D. Kotz. 2000. “A Survey of Context-aware Mobile Computing Research.”

Dodge, S. 2011. “Exploring Movement Using Similarity Analysis”. PhD thesis, Universitaet Zuerich. https://www.zora.uzh.ch/59842/.

Dodge, S., R. Weibel, and A.K. Lautenschütz. 2008. “Towards a Taxonomy of Movement Patterns.” *Information Visualization* 7 (3-4): 240.

Giannotti, F., G. Giannotti, and D. Pedreschi. 2008. *Mobility, Data Mining, and Privacy: Geographic Knowledge Discovery*. Springer.

Güting, R.H., and M. Schneider. 2005. *Moving Objects Databases*. Morgan Kaufmann Pub.

Huang, L., Q. Li, and Y. Yue. 2010. “Activity Identification from Gps Trajectories Using Spatial Temporal Pois’ Attractiveness.” In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, 27–30. http://dl.acm.org/citation.cfm?id=1867704.

Imfeld, S. 2000. “Time, Point and Space- Towards a Better Analysis of Wildlife Data in GIS”. PhD thesis, Universitaet Zuerich.

Jasper, R., M. Uschold, and others. 1999. “A Framework for Understanding and Classifying Ontology Applications.” In *Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling, and Management KAW*, 99:16–21. http://folk.ntnu.no/alexanno/skole/WebInt/Articles/Articles.pdf.

Jensen, C.S. 2002. “Research Challenges in Location-enabled M-services.” In *Mobile Data Management, 2002. Proceedings. Third International Conference On*, 3–7.

Laube, P. 2005. “Analysing Point Motion, Spatio-Temporal Data Mining of Geospatial Lifelines”. PhD thesis, Universitz of Zurich.

Laube, P., S. Imfeld, and R. Weibel. 2005. “Discovering Relative Motion Patterns in Groups of Moving Point Objects.” *International Journal of Geographical Information Science* 19 (6): 639–668.

Lee, J.G., J. Han, X. Li, and H. Gonzalez. 2008. “TraClass: Trajectory Classification Using Hierarchical Region-based and Trajectory-based Clustering.” *Proceedings of the VLDB Endowment* 1 (1): 1081–1094.

Li, Y., J. Han, and J. Yang. 2004. “Clustering Moving Objects.” In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 617–622.

Miller, H.J., and J. Han. 2009. *Geographic Data Mining and Knowledge Discovery*. CRC.

Nanni, M., B. Kuijpers, C. Körner, M. May, and D. Pedreschi. 2008. “Spatiotemporal Data Mining.” *Mobility, Data Mining and Privacy*: 267–296.

Nanni, M., and D. Pedreschi. 2006. “Time-focused Clustering of Trajectories of Moving Objects.” *Journal of Intelligent Information Systems* 27 (3): 267–289.

Niaraki, A. S., and K. Kim. 2009. “Ontology Based Personalized Route Planning System Using a Multi-criteria Decision Making Approach.” *Expert Systems with Applications* 36 (2): 2250–2259.

Ong, R., M. Wachowicz, M. Nanni, and C. Renso. 2010. “From Pattern Discovery to Pattern Interpretation in Movement Data.” In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference On*, 527–534.

Pelekis, N., Y. Theodoridis, S. Vosinakis, and T. Panayiotopoulos. 2006. “Hermes–a Framework for Location-based Data Management.” *Advances in Database Technology-EDBT 2006*: 1130–1134.

Phillips, J., and B. G. Buchanan. 2001. “Ontology-guided Knowledge Discovery in Databases.” In *Proceedings of the 1st International Conference on Knowledge Capture*, 123–130.

Quintas, A.M., J.J. Costa, and V.H. Ribeiro. 2003. “The Use of GIS in the Analysis of Customers Mobility Routes.”

Smith, B., and D. M. Mark. 1998. “Ontology and Geographic Kinds.”

Spaccapietra, S., C. Parent, M.L. Damiani, J.A. De Macedo, F. Porto, and C. Vangenot. 2008. “A Conceptual View on Trajectories.” *Data & Knowledge Engineering* 65 (1): 126–146.

Steiniger, S., M. Neun, and A. Edwardes. 2006. “Foundations of Location Based Services.” *CartouCHe1-Lecture Notes on LBS* 1.

Trasarti, R., S. Rinzivillo, F. Pinelli, M. Nanni, A. Monreale, C. Renso, D. Pedreschi, and F. Giannotti. 2010. “Exploring Real Mobility Data with M-atlas.” *Machine Learning and Knowledge Discovery in Databases*: 624–627.

Tsoukatos, I.I., and D. Gunopulos. 2001. “Efficient Mining of Spatiotemporal Patterns.” *Advances in Spatial and Temporal Databases*: 425–442.

Usery, E. L., M. Azami, and M. P. Kwan. 2004. “Geospatial Ontology Development and Semantic Analytics.”

Yan, Z., J. Macedo, C. Parent, and S. Spaccapietra. 2008. “Trajectory Ontologies and Queries.” *Transactions in GIS* 12: 75–91.