2016

Ontology Based Semantic Knowledge Discovery for Movement Behaviors

Mousavi, Ali

UNIVERSITY OF CALGARY

Ontology Based Semantic Knowledge Discovery for Movement Behaviors

by

Ali Mousavi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOMATICS ENGINEERING

CALGARY, ALBERTA

OCTOBER, 2016

Abstract

The vast collection of movement data through various mobile devices generates a significant and precious amount of information that can provide valuable insight into movement behaviors in various applications. Researchers from different communities have developed models and techniques for mobility analysis, but they mainly are focused on the geometric properties of trajectories. As such, the techniques are good at discovering patterns, but the patterns are difficult to interpret in a particular application domain.

This thesis proposes an ontology based semantic knowledge discovery framework to understand mobility data and semantically interpret trajectory patterns. It consists of three main parts, namely: semantic trajectory ontology modelling, activity recognition, and semantic behavior modelling. First, a semantic conceptual data model is defined, which helps in developing an ontology model. The activity recognition part consists of several steps, namely: data preparation, semantic enrichment process, and semantic features extraction. Activity types were defined as axioms based on the semantic features. The retrieved information from the previous steps are used to populate the ontology model for classifying different activity types by reasoning. Next, the association rule mining algorithm, *apriori*, is applied to extract different behavior patterns. The process considers four different behavior types, namely: semantic, semantic and space, semantic and time, and semantic and space-time.

A system prototype was developed to evaluate the performance of the framework using a simulated dataset and two different real datasets. For evaluating the activity recognition model, the inferred activity types were compared to the activity types declared by users in the feedback. For evaluating the ontology based behavior model, one of the location based services named location based advertisement was considered to test different aspects of the extracted behavior models. In the activity recognition model, it was observed that the accuracy of the results was related to the availability of the points of interest around the places that users had stopped at some parts. In the semantic behavior modelling, the results showed that applying the extracted behavior rulesets could filter relevant services for the users from the number of available services and customize the services based on the rulesets.

Acknowledgements

I would like to take this opportunity to express my deepest appreciation to my supervisory committee, Dr. Andrew Hunter, Prof. Danielle Marceau, and Prof. Derek Lichti, for supporting, guiding, encouraging, and mentoring me throughout my PhD research at the U of C. I also would like to acknowledge my exam committee members Dr. Steve Liang, Dr. Xin Wang, Dr. M.A. Rajabi, and Dr. Lina Kattan, who improved this work through their valuable comments and discussions. Also, I would like to thank the faculty and staff members of the Geomatics Department, Schulich School of Engineering for providing a wonderful educational environment.

Furthermore, my sincere thanks go to all my work colleagues Ehsan Mohammadi and Ebrahim Poorazizi in the GIScience research group for their friendship, support, and productive discussions during these years. I would also want to express my gratitude to Dr. Abbas Sheikh Mohammad Zadeh for shedding light on some of the concepts used in this research. Also, I would like to acknowledge my friends, Dr. Ivan Detchev, Dr. Mohammad Akbari, Dr. Sina Taghvakish, Mahmood Exirifard, and Samira Rizaee for editing the thesis, helping in data collection, and valuable discussion during different stages of my PhD studies. This study would not have been possible without their knowledge and assistance.

Finally, I want to give special thanks to my family. Their moral support has made it possible for me to excel in school and without them I would not be where I am today. I would like to especially thank my parents, Kheirieh and Vali for their unconditional love and trust. I would also like to thank my brothers and sisters for their continuous encouragement over the years. I love you all from the bottom of my heart.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Symbol	Definition
GPS	Global Positioning Systems
GSM	Global System for Mobile
Wi-Fi	Wireless Fidelity
LBS	Location Based Services
POI	Points of Interest
GIS	Geographic Information System
DBSCAN	Density Based Spatial Clustering of Applications with Noise
KDD	Knowledge Discovery in Databases
GKD	Geographic Knowledge Discovery
ADT	Abstract Data Type
AI	Artificial Intelligence
ROI	Regions of Interest
LOI	Lines of Interest
SQL	Structured Query Language
AT	Activity Type
W3C	World Wide Web Consortium
OSM	Open Street Map
MWD	Maximum Walking Distance
UWS	User Walking Speed
MST	Minimum Service Time

OWL	Ontology Web Language
API	Application Programming Interface
XML	Extensible Markup Language
WEKA	Waikato Environment for Knowledge Analysis
RDF	Resource Description Framework
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheets

CHAPTER ONE: INTRODUCTION

1.1 Background

The advancement of location technologies such as Global System for Mobile (GSM), Global Positioning Systems (GPS) in mobile devices, and wireless communication has enabled deployment of a variety of Internet–based services such as Location Based Services (LBS) (Steiniger et al., 2006; Baldauf et al., 2011; Viktoratos et al., 2014). The application domains for these types of services are typically transportation management, urban planning, tourism, location-aware advertising, and integrated information services (Jensen, 2002; Furletti et al., 2013). The widespread use of these applications and services in our daily activities has led to a large number of positioning data that can be represented as trajectories (Ong et al., 2010; Bogorny et al., 2014). Generally, these types of data take the form of an [x, y, t] triplet that represents 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, such as location aware advertising to users, they need to identify relevant customers at the right time, and in the right place. In other words, they need to know where, when, and which services to provide them. Therefore, most systems designed to support LBS 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 (Mehra, 2012; Viktoratos et al., 2014). However, this thesis hypothesises that it could be beneficial to provide proactive services to users that consider their movement patterns and behaviors.

To customize these services, an efficient analysis of movement data from across different application domains is required to identify similar behavior, or discover regularities between users that can be used to predict user's future behavior (Nanni et al., 2008; Karamshuk et al., 2013). 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, what type of information might they want to know?" (Quintas et al., 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 some analytical techniques (Kraak, 2003; Galton, 2005; Miller, 2005; SOOD, 2012; Miller et al., 2015) and computational methods

(Dodge, 2011; Imfeld, 2000; Laube, 2005) for the analysis of movement data. Moreover, data management and database community have focused on the modelling and querying of spatiotemporal data types such as the moving point and moving region (Güting et al., 2000; Guting et al., 2005; Güting and Schneider, 2005; Pelekis et al., 2008, 2008). Some studies also explored data mining algorithms for trajectory pattern discovery (Jeung et al., 2011, 2008; Lee et al., 2008). Different challenges arise when developing new exploratory tools: how should one discover similar trajectories (Lee et al., 2008), periodic movement (Cao et al., 2007; Li et al., 2010), or identify relative motion patterns (Laube et al., 2005)? Accordingly, a large number of studies have established approaches to utilize movement data for various aspects of knowledge discovery, such as trajectory data analysis, movement pattern mining, and exploratory visual analytics (Giannotti, 2011; Imfeld, 2000; Laube, 2005; Mountain, 2005).

1.2 Motivation

The main motivations of this research are:

- To incorporate semantics into geometry
- To propose an ontology based semantic knowledge discovery framework

To explore user behavior, most research in the database community has focused on geometric and temporal characteristics of trajectory data, and the main problem has been the difficulty of correlating extracted patterns with movement behavior [35]. As such, the techniques are good at discovering patterns, but the patterns are difficult to interpret in a particular application domain. For example, when considering only geometric properties, one could discover a dense area where trajectories meet. But without semantics, it is hard to find out why the trajectories meet, and consequently what might attract the users. Moreover, most classical data mining methods focus on the mining step itself, usually, little attention is given to the whole knowledge discovery process, which includes data pre-processing, data transformation, data mining and post processing. This allows for the extraction of many patterns, but it makes pattern interpretation very difficult (Bogorny et al., 2011; Boulicaut and Masson, 2005). Furthermore, mined results can be made more meaningful when the nature of the movement data is considered as context within the mining process (Ong et al., 2010). According to Dodge (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. Therefore, trajectory data need to be reconsidered not only from the geometric view but also from the meaningful semantic view as well to interpret and understand their meaning.

Recently, few research efforts have investigated methods that support trajectories with rich conceptual models where semantics of movement can be clearly expressed and used to understand trajectory patterns (Baglioni et al., 2009a; Bogorny et al., 2011; Spaccapietra et al., 2008; Trasarti et al., 2010). An example is the concept of semantic trajectories as a sequence of stops and moves introduced in (Spaccapietra et al., 2008), wherein a stop is defined as an interesting place in which some moving entities have paused for some time, while a move is defined as the portion of the entities' trajectory between two stops. This model has been adopted in several algorithms as a standard for semantic trajectory data analysis. These works, still have unresolved questions. They only discover stops and moves of moving objects. However, by only identifying stops, it can still be difficult to identify the activity type of the moving object. Therefore, one needs to extract some semantic features such as stop duration and stop frequency, and start begin time from the trajectory data.

Considering former conceptual studies on trajectories, the motivation of this thesis is to further explore semantic methods for analyzing trajectory data, not only from a geometric view, but also from a semantic perspective as well. This can be obtained by means of a semantic process, where raw trajectories are enhanced with semantic information and integrated with geographic knowledge encoded in an ontology. Analysis methods for interpreting the semantics of context within the knowledge discovery process not only leads to discovery of semantic trajectory patterns (Ong et al., 2010), but understanding context also helps with the extraction of behavioral patterns that assist in the understanding of the movement of an object. Using model ontologies, it is possible to improve communication between user and system, and provide intelligent and flexible services that are capable of recommending the most suitable services to a user (Charest and Delisle, 2006).

Therefore, the main idea of this thesis is to propose an ontology based semantic knowledge discovery framework to understand mobility data and semantically interpret trajectory patterns. This not only improves the pre-processing step by associating semantics to raw trajectory data, but also makes the interpretation of discovered patterns easier during post-processing steps. To achieve this goal, it is necessary to define a semantic conceptual data model and a number of computing techniques to utilize a rich model and reconstruct meaningful trajectories from the

movement data. This would be beneficial for extracting activity types and modelling of user behavior.

1.3 Research questions

The following fundamental research questions will be explored and answered in this thesis.

Q1. What are the fundamental modelling requirements for representing trajectory data?

Q2. What are the crucial components of movement, which are essential for defining semantically annotated movement patterns?

Q3. How to semantically annotate trajectory data with auxiliary geographic information and how can the annotated data be used to extract user activity type?

Q4. How to semantically model behavior of a user by considering different activity types?

Q5. How can the extracted behavior model be used in LBS applications such as location based advertisement?

1.4 Research objectives

With regards to the motivation and research questions to develop a semantic approach for understanding movement data, the main objective of this research is to design and develop a framework that enriches user movement data by considering ontology in order to explore and interpret extracted semantic patterns. To accomplish this objective, five sub-objectives are defined:

Objective (a): Semantic conceptual data model definition – The most important component of understanding movement behavior is defining a conceptual model of movement that describes behavior of objects. This conceptual data model helps in developing an ontology model that covers four different dimensions including geometry, geography, theme, and service to address the interactions between them. The purpose of this objective is to address research questions Q1 and Q2.

Objective (b): Activity recognition – Development of an activity recognition algorithm is necessary to integrate various sources of data. This study investigated various extracted semantic features and background information based on the ontology model to extract different activity types. This objective addresses research question Q3.

Objective (c): Semantic behavior modelling – To extract regularities between different user activity types to semantically model behavior of the user. The association rule technique is used to extract semantic behavior models. This objective addresses research question Q4.

Objective (d): Prototype development – The final objective of this research is to develop a prototype to evaluate the proposed framework using one of the LBSs called location based advertisement. This is accomplished by customizing location based advertisement on the extracted behavior models. This objective addresses research question Q5.

The scientific contributions of this thesis parallel the listed research objectives. They will be explained in detail in chapter 3 and are summaries in chapter 6.

1.5 Research workflow

Figure 1-1 illustrates a schematic workflow of the research. It consists of four main parts, namely: semantic trajectory ontology modelling, activity recognition, semantic behavior modelling, and prototype development.

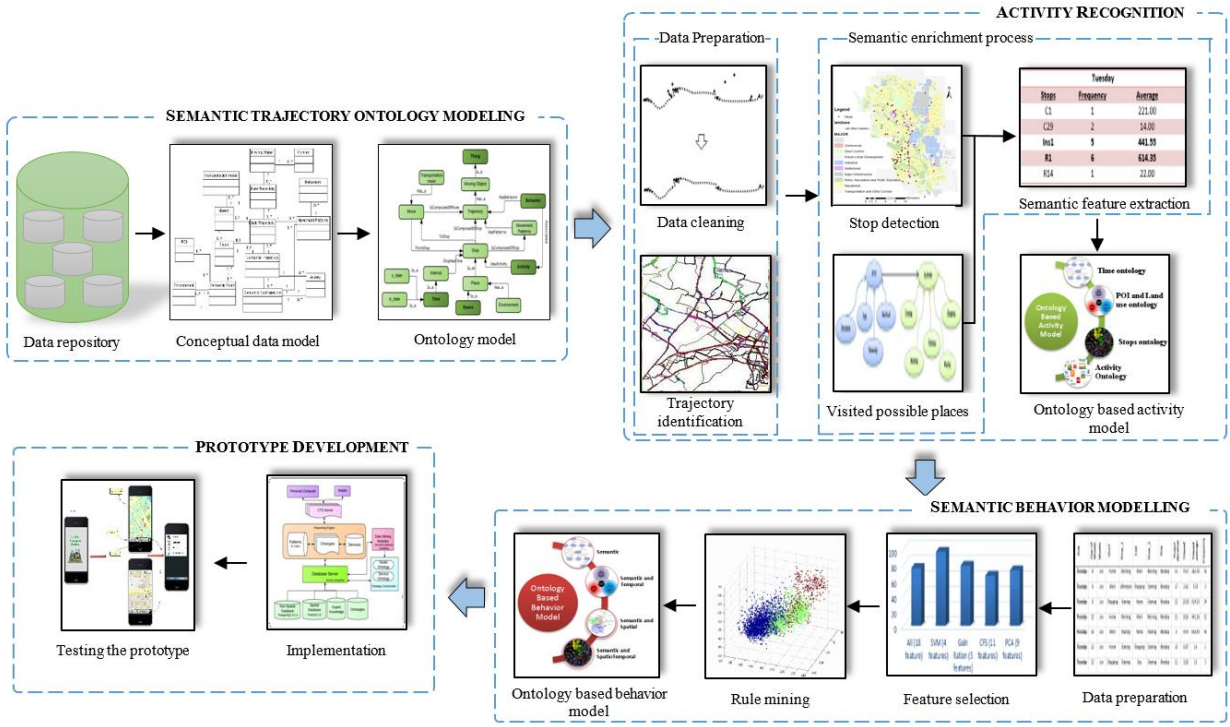


Figure 1-1 Schematic workflow of the research

The first part generally contains maps/layers and an application domain. The data can be specified using expert domain knowledge specific to the application needs such as land use, road network, Point of Interest (POI), and other data types. In this part, a conceptual data model is defined and based on that, a semantic trajectory ontology model is designed. Then the ontology concepts are mapped to the users' trajectory data. Ontologies include geometry, geography, theme, and service.

The activity recognition part consists of several steps. The first step is data preparation, where raw data is cleaned and daily and weekly basis trajectories are identified. The second step is semantic enrichment process, which includes stop detection, finding probable visited places, and extracting semantic features. Once stops are detected, they are annotated with the POI and landuse types and finally, several semantic features such as stop begin time, stop frequency, and average duration are extracted. The retrieved information from the previous steps are used to populate the ontology model for classifying different activity types by reasoning. Extracting activity types helps the system to semantically organize and interpret the trajectories. The extracted activity types are evaluated by users through a web interface, which asks them about the correctness of the inferred results.

Once the data have been semantically annotated and the activities are extracted, association rule mining is applied to extract semantic behavior patterns with respect to the semantics generated in the previous step. This part consists of different steps such as data preparation, feature selection, rule mining, and ontology based behavior modelling. The model considers four different behavior types, namely: semantic, semantic and space, semantic and time, and semantic and space-time.

In the last part, a prototype is developed to evaluate the ontology based behavior model in different aspects by considering one of the LBSs called location based advertisement. The prototype has several essential components such as a database server, an ontology constructor, data mining module, a reasoning engine, LBS server, and a user interface. In this research, the spatial database extension PostGIS for PostgreSQL is used to manage the trajectory data and ontologies. The ontology model is built in the ontology constructor component, and then employed by the data mining module. After that, the results of data mining are parsed as patterns to the service ontology, where the reasoning engine interprets the patterns. The reasoning engine is rules-based and requires both spatial and non-spatial data in order to deliver relevant services.

1.6 Research assumptions

This research assumes that using semantic enrichment process not only can make the interpretation of discovered patterns easier but also can help in customizing the available services in the LBSs. This research assumes that users have to stop in order to perform different activities. Moreover, it is assumed that the users use their own private car to perform their own daily activities. This research hypothesizes a functional relationship for activity types based on several semantic features, namely: POI type, land use category type, stop frequency, stop duration, and stop begin time. It further hypothesises that with an understanding of these features, it is possible to infer activity types through the definition of a set of IF-THEN criteria. To detect POIs nearby users, it is assumed that users would like to walk from their vehicle's parking place to a certain destination in order to perform their activities. Moreover, it is assumed that during a stop users perform only one activity while sometimes more than one activity can be done. For data collection, users are required to install an application, which is developed specifically for this research.

1.7 Research scope and limitations

The aim of this research is to define and develop an ontology based semantic knowledge discovery framework and does not intend to improve the performance of any algorithms in terms of space or time complexity. The word "activity" in this research is what a user is going to do during his/her daily movement such as shopping and going to work. Moreover, the word "behavior" indicates the regularities between user's different activities. In order to extract different activity types, this research only considers the stops in the dataset and moves are not within the scope of this research. The proposed framework is designed to explore individual semantic behavior rules. Therefore, the framework would likely require modification for support in exploring collective behaviors. Activity based planning and trip chaining are some interesting topics regarding activity recognition and behavior modeling, which are out of the scope of this research.

As a limitation, this research currently does not focus on multimodal transportation data and it just concentrates on the users who drive their own cars. The research is limited to two different datasets, from two different cities. Multiple datasets would increase the robustness of the study. Although there exist some trajectory datasets such as GeoLife trajectory dataset or T-Drive taxi trajectories, these datasets lack semantic related information and also there is no access for the users to provide their feedback. Consequently, these trajectory datasets are not fit for experimental

evaluation. Another limitation of this research is the availability of the POIs. This research utilizes the POIs from available third party data sources such as open street map. These kinds of data sources do not provide POIs for all places and they might be incomplete for some cities or even entire countries. Moreover, the domain rules defined in the activity recognition method need to be modified in order to be used in a different city with a different cultural background.

1.8 Organization of the dissertation

The remainder of this dissertation is organised as follows:

- In Chapter 2 the background information regarding the research area is provided and some basic concepts and preliminaries are defined. Then, a summary of various research studies in particular studies on analyzing mobility data is presented.
- In Chapter 3 the proposed conceptual data model, which is used in this research to develop an ontology based model is described. Then the framework is presented, which consists of three different steps, namely: semantic trajectory ontology modelling, activity recognition, and semantic behavioral modelling.
- In Chapter 4 a prototype is developed to evaluate the framework. In this chapter each step of the process is presented: from the ontology model definition to the activity recognition and behavior modelling. Moreover, the main components of the prototype are explained in detail.
- In Chapter 5 the conducted experiments on different datasets are illustrated. A summary of the results is presented and the implications of the evaluation outcomes are discussed.
- In Chapter 6 the research with a summary of the contributions is concluded. In this chapter future research regarding further development is also described.

Figure 1-2 shows the thesis outline and the classification of each topic.

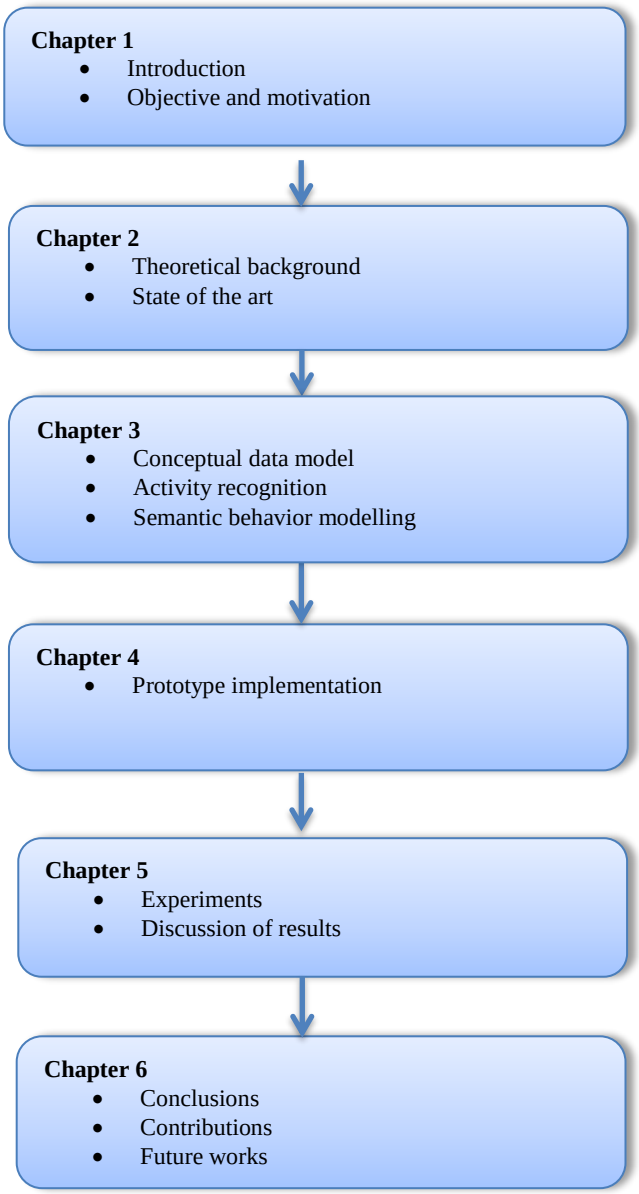


Figure 1-2 Thesis chapters' graph

CHAPTER TWO: THEORETICAL BACKGROUND AND STATE-OF-THE-ART

2.1 Introduction

Mobility is one of the major keywords that characterizes the current development of our society. GPS and other positioning devices enable capturing the evolving position of objects moving within a geographical space, which has created a large amount of tracking data. Therefore, researchers from the database, GIS, visualization, data mining, and knowledge extraction communities have developed models and techniques for analyzing this kind of data. In particular, a new promising approach has been devised to provide applications with richer and more meaningful knowledge about movement. This is achieved by combining the movement data with related contextual data.

This chapter consists of two main parts. The first part (Sections 2.2, 2.3, and 2.4) provides background information regarding the research area and defines some basic concepts and preliminaries used in this research to support the analysis of trajectories and their behaviors. In Section 2.2 some information about movement data and the semantic enrichment process of raw trajectories is provided, which includes trajectory segmentation into episodes and episodes annotation. In Sub-section 2.2.3 and 2.2.4, activity recognition and trajectory behaviors are also discussed, respectively. In Section 2.3 knowledge discovery for movement data is described, which includes movement pattern discovery, trajectory clustering, and trajectory classification. In Section 2.4, semantic trajectory modelling is presented. There are three different types: data type based, design pattern based, and ontology based modelling.

In the second part (Section 2.5) various studies in literature on analyzing mobility data are reviewed. For a better understanding, previous studies are presented in terms of four main perspectives as the research background of this thesis, i.e., semantic trajectory modelling, trajectory segmentation and annotation, activity recognition, and behavior modelling. Different perspectives such as general characteristic of movement data, ontology and knowledge discovery, and semantic based trajectory knowledge discovery are considered in the behavior modelling sub-section. Lastly, in Section 2.7 the chapter is summarized.

2.2 Movement Data

Movement is a vital aspect of almost all organisms and many spatiotemporal processes. Hence, it is crucial to understand movement and gain knowledge about its patterns. Recent advances in positioning technologies provide an increasing access to massive repositories of movement data

and hence challenges arise to develop new exploratory tools and knowledge discovery techniques in order to extract meaningful information, discover interesting patterns, and explore the behavior of moving objects such as humans, vehicles, and animals. A movement basically consists of a temporal sequence of spatiotemporal positions recorded for a moving object. However, depending on the capabilities of the device, additional data, e.g. the instant speed or stillness, acceleration, direction, and rotation may complement the (instant, point) pairs. This research considers movement of moving objects as moving points and it does not address the alternative views of movement such as movement of body parts such as eyes or hands, nor does address deformation issues raised when considering moving objects, such as hurricanes and oil spills, that span over a changing area or volume.

Definition 1 (Movement**):** Movement is defined as "a change in the spatial location of the whole individual in time", that is, as whole-body movement (Nathan et al., 2008). Figure 2-1 represents the movement of a moving object in a schematic way.

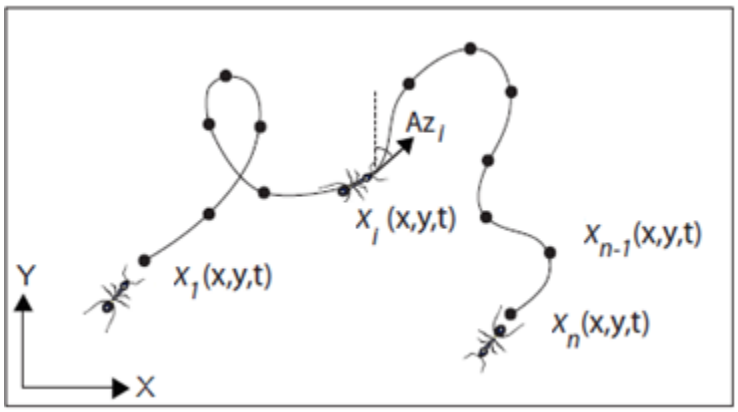


Figure 2-1 Trajectory of a moving object, representative of its movement path over time

Definition 2 (Moving object): Moving object, also called mobile object, is defined as an entity whose position changes over time. In this thesis, the object is a user who carries a mobile device and it is conceptualized by moving points, that is, the location of the object in time. The device is the instrument that collects the trajectory as a sequence of points, and it can be a GPS receiver, a cell phone, a sensor network, or any other device that captures the position of an object in a given time (Bogorny et al., 2014).

2.2.1 From movement to raw trajectories

The movement of an individual is represented by its trajectory, also called geospatial lifeline, as a time-ordered set of positions (Laube et al., 2007; Spaccapietra et al., 2008) (Figure 2.1). Raw trajectories are the segments of the object's movement that are of interest for a given application. Figure 2-2 shows a section of the movement of a moving object and, superimposed as a continuous line, two subsections identified as relevant raw trajectories named T1 and T2. Each trajectory is identified by two specific spatiotemporal positions of the movement, called the Begin and the End of the trajectory. They are the first and the last positions of the object for this trajectory (Spaccapietra et al., 2008).

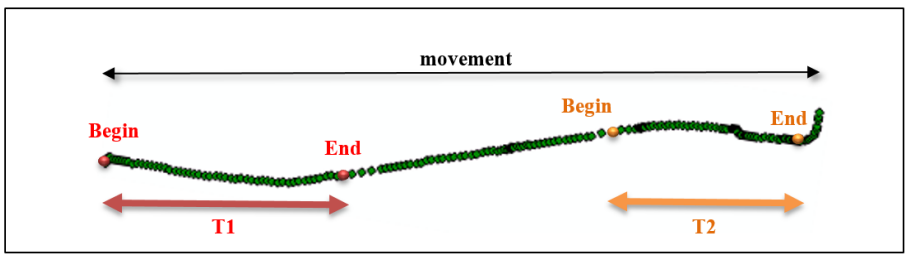


Figure 2-2 Trajectories extracted from a movement visualized as continuous line

Definition 3 (Raw trajectory): A raw trajectory T is an ordered list of points $(p_1, p_2, p_3, \dots, p_n)$, where $p = (x_i, y_i, t_i)$ and $t_1 < t_2 < t_3 < \cdots < t_n$, in which x and y are the spatial coordinates that represent a place and t is the timestamp in which the point was collected (Figure 2-3). The n is the number of sample points recorded during the movement of the object. Therefore, it is the footprint of different positions of a moving object, which is typically represented as a sequence of sample points, describing the spatial and temporal positions detected by a tracking device (R.H. Güting and Schneider, 2005).

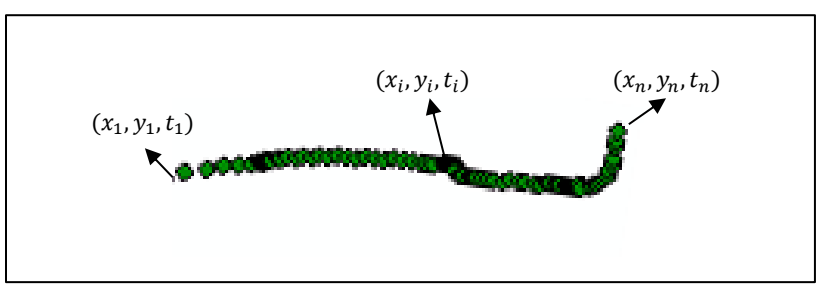


Figure 2-3 Raw trajectory as a sequence of spatiotemporal points

2.2.2 Semantic enrichment process of raw trajectories

Raw trajectories are great for applications that aim only at locating some moving object (e.g., where was the object at 11 am today?) or computing statistics on the spatiotemporal characteristics of trajectories (e.g., which percentage of trajectories show an average speed over 10 km/h?). However, most application analyses require complementing raw data with additional information from the application context. For example, interpreting trajectories of persons within a city requires some contextual information about the features of the city such as POI and ongoing events. The process of adding contextual data to raw trajectories is called semantic enrichment, which is key for supporting mobility and behavioral analyses that are of interest to the application at hand.

Definition 4 (Contextual data): Context is information about the moving object, coming either from the real world where the object moves or more abstractly from the application domain knowledge (Bogorny et al., 2014). It can be considered in terms of various dimensions such as who, what, where, when, why, and how (Moreno et al., 2014). Who, indicates the moving object that is generating the movement data. What, denotes the type of activity performed by a moving object. Where, describes the place where the activity type was made. When, shows the time about the executed activity. Why, states the reason a moving object has executed an activity. How, discusses the way the context information was gained (e.g. from different sensors).

Definition 5 (Semantic trajectory): A semantic trajectory is a trajectory that has been enhanced using contextual information attached either to a raw trajectory as a whole or to some of its sub-trajectories (episodes) (Parent et al., 2013; Bogorny et al., 2014); see definition 6. Figure 2-4 illustrates a raw trajectory, which is a sequence of timestamped locations and a semantic trajectory, which is a sequence of interesting information like departure or destination places of interest, and activities.

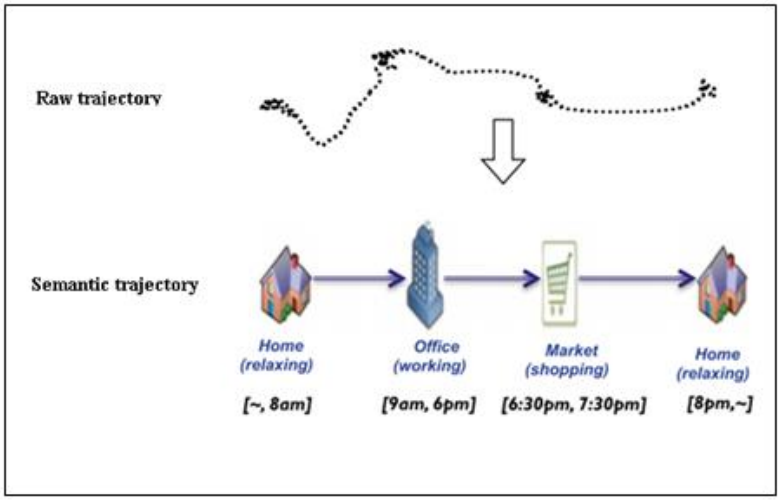


Figure 2-4 Raw and semantic trajectories (Bogorny et al., 2014)

2.2.2.1 Trajectory segmentation into episodes

Trajectory segmentation is the process of segmenting a raw trajectory to meaningful episodes, which is driven by application dependent criteria.

Definition 6 (Episode): An episode e of T is a list of consecutive points $(p_1, p_2, \dots, p_k)$, where $e \subset T, k \geq 1$, and $k + 1 \leq n$. A sub-trajectory can have several definitions, but in general, in the literature it is defined as a segment of a raw trajectory (Bogorny et al., 2014). Therefore, an episode is itself a raw trajectory.

A trajectory may be segmented in various ways corresponding to whatever segmentation criteria are of interest to the application at hand. A popular trajectory segmentation criterion is stillness versus movement, which splits a trajectory into periods of time when the object is considered as stationary and periods where the object is indeed moving. The former periods are denoted as stop episodes while the latter are denoted as move episodes. According to this point of view, a trajectory is a sequence of alternating stop and move episodes.

Definition 7 (Stops and moves): A stop is a part of a trajectory, such that the user has explicitly defined this part of the trajectory to represent a stop. More formally, a stop has a non-empty time interval and the spatial extent is a single point. A move is a part of a trajectory delimited by two extremities that represent either two consecutive stops, or trajectory beginning or end. A move is

represented by spatiotemporal line. As shown in Figure 2-5, the input raw trajectory is segmented into seven episodes ($e_1, \dots, e_7$), among which, there are four stops ($S_1, \dots, S_4$) and three moves ($M_1, \dots, M_3$).

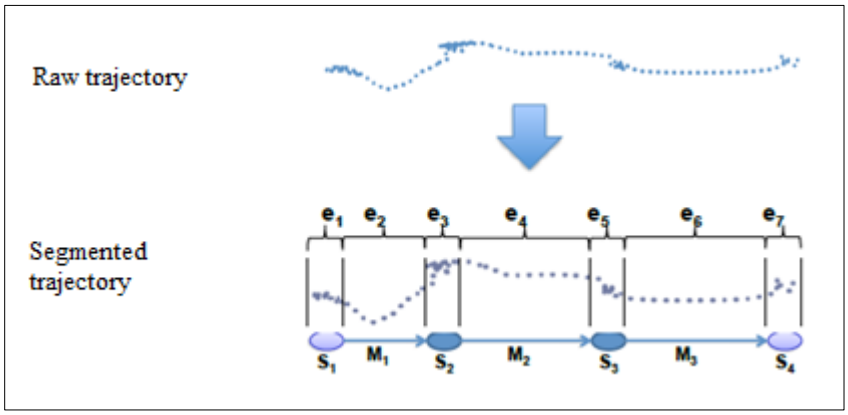
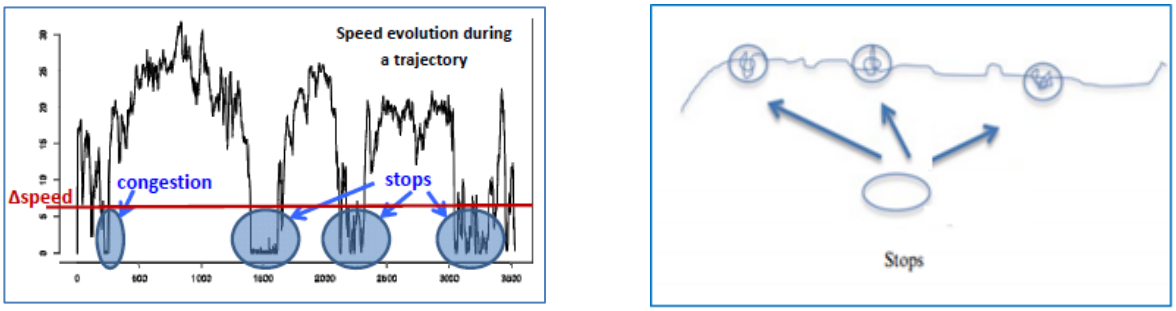


Figure 2-5 A sequence of episodes (Bogorny et al., 2014)

As already stated, identifying stops and moves within a trajectory is the responsibility of the application. The definition of a stop may undergo a number of application dependent constraints, e.g. a limited number of stops are allowed, or stops may have to conform to a minimal/maximal duration or distance between stops (Bogorny et al., 2009; Uren et al., 2006). Therefore, a popular challenge for researchers has been to elaborate ways of identifying stops. A simple approach associates stops with the parts of the raw trajectory where there is either absence of signal of the positioning device (e.g. the GPS is off), or the velocity of the moving object is zero for a given temporal interval (e.g. the car is parked). When segmenting a trajectory with respect to velocity, a speed threshold can be set, so that when the instant speed of a point is below the threshold, it is assumed to be part of a stop, otherwise it is assumed to be part of a move episode. Also, a minimum stop time parameter can be set in order to avoid short-term congestion (e.g. in traffic lights) to be considered as a stop. The weakness of this approach lies on the possibility that an actual Stop may not be identified, as due to signal errors, the measured speed is slightly above zero. To resolve this issue, one can use speed (e.g. less than 1 km/h) and temporal duration thresholds (e.g. more than 5 min) to imply stillness (Figure 2-6a). The threshold values are application dependant.

Stops may be discovered from combining contextual geographic information. For instance, a stop is discovered when a trajectory intersects the geometry of a place (from a set of

predefined POIs) and the duration of intersection is above a given temporal duration threshold. Alternatively, by detecting dense areas of the trajectory points, using e.g. a density-based point clustering method such as Density Based Spatial Clustering of Applications with Noise (DBSCAN), one may identify several clusters. The clusters, which correspond to the 'slower' parts of the trajectory are called potential Stops. Then, the stops are mapped to POIs. The two approaches sketched above are followed by an IB-SMoT (Intersection-Based Stops and Moves of Trajectories) (Palma et al., 2008) and its extension, the CB-SMoT (Clustering-Based Stops and Moves of Trajectories) (Alvares et al., 2007) technique. These approaches can be classified as density-based Stop discovery techniques (Figure 2-6b).



(a) (b)

Figure 2-6 Different trajectory segmentation techniques; (a) threshold (velocity) based and (b) cluster (density) based

2.2.2.2 Episodes annotation

Annotation is the process of attaching any contextual to either a trajectory as a whole or to some of its sub-trajectories. Annotation of stop or move episodes of a user is called episodes' annotation. An annotation value is an attribute value or a link to an object in the contextual data repository. The value may be captured by sensors (e.g. instant speed), computed from raw data, extracted from contextual data, or inferred by reasoning (Guc et al., 2008). Therefore, a raw trajectory can be annotated at different levels of detail (i.e. granularities): from the one extreme as a whole (e.g. a trajectory can be classified as "touristic") to the other extreme at point-by-point level (e.g. a position of the trajectory corresponds to a "hotel"). Obviously, annotating a trajectory as a whole may be inefficient in several cases whereas annotating the positions of a trajectory one-by-one generates a large number of repetitive annotations; it may be more effective to partition the

trajectory into meaningful parts, where each part corresponds to specific activity (e.g. “leisure”). Therefore, annotation is usually performed at the sub-trajectory (episode) level.

Stop episodes can be annotated with application contextual information, a POI that is the closest to the stop or the most likely POI where the moving object stopped, or some other application specific annotation (e.g. the activities pursued during the stop). On the other hand, a popular criterion for annotating a move episode is the means of transportation “walk”, “bus”, “bike”, or “car” used by the moving object. The typical approach is the characterization of each transportation mode, by considering parameters, such as speed (e.g. walking speed is less than 5 km/h), motion continuity (e.g. buses make frequent stops while taxis do not), direction, and route constraints (e.g. buses and trams use move on predefined routes). Instead of identifying the annotations directly from the geographical objects, an alternative approach is to annotate episodes with the activities taking place (e.g. “shopping” instead of specific shopping mall). In the next sub-section, activity recognition is elaborated on.

2.2.3 Activity recognition approaches

Activity recognition is a very active research topic for many domains, such as web log mining in human-computer interaction (Kotiyal et al., 2013) and mining activities from GPS and other mobility data (Xie et al., 2009; Spinsanti et al., 2010; Furletti et al., 2013). Personal GPS tracking trajectories generally consist of tuples that describe the geographic coordinates, time stamp, speed, heading, and other information of the person (Bhattacharya et al., 2012). Moreover, geographic context databases provide the possibility for activity identification based on GPS trajectories. Many studies start to focus on activity inference and annotation using GPS-based trajectory data. The key idea of these approaches is to extract the location (or movement) history of the individual, in conjunction with knowledge about the semantics of the locations (typically from geographic or application data repositories), for inferring the likely higher-level activity of the person. Therefore, the people's trajectory activity is typically recognized in terms of identifying the meaningful and significant locations (also called hotspots or POIs) from their trajectory data. Identifying stops and associated attributes, such as stop duration and stop frequency, can help in identifying different activity types (Huang et al., 2010).

Moreover, many studies have proposed different machine learning approaches from transportation perspective to predict travel times and extract activity types (Chen et al., 2013; Fox et al., 2014; Montini et al., 2014).

2.2.4 Trajectory behaviors

One of the ways to use trajectory data is by querying the data to find facts about some moving objects. Generally, applications focus on analytical querying to discover and analyze information extracted from trajectory databases. For example, trajectories of tourists may be analyzed for creating tourist profiles, which are useful for recommending personalized services, regulating the flows of tourists in the attraction sites, and adjusting the facilities. All that depends on discovering the similarities and dissimilarities among the trajectories, classifying the trajectories into groups of similar trajectories, and extracting the common characteristics that distinguish one group from another. A set of distinguishing characteristics forms a summary description of the group of trajectories, which are called patterns or behaviors (Laube, 2009).

A trajectory behavior is a set of characteristics that identifies a peculiar bearing of a moving object or a set of moving objects. Given the lack of standardized taxonomy or ontology of trajectory behaviors, it is essential for a mobility application to precisely define the predicates for the trajectory behaviors that are relevant to its goals. Existing taxonomic studies for spatiotemporal behaviors include Dodge et al. (2008), Laube and Imfeld (2002) and Thériault et al. (1999). Andrienko et al. (2007) addressed semantic behaviors that take into consideration terrain features (e.g. obstacles) as well as some semantic annotations such as means of transportation.

2.2.4.1 Behavior classification

Based on the characteristics of the trajectories involved in the predicate defining a behavior, behaviors are classified as spatiotemporal and semantic (Parent et al., 2013). A spatiotemporal behavior is a trajectory behavior whose predicate bears only on the spatial and/or temporal data without any contextual data. It also may rely on the spatiotemporal characteristics of the trajectories such as speed for a “speeding” behavior that characterizes vehicles moving above the speed limit. An example of a more generic, application independent behavior, is the “meet behavior” that characterizes a set of trajectories (Dodge et al., 2008). A set of trajectories shows

the meet behavior if every trajectory of the set roughly ends at the same point and at the same instant. A taxonomy defined by Dodge et al. (2008) describes many types of spatiotemporal behavior.

Behavior may also rely on the semantic information conveyed via the annotations and the contextual data that are linked through spatiotemporal relationships to the trajectories. The semantics of space (e.g. the place where a trajectory has passed) and the semantics of time (e.g. morning, afternoon or weekday instead of a timestamp) are considered more meaning is given to a behavior. This may improve the understanding of the behaviors, and may reveal behaviors that would be difficult to identify without using semantics. For example, a behavior of a group of objects defined as "converge in the morning to the city centre" bears more information than a convergence behavior expressing that a group of objects "converge at the same time to the same place". Figure 2-7 (left) shows an example of analyzing raw trajectories and discovering the spatiotemporal behavior "meet": A group of trajectories end at the same area. If the stops of the trajectories are annotated with contextual data as it can be seen in Figure 2-7 (right), frequent semantic behavior "going from home to a park for a festival on Friday evening" can be discovered. Therefore, a semantic behavior is a trajectory behavior whose predicate bears on some contextual data and possibly some spatial and/or temporal data.

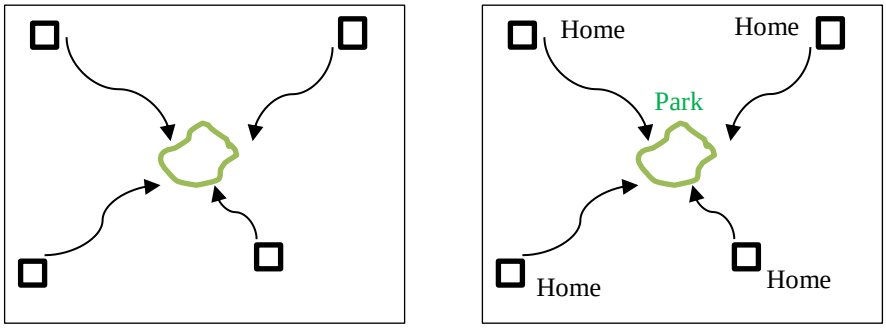


Figure 2-7 (left) Spatiotemporal behavior "meet". (right) Semantic behavior "going from home to a park for a festival on Friday evening"

One important characteristic of trajectory behaviors is whether they apply to single moving objects (individual behaviors), or to groups of moving objects (collective behaviors). An individual trajectory behavior is a trajectory behavior in which researchers focus on the dynamics

of single trajectories in terms of which regions they traversed. This leads to the definition of sequence behaviors. For this class, literature mainly focuses on the sequence of regions traversed by the trajectories. A sequence behavior is a trajectory behavior whose predicate specifies a sequence of component predicates that have to be satisfied in a specified temporal order. A collective trajectory behavior is a trajectory behavior in which researchers focus on characterizing behaviors implying the participation of several moving objects and showing specific interactions or coordination. The speeding and tourist behaviors are individual behaviors that characterize individual trajectories. On the opposite, the meet behavior is collective. This research focuses on individual behaviors. Different knowledge discovery methods are used to discover semantic behavior patterns.

2.3 Knowledge discovery for movement data

Knowledge Discovery in Databases (KDD) on movement data is a productive community, which develops a large amount of algorithms and methods (Giannotti et al., 2008). It is defined as a "process of obtaining information through data mining, and distilling this information into knowledge through interpretation of information and integration with existing knowledge" (Miller and Han, 2009). According to this definition, the KDD process consists of several steps; namely, data selection; data pre-processing; data enrichment; data reduction and projection; data mining and pattern recognition; and reporting. KDD is a nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data (Fayyad et al., 1996). In addition, data mining can be used as a step in the KDD process to reveal hidden information and patterns in a database (Fayyad et al., 1996).

Moreover, Geographic Knowledge Discovery (GKD) is a special case of KDD that deals with geographic data. Geographic data consists of different properties, such as high dimensionality, topology, geometry, spatial dependency, and spatial heterogeneity, which need to be considered in the knowledge discovery process (Miller and Han, 2009). Different attempts have been made to extend many techniques for knowledge discovery from trajectories such as classification, clustering, movement pattern discovery, location prediction, and time-series analysis (Bashir et al., 2007; Blythe et al., 1996; Fraile and Maybank, 1998). Knowledge discovery from trajectories aims at identifying behaviors, either among individual trajectories or groups of trajectories. The discovery of common behaviors among either individual trajectories or groups of

trajectories is based on similarity evaluations that allow defining meaningful classes of trajectories. Three machine-learning techniques are very popular in this context: movement pattern discovery, clustering, and classification.

2.3.1 Movement pattern discovery

Movement pattern discovery is referred to as the process of finding interesting patterns in a large movement dataset by applying data mining methods such as exploratory data analysis, descriptive and predictive modelling, and mining association rules. Fayyad et al. (1996) relate pattern extraction to fitting a model to data, finding structure, or making any high-level description from the data.

2.3.2 Trajectory clustering

Trajectory clustering is a well-known generic unsupervised learning technique that is applied to determine relevant groups of trajectories from the analysis of a large data set, without any a priori knowledge of the targeted grouping. Moreover, it can be applied to identify typical trends in datasets and supports deviation analysis to detect outliers and anomalies in data (Miller and Han, 2009). It is an exploratory data mining technique that facilitates the study of movement data by reducing its complexity, which aims at discovering the similarity in a set of movement trajectories, grouping similar trajectories into the same cluster, and finding the most common movement behaviors. Groups are computed based on similar characteristics such as shape, speed, distance, time, and direction that should be shared by trajectories either along the entire trajectories or along some sub-parts of the trajectories. In spatial and temporal clustering, the usual distance measures are spatial and temporal distances, respectively.

2.3.3 Trajectory classification

Trajectory classification is a complementary supervised learning technique that determines the appropriate grouping when the targeted classes are defined a priori. It is denoted as “finding rules or methods to assign data items to pre-existing classes” (Miller and Han, 2009). Accordingly, trajectory classification is defined as the process of constructing a model, and then applying

segmentation algorithms for identifying the type of trajectory based on the movement of an object (Lee et al., 2008).

2.4 Semantic Trajectories Modelling

Semantic trajectory modelling is a main task of the semantic trajectory construction. It is the process of defining and analyzing data requirements to support the application of trajectories. Most of the research on semantic trajectory has originated by a community within the GeoPKDD [1]group, whose original focus was on privacy awareness exploitation of spatiotemporal data. To continue the investigation on the discovery of knowledge and exploitation of moving object data, the group has been followed by MODAP [2](EU Coordination Action, FET OPEN, 2009) and more recently by SEEK [3](EU Marie Curie Project N 295179, 2009). The same community has recently presented a survey of the research in this area (Parent et al., 2013). Among the active initiatives aiming at boosting the research on moving object modelling, analysis and visualization, a notable contribution has originated also by the COST Action MOVE (European Science Foundation, 2009). For the representation and modelling of semantic trajectories, there are three different approaches; namely, data type based; design pattern based; and ontology based modelling.

2.4.1 Data type based modelling

The first approach is adopted in Zheni et al (2009), where the authors introduce an algebraic model that represents a spatiotemporal trajectory as an Abstract Data Type (ADT) that encapsulates the semantic dimension. A series of trajectory states is potentially observed and measured, and the ADT representation combines a formal definition with manipulation operations, allowing the user to formulate queries on the semantics of the spatiotemporal trajectory data type. Close to this approach one can also consider the work of Pfoser et al. (2003) that generates synthetic datasets of semantic trajectories.

2.4.2 Design pattern based modelling

The second approach has been presented by Spaccapietra et al. (2011), the same group that originally proposed the first conceptual model for the representation of semantics in trajectories

[1] Geographic Privacy-aware Knowledge Discovery and Delivery
[2] Mobility, Data Mining, and Privacy
[3] SEmantic Enrichment of trajectory Knowledge discovery

(Spaccapietra et al., 2008). This model has become a reference model for several trajectory data analysis research (Alvares et al., 2007; Bogorny et al., 2009; Spinsanti et al., 2010; Trasarti et al., 2011). As can be seen in Figure 2-8, Spaccapietra et al. (2008) specifically designed a comprehensive conceptual view on trajectory, which aims at exploring semantic for modelling trajectory data.

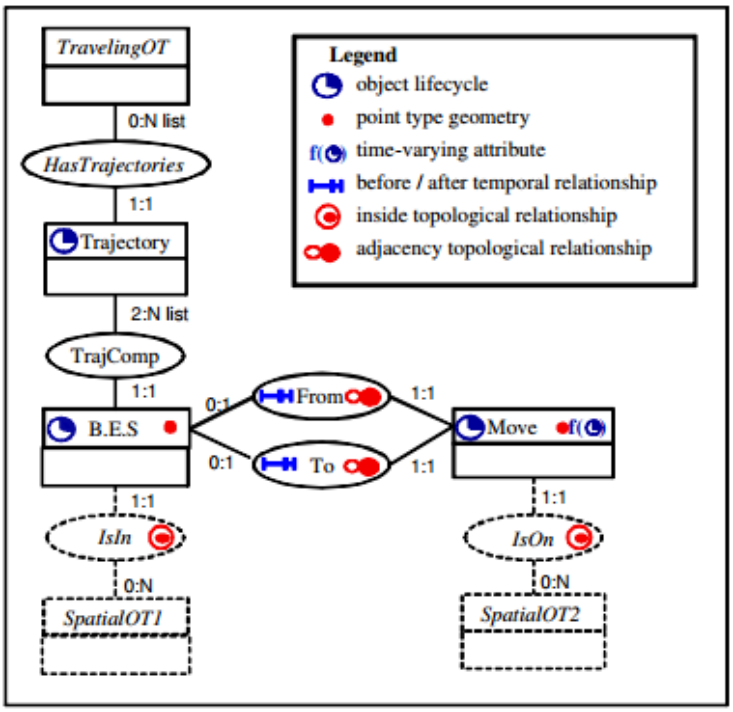


Figure 2-8 Conceptual view on trajectory (Spaccapietra et al., 2008)

The model relies on the conceptualization of stops and moves in trajectories. In a single trajectory, stop and move are the two crucial trajectory components. During the temporal extent of stop, namely $[t_{beginstop}, t_{endstop}]$, the spatial range of the trajectory is a single point; whilst, during the temporal extent of move, namely $[t_{beginmove}, t_{endmove}]$, the spatial range of the trajectory is a spatiotemporal line.

2.4.3 Ontology based modelling

Ontology has become a major research issue for most semantic-aware GIS studies. Ontology is a conceptual framework to formally model the semantic hierarchical structure of a system. They are

the relevant entities and relations that emerge from the observation of the world, and which are useful to the application domain (Guarino et al., 2009). A formal ontology is a logical theory specifically designed to capture the description of the world corresponding to the conceptualization. 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 (Bogorny et al., 2011). Since ontology is another important part of this research, more detail is provided in this regard.

2.4.3.1 Ontology concepts

In computer and information science, ontology is an explicit specification of a conceptualization (Gruber, 1995), consisting of a concept hierarchy and relationships between concepts. Neches et al. (1991) defined ontology as: "…basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary." Moreover, Studer et al. (1998) presented an alternate definition: "…a formal, explicit specification of a shared conceptualization". "Conceptualization" refers to an abstract model of some phenomenon in the world, through the identification of relevant concepts of that phenomenon. "Explicit" means the types of concepts used, and that the constraints to their use are unambiguously defined. "Formal" refers to the fact that the ontology should be machine-readable. The term "shared", reflects the notion that ontology captures consensual knowledge that is, it is not private to some individual, but accepted by a group.

Furthermore, Jasper and Uschold (1999) extended the definition of an ontology such that it 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". Ontologies are coupled with semantic approaches (Gruber and others, 1995; Guarino, 1998) to aid sharing and processing of information, either by automated tools, or by people. Ontologies have been used recently in many fields such as the semantic Web, Artificial Intelligence (AI), databases, expert systems, conceptual modelling, and so forth (Brisson, 2007; Guarino, 1998). They also play 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). Furthermore, the use of ontologies in GIS development has been discussed by Frank (1997) and Smith and Mark (1998). They state that "the use of ontologies will contribute to better information systems by avoiding problems such as inconsistencies between ontologies built in GIS, conflicts between the ontological concepts and their implementation, and conflicts between the common-sense ontology of the user and the mathematical concepts in the software". Spatial ontologies (Dellis and Paliouras, 2007; Grenon and Smith, 2004; Spaccapietra et al., 2004) become a major research issue for most semantically aware GIS studies, especially in some standardization organizations like the Open Geospatial Consortium (2006) and ISO Technical Committee, where the objective is to determine a set of formal and sharable concepts of geographic data. Regarding temporal ontology, instant and interval are the two fundamental concepts, together with a couple of temporal relations (e.g., begins, ends, before, after) that need to be specified and determined (Hobbs and Pan, 2004).

2.4.3.2 Main components of an ontology

Gruber (1992) identified that knowledge in ontologies can be formalized using five types of components: concepts, relations, functions, axioms, and instances. Concepts are used in a broad sense and can be abstract or concrete, elementary or composite, real or fictitious. Relations describe a type of interaction between concepts of the domain. Taxonomies are widely utilized to organize ontological knowledge in the domain using a generalization or specialization relationship through which simple, or multiple inheritances could be applied. Axioms are employed to model sentences that are always true. Instances or individuals are used to show elements of a given concept. In general, ontologies are categorized based on abstraction level and the subject of the conceptualization. Different levels of ontologies can be used to guide processes for the extraction of more general or more detailed information. The use of multiple ontologies allows the extraction of information in different stages of a classification (Fonseca et al., 2002). As shown in Figure 2-9, Guarino (1997) classifies ontologies according to their dependence on a specific task or point of view.

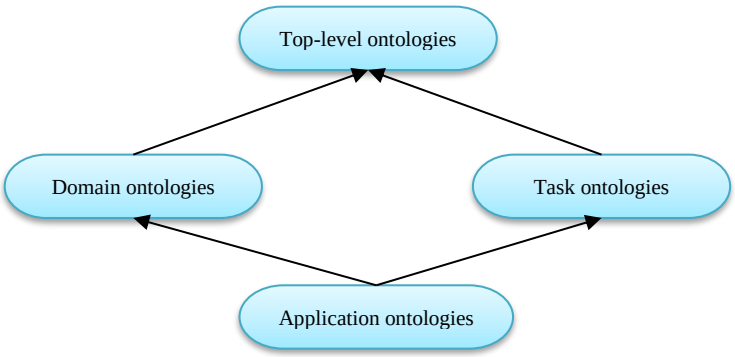


Figure 2-9 Ontology classification by Guarino (1997)

Ontologies can be classified into different types, based on the way they are used. Top-level ontologies offer a comprehensive view of the world by explaining general concepts. On the contrary, domain ontologies provide a specific domain, which represents a part of the world. They describe the vocabulary related to a generic domain and store knowledge about the existing datasets such as relationships between datasets (Valente and Breuker, 1996; Yan et al., 2010). Task ontologies explain the vocabulary related to a generic task or activity by specializing the concepts of top-level ontologies and they may involve more than one domain. Application ontologies describe concepts depending on both a particular domain and a task, and are usually a specialization of both the domain and the task. They are suitable to be used directly in reasoning engines. The remainder of this chapter briefly summarizes the state of the art of movement studies tailored to the scope of the research.

2.5 Movement Research in GIS

Generally, analyses of movement and related features are essential elements in various real world applications such as behavioral ecology of animals (Hunter, 2007; Nathan et al., 2008; Shiode et al., 2002), environmental hazards (Elsner and Kara, 1999; Sinha and Mark, 2005), traffic management (Renso et al., 2013), human mobility studies (Gonzalez et al., 2008; Miller, 2005; Mouza and Rigaux, 2005) and modelling and simulation of movement data (Galton, 2005).

Moreover, the importance of movement analysis has attracted a range of studies in GIS and related disciplines including investigation of space-time paths (Miller, 2005), modelling moving objects and their collective dynamics (Galton, 2005; Kraak, 2003; Miller, 2005), development of new analytical methods for movement pattern discovery (Imfeld, 2000; Laube,

2005), (Dodge, 2011), exploratory data analysis, and visual analytic techniques for movement (Baglioni et al., 2009a; Güting and Schneider, 2005; Miller and Han, 2009; Pelekis et al., 2006b). As seen in Figure 2-10, there are different movement research in this area, knowledge discovery, ontology, and modelling of movement data are the important topics in this work.

Figure 2-10 Interdisciplinary research on movement

The database community has focused on the definition of spatiotemporal data types, in particular, moving point and moving region data types for modelling and querying moving objects. Güting et al (2005) paid significant attention to moving object databases by proposing different data types and operations for querying objects on road networks. Pelekis et al (2006) developed the HERMES prototype to exploit spatial data types in Oracle. Moreover, Parent et al. (2006) proposed a multi-dimensional conceptual data model MADS (Modelling of Application Data with Spatiotemporal features), which provides a thorough and shareable infrastructure for spatiotemporal applications. MADS conceptual model integrates a wide span of topics in temporal and spatial database technologies. The MADS can be taken as an object plus relationship spatiotemporal conceptual data model. However, most research has focused on geometric and temporal characteristics of trajectory data, and the main problem has been the difficulty of correlating extracted patterns with movement behavior (Baglioni et al., 2009a). From

aforementioned discussions, methods used for spatiotemporal systems have already explored spatial and temporal semantics. Although these methods provide some support on modelling spatiotemporal dimensions, they are still missing the high-level conceptual model for moving objects and trajectories.

2.5.1 Semantic trajectory modelling in GIS

As a data type based model approach, the research presented in Zheni et al. (2009) introduced an algebraic model that represents a spatiotemporal trajectory as an ADT, encapsulating dynamic and semantic features. The ADT was designed in a way that if it got integrated in any database management system, it acquired the same status as built-in data structures. It also supported operations covering its spatiotemporal and semantic properties. However, a conceptual model supporting the various requirements of the applications of semantic trajectories was still needed. The desired requirements for such a model cover the characterization of trajectories with attributes, semantic and topological constrains and links to application objects. To fill this gap, the authors in (Marketos et al., 2008) introduced dedicated data types. They brought the minimal information common to all trajectories like the begin, end, moves, stops, as well as their sample points and interpolation functions, and encapsulated them in a generic data type. Therefore, the application specific information that cannot be encapsulated in the generic data type was modeled explicitly using dedicated data types. Those data types contain attributes representing the travelling object or its trajectories and have relationships linking them to the application objects.

In Zhou et al. (2004), the authors introduced a design pattern based model relying on the MADS model (Parent et al., 2006). The model aims at the explicit representation of trajectories and their components (such as stops, moves, begin, and end) as object types in the database schema and linking those components with application objects. Moreover, design pattern based modelling has been significantly adopted in the GeoPKDD project, where many mining techniques have used this stop-move concepts, for example: enriching trajectories with semantic geographical information (Alvares et al., 2007; Wachowicz et al., 2011), a clustering-based approach for discovering interesting places in trajectories (Palma et al., 2008), analyzing trajectories using background information (Kuijpers et al., 2009), aggregation languages for moving object and places of interest (Gómez et al., 2008). The proposed model in Yan et al. (2011) encapsulated both the geometry and semantics of mobility data into three different models: namely, raw data model,

conceptual model, and semantic model. They also introduced a computational platform for the progressive construction and evolution of those three models.

One example of a model based on ontologies is represented in Yan and Spaccapietra (2009), where the authors analyzed modelling requirements for trajectory modelling and proposed a trajectory model. In this approach, they used different ontologies to add semantics to the trajectories. In Wannous et al. (2013a), a case study was presented on the use of an ontological-based approach for modelling semantic trajectories. The modelling approach is based on two main components: domain ontology and time ontology. The ontologies represent basic domain and time concepts for the application and show the relationships between them. Along with the ontologies, rules were defined to understand temporal relationships. Similarly, in Amine et al. (2013) and Zhixian (2011), the authors presented an ontological approach for modelling semantic trajectories, which integrated domain ontologies with spatial ontologies to answer queries based on spatial relationships.

Choosing the right modelling approach for semantic trajectories depends on several factors. Among them is the application used, the availability of the domain's ontology, the level of trajectory abstraction required, and the extent of intervention required by the database designers. Data type based models are generic models that fit into a wide range of applications. They can be made persistent by extending a database model, and can be queried by extending Structured Query Language (SQL). Data type modelling approaches alone are not sufficient to support the semantic trajectories application requirements. This is due to the inefficiency of using a generic data type for all application domains. Design pattern based models are even more generic than the data type based models, as they are not restricted to a specific data type. Instead, a dedicated type relevant to the application in hand can be added to the generic data types but will need the help of a database designer. Therefore, this model requires from the designer to add the semantic information specific to the application. The model provides the designer with a predefined sub-schema that supports the basic data structures for data modelling. On the other hand, ontology-based models are application specific, as the ontology needs to reflect the application domain. They can represent richer semantic information, and involve any kind of semantic annotations. In contrast to data type models, ontological models are naturally extensible because ontologies are designed to extend.

2.5.2 Trajectory segmentation and annotation

Ashbrook et al. (2003) identified stops from a car trajectory where either there is an absence of signal, for instance, the GPS is off, or the velocity is zero for a given temporal interval. They annotated stops with the POIs within a given radius defining a spatial buffer around the stop. The use of the buffer relied on the assumption that the GPS signal is subject to measurement errors. Krumm and Horvitz (2006) used some criteria to split a car movement track into trajectories. These criteria could similarly identify stops within a trajectory; first, a five-minute gap indicating that the car was not moving, and second, at least five minutes of very low speed, suggesting car stillness while the GPS has kept sending signals giving the same position with minor deviations to noise. The approach of Andrienko et al. (2007) used a longer temporal threshold (i.e., 2 hours) to classify the important places visited by a person. They annotated stops with the POI that corresponds to the places visited by the user of the moving car. They determined the POI type by examining the temporal profiles of visits and using background map and knowledge about the area. Zheng et al. (2011) detected stops as sequences of consecutive GPS positions when their spatial distance is below a threshold while the temporal duration is above another threshold. Moreno et al. (2014) proposed a hierarchical approach, which combines clustering and classification tasks to detect important parts of trajectories. They used different context information to enrich the detected parts. Alvares et al. (2007) presented SMoT to identify stops by using a combination of raw data, contextual geographic information and application information. In their work, a stop was a position where a trajectory holds on for a certain time duration and that matches to the position of one of the Regions of Interests (ROIs) defined by the application. Moreover, Cao et al. (2010) and Guc et al. (2008) used the same approach to extract stops. But, they annotated stops from pre-encoded POIs crossing the moving object trajectory. Another criterion to find stops was the change of direction. For example, Rocha et al. (2010) characterized a stop for fishing boats by a sudden change of direction in contrast to the move where the boat goes in a nearly straight line.

Meanwhile, Griffin and Huang (2008) used a density based clustering technique to detect clusters of GPS points, and then combined them to activity spots. Also, Palma et al. (2008) proposed CB-SMoT to calculate stops based on the speed variation of trajectories. The stops were those parts of the trajectory that the speed was lower than the average speed of the trajectory. They also annotated a stop with the POI located at the stop. The stop episodes that do not cross any POI are annotated as unknown stops. Xie et al. (2009) used an enhanced geographical context to

segment the trajectories based on the proximity to POIs. They described an episode as part of the trajectory whose positions are "influenced" by a given POI. A POI influences a position of the trajectory if the distance between the POI and the position is smaller than the distance between the position and any other POI. Therefore, each episode is annotated with the corresponding POI.

Recently, Buchin et al. (2010) described a theoretic framework, which uses different criteria such as speed, direction and location to compute optimal segmentation. However, there was no experimental study to validate such a framework. Spinsanti et al. (2010) proposed another method that relies on additional information about the POIs including common sense constraints. They annotated each stop with the list of probable POIs visited by the moving object. The POIs were designated based on two different criteria. Some researchers such as Furletti et al. (2013) used gravity model in their algorithm as a probability measure to annotate stops with the POI category types. They measured the probability for each POI category type using the minimum distance of the POIs around each stop. Gravity models are used in various areas such transportation planning (Kincses and Tóth, 2014), trip distribution (Baldauf et al., 2011), and analyzing trade activities (Binh et al., 2011) to predict and describe certain behaviors that mimic gravitational interaction as described in Newton's law of gravity.

Instead of concentrating on one method, Yan et al. (2011, 2013) presented a computing platform, which supports the segmentation of the trajectories. Stops and moves can be computed by considering some spatiotemporal criteria like position density, velocity, and direction. They proposed a framework named SeMiTri, which is able to annotate episodes using POI, ROI, and Lines of Interest (LOI) as depicted in Figure 2-11. ROIs and POIs used to annotate stops and LOIs used to annotate moves.

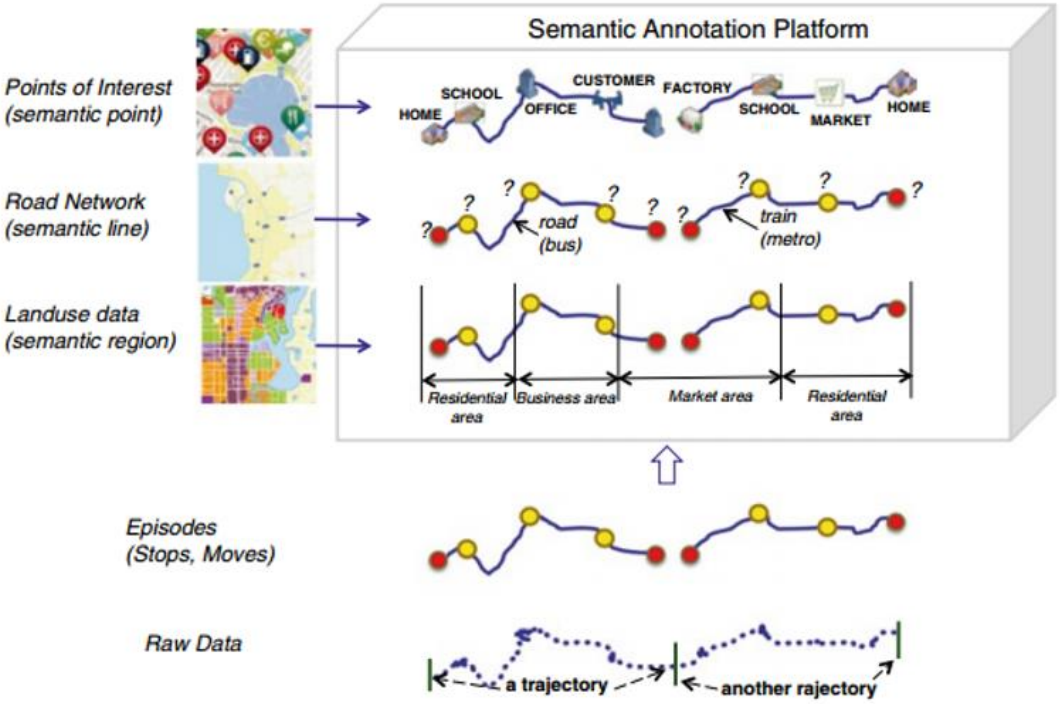


Figure 2-11 Trajectory annotation platform in SeMiTri framework (2013)

Liao et al. (2006a) used a Gaussian mixture model based on the speed, which was categorized into three speed ranges such as walking, high speed and low speed. Therefore, a new segment was created when a switch in the speed range was detected. A similar approach was developed in Zheng et al. (2010). The authors used speed, acceleration and speed change rate to detect the positions where the movement switches between walking and non-walking. Next, they refined the non-walking segments into segments characterized by the other transportation modes such as biking, busing, and driving. They used a combination of techniques, from supervised learning to decision tree inference, and added a post-processing step, which relied on some common sense constraints about typical human behaviors to improve the accuracy of the process. Beyond using raw data and its derived features such as velocity and acceleration, the approach in Yan (2011) used also contextual data such as road categories and public transport networks. For example, using a bike path specifies that the transportation mode is either biking or walking, definitely not driving a car, busing, or travelling by train.

Some researchers used ontology in their works. For example, Baglioni et al. (2009b, 2008) represented annotated trajectories in an ontology including geographical and application domain knowledge. Different kinds of stops were considered, and temporal knowledge was used

to discriminate among them. Similarly, Yan et al. (2008) used an ontological approach for the representation of semantic trajectory. They defined different ontology modules for representing geometry, geography, and application domain. They tested their approach with traffic management application. Based on space time ontology and events approach, Boulmakoul et al. (2012) proposed a generic meta-model for trajectories of moving objects to allow independent applications processing trajectories data to benefit from a high level of interoperability, information sharing as well as an efficient answer for a wide range of complex trajectory queries. The work of Wannous et al. (2013a) was a case of moves annotation for animal trajectories, specifically "seals", to distinguish states such as travelling, resting, and foraging. They used ontologies to integrate the time knowledge to reason different travelling states, which were distinguished on duration and were defined in terms of temporal axioms.

2.5.3 Activity recognition

Most of the approaches for activity recognition are data-driven (Huang and Li, 2010; Liao et al., 2006a; Patterson et al., 2003). Activity recognition studies usually use spatial distance based and statistics based methods. The basic idea of spatial distance based method is to assign closest POIs to the stops where activities have happened. Bohte et al. (2009) defined rules to restrict the candidate POIs and possible activity types, then detected the POIs with the smallest distance to activity centers in trajectories. Xie et al. (2009) proposed influence and influence duration of POIs on trajectories. They constructed a Voronoi diagram using POIs as Voronoi sites and define each Voronoi cell as the POI's influence region. This method actually selects the POIs closest to the polyline geometry of the trajectory.

Some researchers extracted regular activity patterns in trajectories to identify the activity types. Huang and Li (2010) introduced a multivariate analysis approach to identify activities using vehicles trajectories. Time constraints, network distance, activity chains, and POIs were used as four inputs. Scores for candidate POIs selected around the locations where vehicles stopped were calculated using a neural network framework. Some works (Liao et al., 2006a; Patterson et al., 2003) used Markov networks to classify activities into six pre-defined types. The technique was an extension of Markov networks for sequence matching. Moreover, Liao et al. in (2007, 2006b) proposed the machine learning and probabilistic reasoning methods in particular conditional random field method to identify daily activities from GPS data. Yuan et al. (Yuan et al., 2012)

proposed an approach to understand users' trajectories in order to have a better understanding about the complexity of a metropolitan area. They used domain data such as POIs and road network topology together to define functional regions.

Studies like Alvares et al. (2007), Xie et al. (2009), and Moreno et al. (2010) designed relevant spatiotemporal join method to infer activities from trajectories, by computing the topological relationships between the trajectory data and a small set of predefined activity hotspots, together with the time constraints. Spinsanti et al. (2010) proposed a series of rules to detect overall activity of the trajectory. Considering a trajectory whose stops are annotated with possible activities, the authors annotated the trajectory with the most probable global activity. A similar approach was proposed by Renso et al. (2013). This work annotated trajectories with different activities such as tourist, home-to-work commute. Montini et al. (2014) proposed a machine learning approach, random forests, to improve trip purpose identification. The analysis was based on GPS data and accelerometer data. Miller et al. (2015) designed an activity based travel demand model in order to constitute the fundamental behavior of people. They discussed the advantages of the disaggregate activity based approach to travel demand modeling.

The majority of the mentioned activity recognition works still have unresolved questions. They only discover stops and moves of moving objects and annotate them with contextual data. However, by only doing that, it can still be difficult to identify the activity type of the moving object. For instance, as it can be seen in Figure 2-12, there are two stops identified for the same moving object; one on a residential land use and the other on a commercial land use. However, it is difficult to determine whether the residential stop is the residence of the moving object, or if the commercial stop is a work place. The moving object could for example be visiting a friend, or shopping. To draw stronger inferences, one needs to extract additional information and semantic features such as stop duration, stop frequency, and stop begin time from the trajectory data.

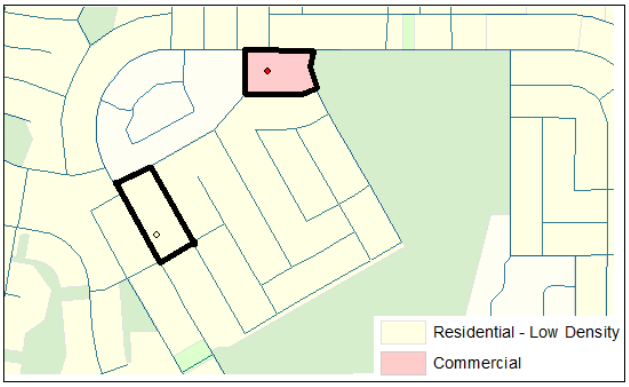


Figure 2-12 Residential and commercial land use types as two user stops

2.5.4 Behavior modelling

Several works used knowledge discovery techniques to identify behaviors from trajectories. They can be classified to general characteristic of movement data (Sub-section 2.5.4.1), ontology and knowledge discovery (Sub-section 2.5.4.2), and semantic based trajectory knowledge discovery (Sub-section 2.5.4.3).

2.5.4.1 General characteristic of movement data

General characteristic for trajectory knowledge discovery includes trajectory classification, trajectory sequential, association rule mining, and trajectory clustering patterns.

2.5.4.1.1 Trajectory classification pattern

A number of studies applied classification techniques for modelling moving object trajectories in imagery and video surveillance databases (Fraile and Maybank, 1998), and behavior studies of individuals (Bay and Pazzani, 2001; Blythe et al., 1996). Many machine learning techniques (in particular the Hidden Markov Model) have been applied in trajectory classification (Sbalzariniy et al., 2002) for biological trajectories, (Bashir et al., 2007; Jeung et al., 2007; Nascimento et al., 2010) for human motion trajectories using unsupervised learning methods on the coefficients of the basic functions. Lee et al. (2008) applied classification to group trajectories of ships into three predefined classes: trajectories of fishing ships, trajectories of passenger ships, and trajectories of cargo ships. The authors work on trajectories partitioned into segments. After using clustering to find the interesting segments, they defined a class for each cluster and apply classification to populate the classes.

2.5.4.1.2 Trajectory sequential pattern

The discovery of trajectory sequential patterns can show the cumulative and consecutive behavior of moving objects and help understanding the mobility-relevant repeated sub-sequences (Giannotti et al., 2007; Giannotti and Pedreschi, 2008). A couple of extensions on traditional sequential mining algorithms such as GSP (Generalized Sequential Patterns) (Srikant and Agrawal, 1996), PrefixSpan (Pei et al., 2001), T-pattern (Giannotti et al., 2007), and SPADE (Sequential PAttern Discovery using Equivalent Class) (Zaki, 2001) have been proposed for trajectory sequential mining (Cao et al., 2007; Gidófalvi and Pedersen, 2009; Han et al., 1999; Tsoukatos and Gunopulos, 2001). Among these studies, T-pattern is being widely cited for mining frequent movement behaviors. Both space and time are considered in the T-pattern approach. Giannotti et al. (2007) additionally designed a tree hierarchy based on the connection of many T-Pattens, and provide the functionality of location prediction by using the tree structure.

Another interesting sequential study on trajectory is mining periodic behaviors of moving objects (Cao et al., 2007; Li et al., 2010). Periodic pattern can be roughly defined as the repeating activities happened in certain locations with regular time interval at (between) periodic time instances; therefore, it can be significantly used for location and behavior prediction. For example, Li et al. (2010) designed a two stage algorithm called “Periodica” for mining such periodic trajectory patterns. Their basic assumption was that periodic behaviors are associated with a corresponding frequently visited ROI. Firstly, the periods were detected by reference spots using Fourier transform and autocorrelation. Secondly, hierarchical-based clustering used to summarize periodic behaviors. Giannotti et al. (2007) pursued the same goal, recognizing sequence behaviors, but the trajectories they considered do not need to have the same duration and belong to different moving objects. The algorithm they proposed for extracting sequence behaviors from groups of trajectories computes a sequence of regions, frequently visited in a specified order and with similar transition times. Interesting regions are computed using density-based clustering techniques of trajectory positions in space, not in time. Having detected the ROIs, the method filters the trajectories that intersect the regions and computes the travel time in each region. Therefore, a sequence behavior is made if the sequence of regions is visited by at least a minimal amount of trajectories.

2.5.4.1.3 Association rule mining

Association rule mining is a powerful method in the data mining domain (Shen et al., 2006). Apriori (Agrawal and Srikant, 1994) as one of the association rules mining methods, has been used for mining association rules. Several works have focused on this challenge resulting in rule pruning (Toivonen et al., 1995), grouping (An et al., 2003), clustering and rule visualization (Shen et al., 2006) extensions to the original method. The most critical method used to resolve this problem was the association rule retrieval method (An et al., 2003; Morzy and Zakrzewicz, 1998; Tuzhilin and Adomavicius, 2002; Tuzhilin and Liu, 2002). For instance, Hipp et al. (2002) used a rule cache to postpone some operations to reduce the retrieval time of association rules. Tuzhilin and Adomavicius (2002) proposed the use of several rule evaluation operators to filter and browse data, and data inspection operators that simplify reduction of a large numbers of rules to achieve useful interpretable results. Appice et al., (2005) reduced the number of rules by considering user specified pattern constraints, which were inefficient, since pattern constraints are used in post processing steps. Generally, these methods have considered low level information in association rules mining and ignored semantic knowledge of data in the discovery process (Shen et al., 2006). Therefore, it might be difficult for the data miner to analyze all rules to discover if they are really interesting or not.

2.5.4.1.4 Trajectory clustering pattern

A large number of trajectory clustering approaches for discovering similar trajectories or dense regions have been proposed (Fu et al., 2005; Giannotti et al., 2008; Miller and Han, 2009; Rinzivillo et al., 2008). Many studies have focused on extending the well-known clustering algorithms and applying them in trajectory data, such as k-means (Lloyd, 1982), BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) (Zhang et al., 1996), DBSCAN (Density Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999), and STING (STatistical IN-formation Grid based clustering method) (Wang et al., 1997).

Pelekis et al. (2012) defined a method that groups trajectories using various distance functions based on motion properties such as spatial location, speed, acceleration and direction. The method evaluated distances over the entire trajectories. In contrast, Lee et al. (2007) proposed a clustering algorithm, which looks for similarity between segments of trajectories. It didn't require

a similarity hold for the entire trajectory. The advantage of using segment similarity is the possibility to fine-tune the grouping of trajectories. In (Han et al., 2012) the authors discovered clusters of segments of car trajectories data by considering sequences of nearby road segments that are followed by continuous traffic flows. Clustering was also used by Cao et al. (2010) for extracting semantic locations and also by Palma et al. (2008), who adopted spatiotemporal clustering to classify trajectory with respect to their speed. Moreover, trajectory clustering has been used in several contexts as trajectory searching and querying (Panagiotakis et al., 2012), trajectory visualization (Rinzivillo et al., 2008), and processing of trajectory uncertainty (Pelekis et al., 2011). Therefore, most research in this area has focused on geometric and temporal characteristics of movement data for behavior modelling, where the main problem has been the difficulty of correlating extracted patterns with movement behavior. Classical data mining methods focus on the mining step itself, and usually little attention is given to the entire knowledge discovery process. In addition, most of the proposed GKD algorithms have been concentrated on mining the geometric properties of trajectories.

2.5.4.2 Ontology and knowledge discovery

Several works have demonstrated the effectiveness of using ontologies for supporting the knowledge discovery process (Charest and Delisle, 2006; Nigro et al., 2008; Wang et al., 2010). Bernstein et al. (2002) proposed an intelligent data mining ontology assistant to discover an appropriate ranking for data mining processes. Phillips and Buchanan (2001) have used ontologies to conduct a feature selection step during the knowledge discovery process. Bauer and Baldes (2005) used an ontology based interface to help non-expert users understand a machine learning system. Canataro and Camito (2003) demonstrated the use of a data mining ontology in the area of grid computing to simplify the progress of distributed knowledge discovery applications. Moreover, data mining and ontologies have also been applied to define some constraints to filter discovered patterns (Bellandi et al., 2007). Bittner and Winter (1999) also identify a role for ontologies in the modeling of spatial uncertainty. Chaves et al. (2005) defined a meta model, named Geographic Knowledge Base (GKB) to define an ontology for geographic data, where spatial integrity constraints were represented by properties of geographic data.

However, they mostly focused on data dictionaries, data interoperability, and data presentation and none of them have considered an ontology model that encompasses space, time,

and application domain in the knowledge discovery process. Therefore, an ontology model that has integrated analysis from multiple forms of information and use of explicit knowledge is required. Such an ontology should enable discovery of complex relations among entities and also enable meaningful interpretation of multimodal information across different domains as they relate geospatially.

2.5.4.3 Semantic based trajectory knowledge discovery

Bogorny et al. (2010, 2009) formalized the idea of semantic trajectory pattern mining to boost data preprocessing and to mine data at a higher abstraction level. They presented a Semantic Trajectory Data Mining Query Language (ST-DMQL), which allows users to identify the process of semantic enrichment of trajectories with domain knowledge. Three different kinds of semantic patterns such as discovery of frequent and sequential patterns and association rules from trajectories could be mined in this work. Another tool that analysed semantic trajectories to infer behavior was M-ATLAS (Giannotti et al., 2011). It offers support for raw and semantic trajectories, and can easily integrate with different mining algorithms such as clustering and classification techniques. The algorithms can be merged using a data mining query language, which allows to discover different behavior types. Moreover, Baglioni et al. (2013) integrated this tool with the Athena system, which enables trajectory annotation with behaviors such as home-to-work, general commute or tourist sightseeing.

Lucca Siqueria and Bogorny (2011) defined chasing behavior and provided an algorithm that evaluates if an individual called the stalker intentionally follows another individual called the target. The stalker must follow the target for a certain time period, and during this period, the movement of the two individuals must remain with similar speed and direction. Moreover, the target must always be in front of the stalker. Alvares et al. (2011) defined avoidance behavior and present an algorithm for identifying the trajectories that avoids a static object. For example, when analyzing human trajectories an avoidance of street cameras may reveal a suspect behavior. A trajectory shows an avoidance behavior when it moves towards a target geographic object, turns around this object without intersecting it, and after avoiding the target object, the trajectory returns to its original path. Therefore, the main objective of this thesis is to propose an ontology based semantic knowledge discovery framework to better understand mobility data.

2.6 Summary

In this chapter the background and related works as the context of this thesis are reviewed. A literature review is presented related to movement research in GIS, semantic trajectory conceptual model, knowledge discovery for movement data and ontology based models. Later, all works related to activity and behavior modelling are described. Different methods to study the high-level movement behaviors for understanding mobility data are explained. Generally, most of the methods used in the classical data mining are focused on the mining step itself and just a few researchers have considered some contextual data as semantic information in their research to help users understand trajectory patterns. However, they only discover stops and moves of moving objects, which still, it can be difficult to identify the activity type of the moving object. In the next chapter, the methodology that has been used in this research is presented.

CHAPTER THREE: METHODOLOGY

3.1 Introduction

This chapter consists of two main parts. In the first part (Section 3.2), a proposed conceptual data model that has been used in this research is described. Some definitions regarding moving objects and their components are provided to illustrate how moving objects can be semantically interpreted. In the second part (Section 3.3), the framework is described, which consists of three different steps, namely: semantic trajectory ontology modelling, activity recognition, and semantic behavior modelling. In the first step an ontology model is built based on the proposed conceptual data model. In the activity recognition step, activity types are extracted from the given raw trajectories and semantic trajectories are computed. In this regard, various activity types are defined by ontology axioms, i.e. logical expressions built over semantic features. In the semantic behavior modelling step, data mining algorithm, *apriori*, is executed on the semantic trajectories in order to extract semantic behavior patterns.

3.2 A Conceptual Data Model for Semantic Trajectories

The essential part of this research is a conceptual data model of semantic trajectories, which is illustrated in Figure 3-1. This model is an extended version of the conceptual model (the green colored boundary) introduced in Spaccapietra et al.(2008), which relies on the conceptualization of stops and moves in trajectories. The conceptual model contains information related to: moving object, raw trajectory, sub-trajectory, semantic sub-trajectory, semantic trajectory, semantic place, stop, move, activity type, and behavior type. As explained earlier in Chapter 2, a moving object generates raw trajectory, which based on different criteria it can be divided into sub-trajectories. Giving meaning to the sub-trajectories, semantic trajectory is a combination of different semantic sub-trajectories. Each sub-trajectory is composed by stops and moves. Each stop, as shown in the figure, could have different semantic features such as stop frequency, average duration, stop type, and other features, which can help to infer the type of activity that occurs at the stop. Behavior type is a combination of different activities and their attributes. In the following some definitions are provided.

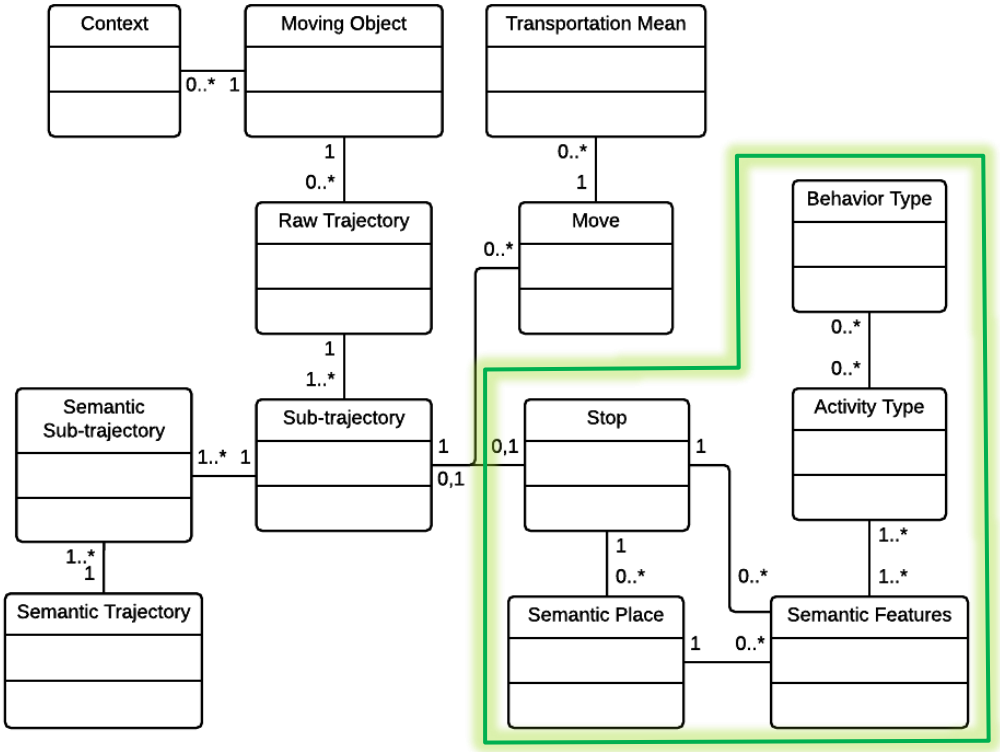


Figure 3-1 Extended conceptual model of semantic trajectory used in this research

Definition 8 (Semantic place): A semantic place is a set of positions where a stop is located. It includes land use type and POI type, which cover environmental information related to the stop.

Definition 9 (Semantic features): Semantic features are extracted from the stops and are divided into five different types such as stop frequency, average duration, stop land use type, stop POI category type, and start begin time, which can help to infer the type of activity that occurs at the stop.

Definition 10 (Activity type): An activity is what the moving objects are going to do during their movement. In other words, it is the objective of the movement, which has a start time and an end time, and it can be relative to the entire trajectory or part of the trajectory (the semantic sub-trajectory). Activities can be represented as taxonomy, from the more specific to the more general. Figure 3-2 shows an activity classification of a user. It is classified into four different major activity types such as recreation, profession, shopping, and other activity types.

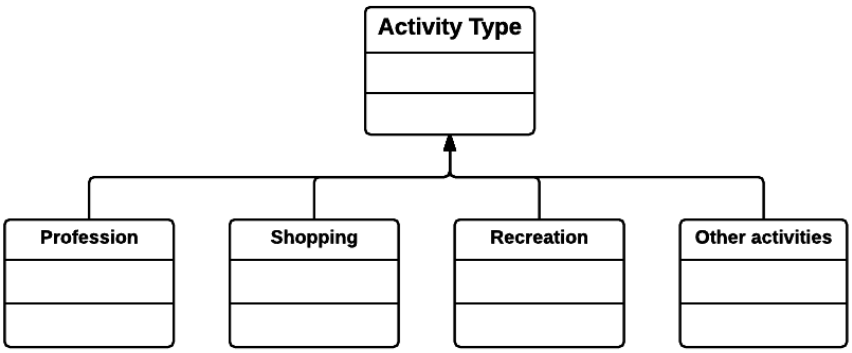


Figure 3-2 Activity types classification

Profession activity types could be working, getting involved into different jobs, etc. Shopping refers to time spent at different stores for buying food and drinks, or groceries in general required for one's daily needs. Recreation might include going to theatres, pubs, gyms, and other places related to leisure. Other activity type might include relaxing at home, cultural or religious activities, etc.

The inference of the activity types is based on the semantic features. Activity types are highly influenced by users' location. For instance, if a person is close to a university, the most probable activity types would be studying, teaching, or working. In order to capture this dependency, first it is needed to model which activity types can be performed or hosted within or nearby every place (e.g., eating is possible in a restaurant, while shopping is possible in a mall). Therefore, there is an association between places and activity types and according to the conceptual model, an activity is typically performed in a place.

As shown in Figure 3-3, the restaurant is both a place for eating and a working place expressing the fact that a restaurant may be a kind of work place for people working there (the cook, the waiter etc.), or a place to represent the fact that typically restaurants are attended by people for dining. Therefore, activity types are also correlated to time and, in particular, to the time of day and the duration that a user spends at each place. Different activity types might have different timetables and durations.

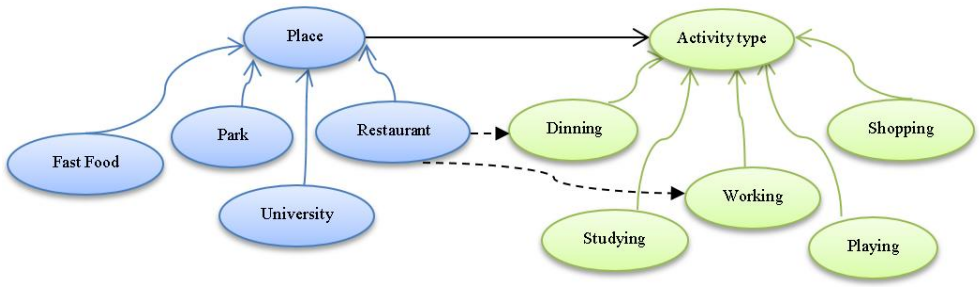


Figure 3-3 Association between places and predefined activities

For instance, if the place is a restaurant, different time periods may be interpreted as different activity types. For instance, the period of 15 to 30 minutes would be interpreted as a delivery, since it is not enough time to stay in and eat. If the time period is between 30 minutes and 3 hours, it would be interpreted as dining and if the time period is between 3 hours and 8 hours, it would be interpreted as working.

Also, stop frequency and average duration are important features to find out the type of activity (Mousavi and Hunter, 2012a). For instance, a moving object is considered a worker only if he visits the same region at least four times a week. As a general example, if the stop frequency is more than five days a week and the average duration is more than eight hours, it could infer that the place would be either where the person works or lives. Therefore, this research hypothesizes a functional relationship (3-1) for Activity Types (AT) based on the semantic features as shown below.

$$AT = f(P, L, S_f, T_b, S_d) \qquad (3\text{-}1)$$

Where:

- P is the POI type that is around the stop
- L is the land use type where the stop has occurred
- S_f is the frequency of the stop in a week
- T_b is the begin time of the stop in the place
- S_d is the average duration of the stop in a week

This research further hypothesise that having an understanding of these features; activity types can be inferred through the definition of a set of IF-THEN criteria. For instance, for a specific

stop, if the land use type is residential, the POI type is null, the begin time is evening, stop frequency per week is more than six and the average duration is more than ten hours per week then the moving object is 'spending time at home', i.e., AT= Return Home. At this step some rules can be defined on the captured data in order to extract different activity types.

Definition 11 (Behavior type): A behavior type may be defined and inferred based on the spatiotemporal characteristics of the trajectories, or can rely on the semantic information, for instance the place the object visits. Therefore, considering the semantics of space (e.g. the place where a trajectory has passed) and the semantics of time (e.g. morning, afternoon or weekday instead of a timestamp) gives more meaning to a behavior. As shown in **Error! Reference source not found.**, a behavior type is composed of different activities and their attributes and it indicates the regularities between users' activity types, since one activity may or may not be followed by another (i.e. "return home" is typically followed by "being at work").

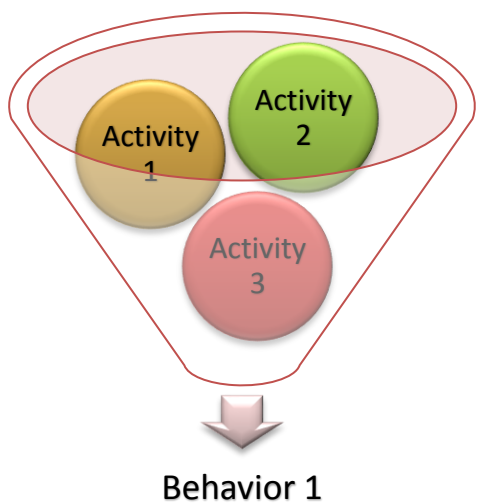


Figure 3-4 A behavior is composed of different activities and their attributes

Therefore, a Behavior Type (BT) is a function of different activity types as shown in equation (3-2).

$$BT = f(AT_1, AT_2, \ldots, AT_n) \qquad (3\text{-}2)$$

The behavior of a trajectory in general, is computed with intelligent methods such as data mining algorithms. Different visions on the features of trajectories lead to different types of their behaviors. As shown in Figure 3-5, in this research, taxonomy of movement behavior is classified into four different types, namely: semantic, semantic and temporal, semantic and spatial, and finally, semantic and spatiotemporal. More information is provided in Sub-section 3.3.3.

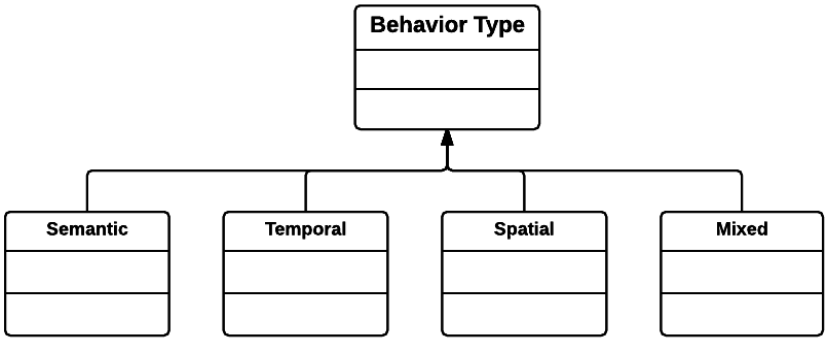


Figure 3-5 Taxonomy of movement behavior type

3.3 The Proposed Framework

This section describes an ontology based semantic knowledge discovery framework for handling semantics of movement patterns in order to interpret users' behavior. As depicted in Figure 3-6, it consists of three different steps: namely, semantic trajectory ontology modelling, activity recognition, and semantic behavior modelling. The first step, consists of raw trajectory data, maps and different geographical layers. A Semantic Trajectory Ontology Model (STOM), which is the essential part of the framework is built based on the proposed semantic conceptual model, which was presented in Section 3.2.

In the activity recognition step, which is the heart of the framework, first, the raw trajectory data is cleaned and trajectories are reconstructed. In the semantic enrichment process, stops are identified and they are annotated with probable visited places. Next, some semantic features are extracted from the annotated stops. Then, the ontology model is populated with the extracted features and the predefined axioms. The ontology inference engine is executed and the axioms are interpreted to classify the ontology instances using the appropriate concepts on the activity types. Once the activity types are extracted, semantic trajectories are created.

In the semantic behavior modelling step, data is preprocessed and different features are selected from the semantic trajectories. In this research, association rule mining is used to generate semantic behavior patterns as IF-THEN rules from the semantic trajectories. Finally, the rules are applied into the STOM. Each of the steps is explained in more details in the following sections.

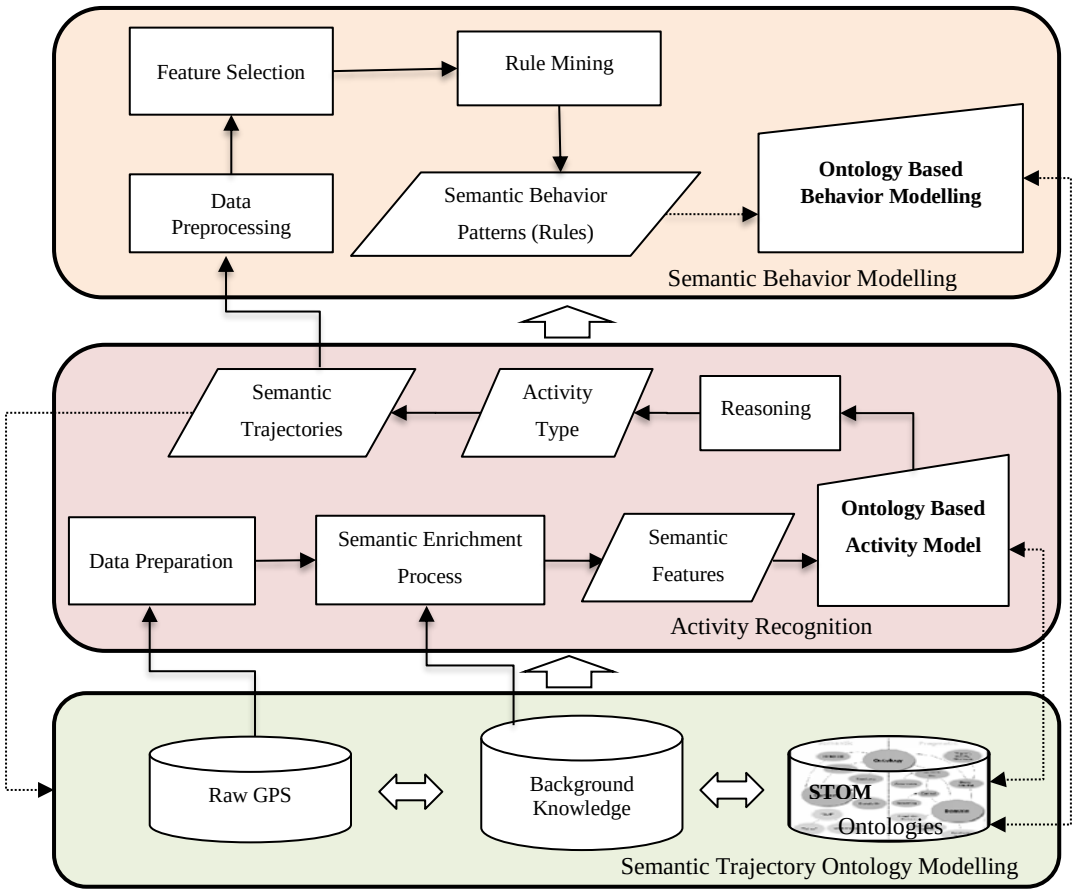


Figure 3-6 The proposed framework to model users' behavior

3.3.1 Semantic trajectory ontology modelling

Semantic trajectory ontology modelling step is used to manage raw GPS data, maps/layers, mined patterns, and ontologies. In this research, an application was developed to gather user's GPS data (see Chapter 4). The data consists of the raw (x, y, t) observations, and other relevant attributes such as speed, and direction. The background knowledge database includes land use, road network, and POI layers. Moreover, various data sources can be specified using expert domain knowledge

specific to an application needs. Section 3.2 described the conceptual model, which addresses the modelling requirements with the goal of analysis of semantic trajectory data. In this step an ontology model named STOM is built based on the proposed conceptual data model. As shown in Figure 3-7, the ontology model in this research includes geometry, geography, theme, and service ontologies.

The geometry ontology is composed of spatial ontology, temporal ontology, and trajectory ontology models. The geography ontology describes the places where people move through and includes a variety of land use types, road networks, and POIs layer. The theme ontology gathers all the application dependent concepts such as user activity type and behavior type ontologies. Integrating these ontologies together leads to the semantic trajectory ontology model, which helps to the discovery of more semantics on trajectories. These ontologies are integrated into a unique ontology by setting up rules between them. The service ontology focuses on available services. It represents the available services that a system would offer to users.

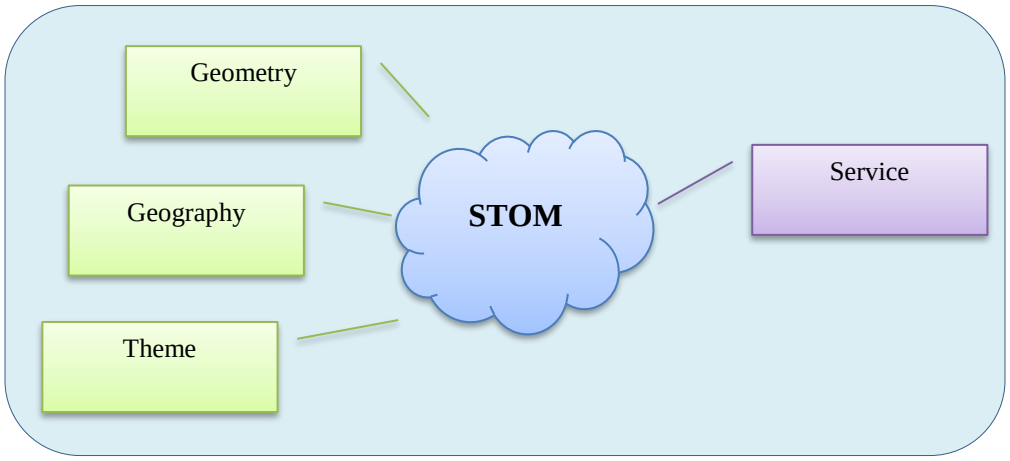


Figure 3-7 The STOM consists of different ontologies

Therefore, the aim of this model is to represent the concepts and relations of the movement domain where trajectory data and semantic movement patterns are to be interpreted along with the behavior types. In the following, each of the ontologies is explained.

3.3.1.1 Geometry ontology

The geometry ontology model includes spatial ontology, temporal ontology, and trajectory ontology models. In the following figures, shapes represent the main concepts whereas arrows represent relationships between two concepts.

3.3.1.1.1 Spatial ontology model

In this research, the OwlOGCSpatial ontology developed by Wannous (2013b) is chosen to model the spatial data of a trajectory. An extract of this ontology is shown in Figure 3-8. The model holds generic concepts for the description of the geometry component of a trajectory. It includes spatial concepts used to specify spatial features needed for a description of application data. The Geometry concept generalizes any kind of simple spatial features, like points, lines, and surfaces. This enables the model to state, the POIs and the land use types are represented as points and areas, respectively. The spatial ontology represents spatial relationships as a set of rules. PostGIS is used to implement these relationships.

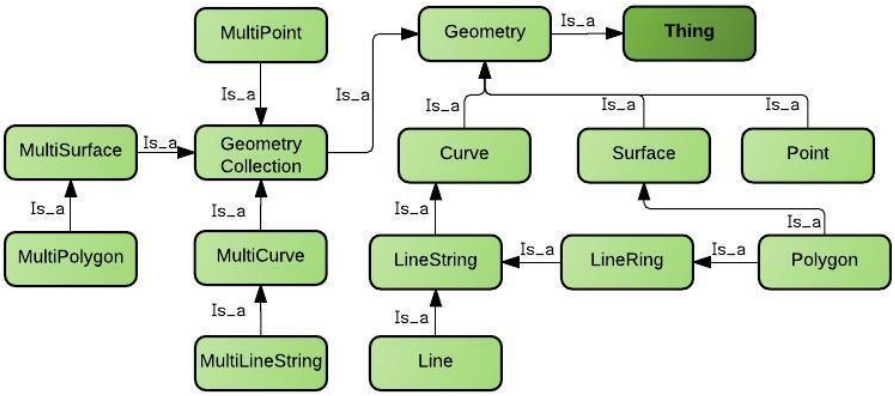


Figure 3-8 View of the OwlOGCSpatial ontology model

Spatial relationships are usually classified as topological, directional, and metric relationships. In this research, topological relationships are considered: Equals, Disjoints, Intersects, Touches, Crosses, Overlaps, Interacts, Within, and Contains. More information is provided in Chapter 4.

3.3.1.1.2 Temporal ontology model

Temporal ontology is another source of information, which integrates time concepts and rules for modelling semantic trajectories. OwlTime ontology (W3C n.d.), which is being developed by the World Wide Web Consortium (W3C) is chosen. An extract of the ontology is shown in Figure 3-9. The main class has several subclasses i.e. TemporalEntity, DurationDescription, TemporalUnit, DayOfWeek, and TimeZone. It includes the concept of a temporal entity, which can be either instantaneous, and is called an instant, or of some duration, and called an interval. The duration of an interval is defined with an instance of the DurationDescription class, and it can be bound by instants with the hasBeginning and hasEnd properties. The DurationDescription instance uses predicates such as years, weeks, days, hours, minutes, and seconds to specify the duration. The DateTimeDescription also allows for the definition of instants. Therefore, this ontology model is able to define temporal discretization such as durations or absolute intervals.

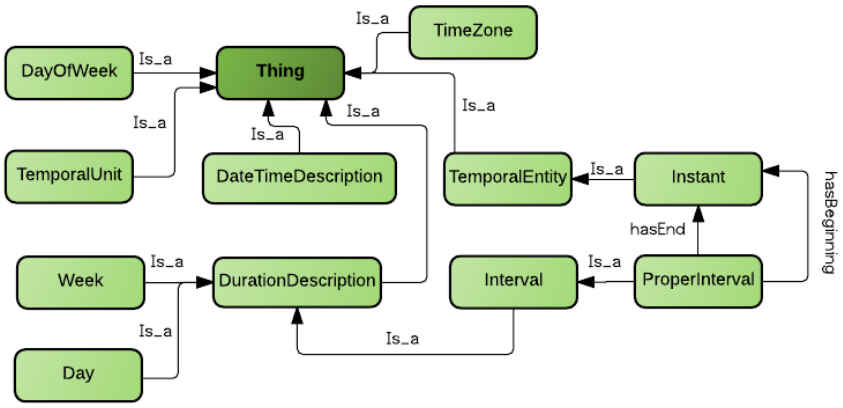


Figure 3-9 OwlTime ontology model

3.3.1.1.3 Trajectory ontology model

Trajectory ontology covers the concepts that help in finding how movement can be identified as a set of structured trajectories. It includes stop and move concepts, which define a segmentation of a trajectory. This set of concepts allow to define a design pattern for structured trajectories and supports linking trajectory elements to application elements, thus allowing semantic enrichment on the trajectory data for providing better understanding. Figure 3-10 shows the trajectory ontology model contains trajectory, stop, and move concepts and their relationships.

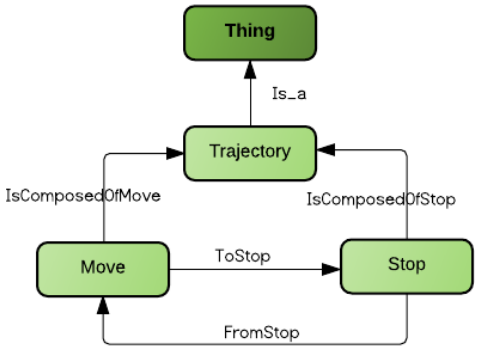


Figure 3-10 Trajectory ontology model

3.3.1.2 Geography ontology

To achieve the semantics of trajectories, it is required to connect the trajectory ontology model concepts into meaningful geographical concepts. The geography ontology holds the concepts about the geographical environment, which participate in the application description. Concepts are likely to include those describing the topography of the land, road networks, land use types, POIs, landmarks, and anything else that is of interest to the application. As seen in Figure 3-11, the geography ontology model contains different types of layers; POI, road network, and land use. The POI represents the specific categories such as shopping center, park, etc. The road network represents the interconnections of different road types designed within urban areas. The land use represents different regions and their utilization such as agricultural, residential, recreational or other purposes. Therefore, this ontology is used to aid potential interpretation of each stop, i.e., why the moving object stopped.

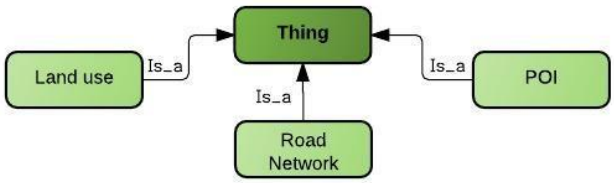


Figure 3-11 Geography ontology model

3.3.1.3 Theme ontology

The theme ontology model gathers a wide range of application dependent concepts. The understanding of trajectories profoundly depends on their relationships to application objects not just the moving object itself. As can be seen in Figure 3-12, the model describes the concepts of

activity type and behavior type that are of interest within a particular application context, which considers users' trajectory.

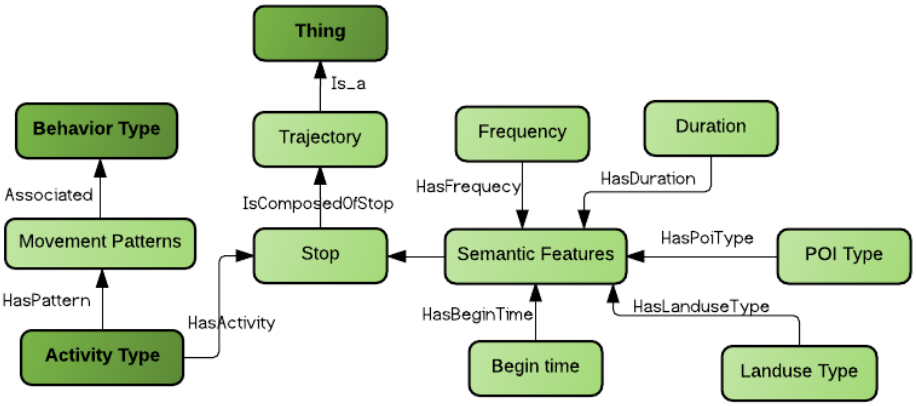


Figure 3-12 Theme ontology model describing activity and behavior type

It includes stop, trajectory, semantic features, movement patterns, activity type, and behavior type concepts. The activity type is composed of stops and their features. The semantic features are composed of begin time, duration, stop frequency, POI type and land use type.

3.3.1.4 Semantic Trajectory Ontology Model (STOM)

Combining the trajectory ontology, geography ontology, and the theme ontology together leads to the STOM. This ontology provides the semantic description of application-relevant trajectories with their domain specific semantic meaning. Figure 3-13 shows a very partial version of this ontology with only the most important concepts and relationships. The necessary concepts have been identified and discussed in the conceptual data model, which specify that spatial, temporal and domain related information is required to enhance trajectories.

Every trajectory is composed of stops and every stop is connected to other stops by moves. Moves can be identified by transportation mean type. Every stop is connected to an interval that represents the time of the stop. It includes begin time and end time concepts that identify when the trajectory starts and ends. Moreover, each stop is linked to a place and an activity type that can be performed in the place in order to enrich the stop entity. For instance, a stop at a restaurant can be associated to a food activity. An activity may in turn be associated to a behavior, which is associated to the trajectory concept, thus representing the fact that a trajectory expresses a behavior

through the performed activities. Movement patterns are composed of different activity types, which are some of the main parameters to identify movement behavior.

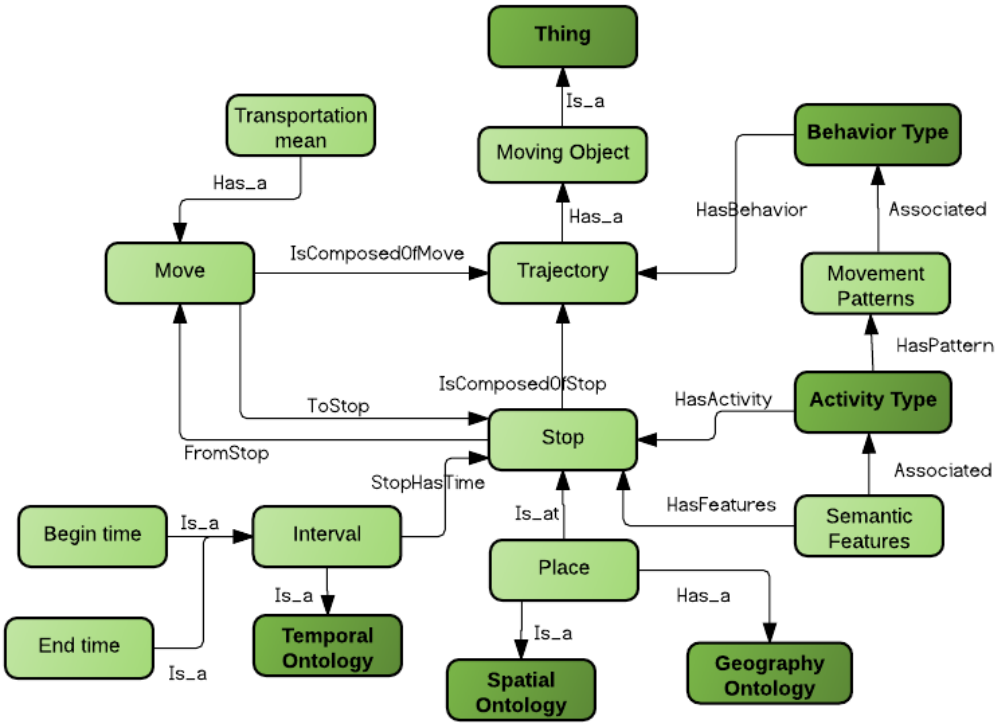


Figure 3-13 Semantic trajectory ontology model

Figure 3-13 presents an extract of this ontology, where:

- Moving object is a user object, who is equipped with an enabled-GPS device;
- Trajectory is a logical form to represent a set of stops and moves;
- Stop is the spatial part in trajectory ontology;
- Move is defined as the maximal subsequence in between two consecutive stops;
- Place is a description of the location that user visited/stopped;
- Transportation mean refers to the type of transportations that objects use to move from one stop to another stop;
- Begin time is the time at which the activity starts. The time that user arrived at the location;
- End time is the time at which the activity is finished. The time that user leaves the location;
- Duration is the amount of time that a user has stayed at the location;

- Activity is the semantic part representing user activity types for a stop;
- Movement patterns are the regularities and common things that happen in the movement data;
- Behavior is a pattern among different activity types.

Besides these concepts, the model defines relationships such as:

-IsComposedOfStop is an object property between a stop and a trajectory;

-IsComposedOfMove is an object property between a move and a trajectory;

-HasPattern is an object property between a trajectory and movement patterns;

-HasDuration is a data property for time that is connected to a stop;

-Associated is an object property between an activity and a behavior;

-HasActivity is an object property between a stop and an activity;

-FromStop is an object property between a stop and a move;

-ToStop is an object property between a stop and a move;

-HasBehavior is an object property between a trajectory and a behavior;

-StopHasTime is an object property between an interval and a stop.

These relationships define connections between different concepts. For instance, the relationship named HasActivity connects a stop concept to an activity concept, which means at each stop some activities might occur.

3.3.1.4.1 STOM structure

Figure 3-14 shows the structure of the STOM, which includes upper and domain ontologies. The upper ontology defines the basic concepts such as time, space, geography, activity type, and behavior type. The domain ontology defines the details of general concepts and their properties in the domain.

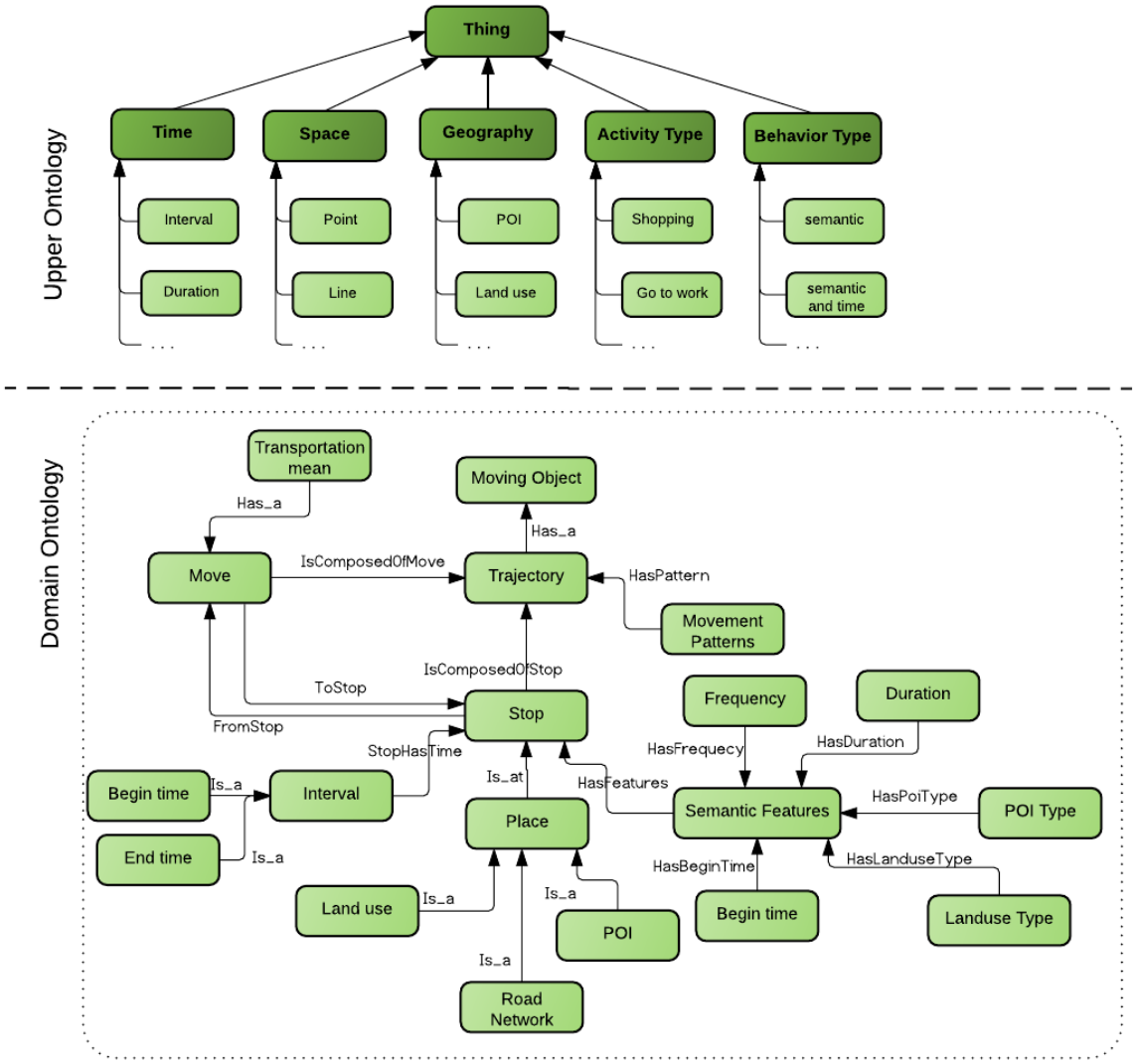


Figure 3-14 The STOM structure

3.3.1.5 Service ontology

Once user behaviors are extracted from their history movement data, it's time to connect the model to service ontology. For example, groups of users can be identified exhibiting similar behavior. These groups can be characterized based on various attributes of the group members or the services they requested. Sequences of service requests made by users can also be analyzed to discover regularities in such sequences. Later these regularities can be exploited to make intelligent predictions about user's future behavior given the requests the user made in the past.

As an example of type of knowledge, the mining patterns might find that if a user with a certain profile goes to work place every day then he/she usually shops at certain days either it could be after work or before work. Considering available services that exist in the system, the user will get relevant services. For instance, as the user is approaching a mall, he/she might receive a message on his/her terminal saying: “there’s a mall in the neighborhood, would you be interested to go shopping?”

As a result, each user receives a service customized to the user’s specific preferences and needs and current situation. As seen in Figure 3-15, there are different services available such as spatial, temporal and spatiotemporal services. The service ontology has different categories and parameters.

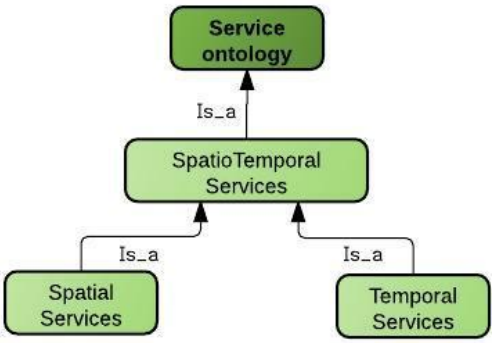


Figure 3-15 Service ontology taxonomy including spatial, temporal and spatiotemporal services

3.3.2 Activity recognition process

The approach presented here aims at enriching people’s movements, represented as trajectories, with semantic information about the activities performed during her/his travel. In this context, a user activity is inferred through the activity based ontology model. As shown in Figure 3-16, the activity recognition process consists of three steps. The first step is the data preparation, where the GPS data are cleaned and daily and weekly basis trajectories are identified. The second step is the semantic enrichment process, which includes stop detection, finding probable visited places and extracting semantic features. Once stops are detected, they are annotated with the POI and land use types. Next, several semantic features such as stop begin time, stop frequency and average duration are extracted. The final step is the ontology based activity model. The retrieved

information from the previous steps is used to populate the STOM for reasoning activity type. Each step is explained in the following sub-sections.

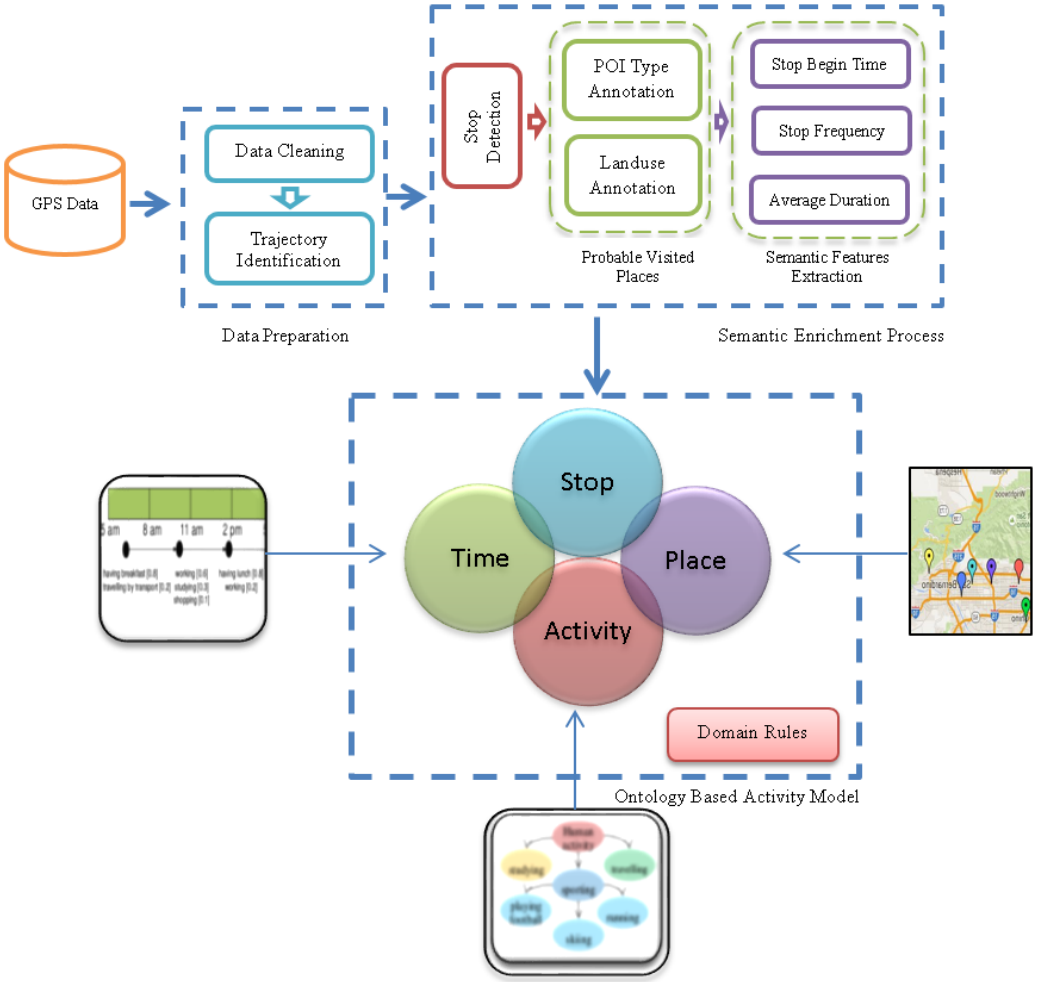


Figure 3-16 Activity recognition process

3.3.2.1 Data preparation

Due to problems in GPS data collection and sampling errors from mobile devices, the recorded positions usually contain errors (Zhang and Goodchild, 2002). Therefore, it is needed to apply techniques to identify possible causes for such errors and further remove them or reduce their influence. Data preparation consists of two steps. First, data has to be cleaned from any inconsistencies such as empty values, duplicates, and outliers. As shown in Figure 3-17, there are some GPS records with outliers (e.g., the data point in the triangle) that are far away from the data real trajectory.

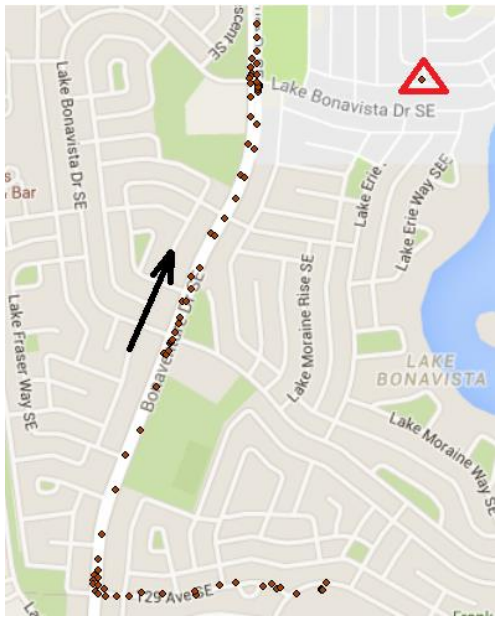
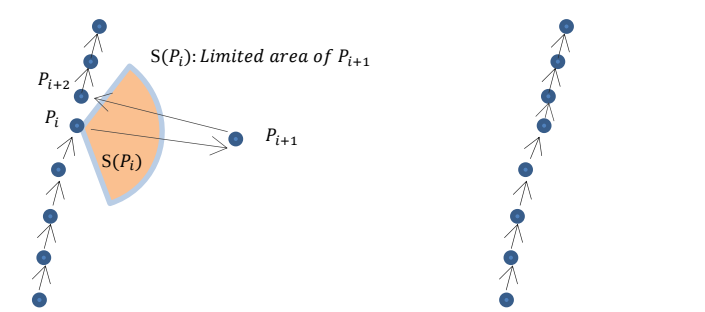


Figure 3-17 A raw GPS example with an outlier

To remove such an outlier, a data filtering strategy is used. The idea is to build a "limited area" $S(P)$ for a GPS point (P) that needs to be checked, and according to the inside topological relationship between the point P and the area $S(P)$ to determine whether P is an outlier or not. For example, to check point P_{i+1} in Figure 3-18-(a), the "limited area" $S(P_i)$ based on the previous point P_i is computed by using the maximum speed, time duration $t_{i+1} - t_i$, and the previous moving direction. If P_{i+1} is out of this area (i.e., P_{i+1} = 2 $S(P_i)$), it needs to be removed as an outlier. After removing such an outlier point, a new filtered GPS data is achieved as shown in Figure 3-18-(b).



(a) Raw GPS with outlier (b) Raw GPS with no outlier

Figure 3-18 Filtering the outliers in raw GPS data

Second, trajectory identification is applied for dividing the cleaned GPS data into daily and weekly basis trajectories (Figure 3-19). In this step, first unrealistic attributes such as trajectories with a short travel time (e.g., 10 seconds duration) is eliminated. Furthermore, if the data contain trips outside the area, they are eliminated by identifying the last road network link where the vehicle was observed.

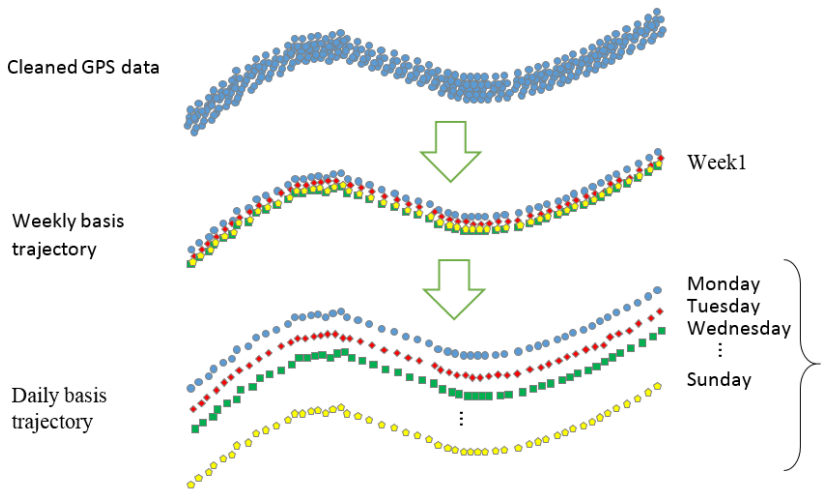


Figure 3-19 Daily and weekly basis trajectory identification

In order to identify the daily and weekly basis of trajectories, time constrain is used. The daily basis trajectories are used in order to annotate the stops with the land use and POI category types and the weekly basis trajectories are used to extract the feature such as stop begin time, stop frequency, and stop duration.

3.3.2.2 Semantic enrichment process

The enrichment process aims at extracting stops from the cleansed trajectory data and annotating them with the environmental information around them in particular by exploiting nearby POIs and land use types. It has three different steps. First, stops are detected using the Temporal Velocity Based (TVB) algorithm. Second, they are annotated with the most probable visited POI category type and land use type. Finally, some semantic features are extracted.

3.3.2.2.1 Stop detection

Given the cleaned movement data, the first step in semantic enrichment process is to identify the places where people would stop. In this research, the GPS data, which was collected by users using our application has temporal gaps (the application was turned off automatically or manually when users did not move for a certain time or when they entered into a building). Therefore, considering the format of the GPS data, the TVB algorithm is used to extract stops. Algorithm 1 provides pseudocode for determining stops. Given the speed threshold Δ_{speed} and time interval $\Delta_{duration}$, for any two consecutive GPS records $p_i(x_i, y_i, t_i)$ and $p_{i+1}(x_{i+1}, y_{i+1}, t_{i+1})$, if the speed of p is lower than Δ_{speed} and the temporal gap $t_{i+1} - t_i > \Delta_{duration}$ then p_i is the end point of the current trajectory while p_{i+1} is the starting point of the next trajectory. Therefore, p_i is considered as a stop.

Algorithm 1. TVB

Input:
Cleaned raw trajectory $T_{raw} = \{p_1, p_2, \ldots, p_n\}$
Speed threshold Δ_{speed}
Time gap $\Delta_{duration}$

Output:
Stops $T_{stops} = \{s_1, s_2, \ldots, s_m\}$

```
begin
    for all p_i = (x_i, y_i, v_i, t_i) do
        if (v_i < Δ_speed AND t_{i+1} − t_i > Δ_duration) then
            T_stops = p_i(x_i, y_i, t_i, t_{i+1}, Δ_duration)
    return T_stops
end
```

Figure 3-20 shows how stops of a vehicle can be determined by the given parameters. It presents the speed evolution of a vehicle over time and a constant speed applied all across the trajectory.

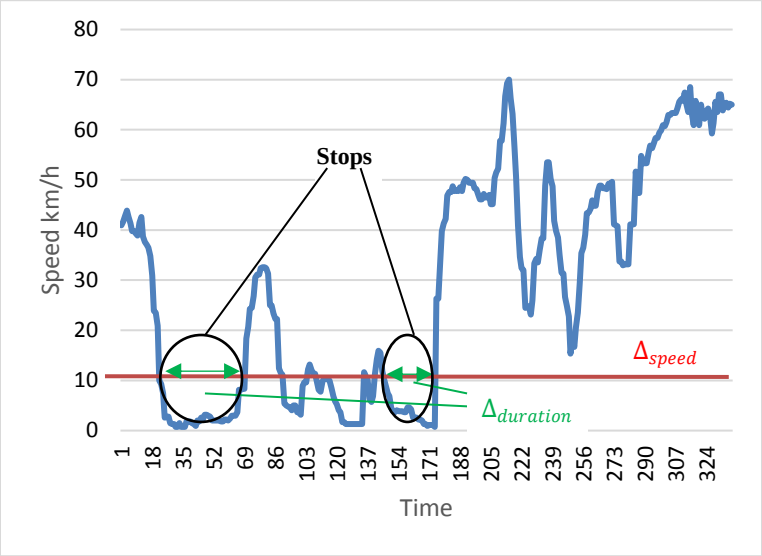


Figure 3-20 Stop detection using the TVB algorithm

3.3.2.2.2 Probable visited places

The objective of this step is to find probable places that can be visited by a user at a stop. This step utilizes available third party data sources such as Open Street Map (OSM) to gather contextual data for each stop. Therefore, two different algorithms are applied to annotate stops with the land use types (Algorithm 2) and the POI category types (Algorithm 3). As shown in Figure 3-21, land use and POI types are extracted to enrich the stop. The user has stopped in an area with a commercial land use type, which has several POIs inside. Therefore, "commercial" is assigned to the stop as a land use type and the probable POI category type is computed for the stop among different nearby POIs.

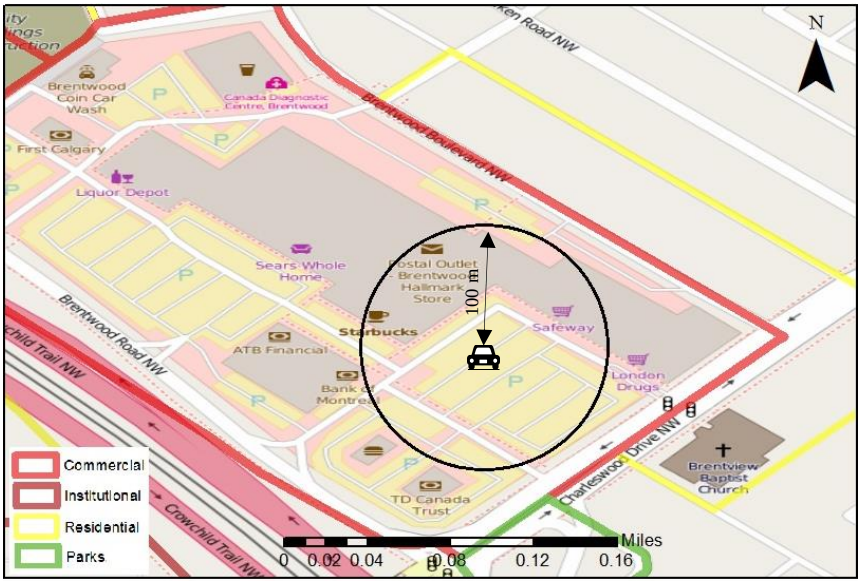


Figure 3-21 Annotating stops with land use and POI category types

(1) Annotation with land use types

This sub-section enables the annotation of stops with land use data, which includes meaningful semantic regions. Table 3-1 provides a list of land use types.

Table 3-1 Different land use types that are considered in this research

Land use type	Residential
	Parks and recreation
	Urban development
	Commercial
	Institution
	Industry
	Transportation

Algorithm 2 shows the pseudocode for the annotation procedure. For this purpose, the topological correlation is measured using the spatial join between each stop and the semantic regions. If a stop either intersects with or is nearby any regions, the stop is annotated with that region. If a stop is located out of the boundary of the city, it is annotated as an unknown area.

Algorithm 2. Select Land use Type

Input:
Stops $T_{stops} = \{s_1, s_2, \ldots, s_m\}$,

Land use layer $A_{landuse} = \{r_1, r_2, \ldots, r_q\}$

Output:
Land use type for each stop $L_{Region} = \{l_1, l_2, \ldots, l_n\}$

```
begin
    for all s_i = (x_i, y_i, t_i, t_{i+1}, Δ_duration) do
        if T_stops intersects A_landuse then
            L_i = (x_i, y_i, t_i, t_{i+1}, Δ_duration, r_i)
        else
            find nearest A_landuse to T_stops
            L_i = (x_i, y_i, t_i, t_{i+1}, Δ_duration, r_i)
    return L_region
end
```

As an example, Figure 3-22 shows a user's daily trajectory on Saturday with three different stops: S1, S2, and S3 (blue star symbols), annotated with various kinds of land use types. The user starts from S1 and then stops at S2 for 1 hour and 30 minutes; next, he stops at S3 for 45 minutes, and finally, returns to the S1 again and stays there for 11 hours. Therefore, the enriched trajectory is represented as below:

S1-Residential → S2-Commercial (stayed 1.5 hours) → S3-Park (stayed 45 minutes) → S1-Residential (stayed 11 hours).

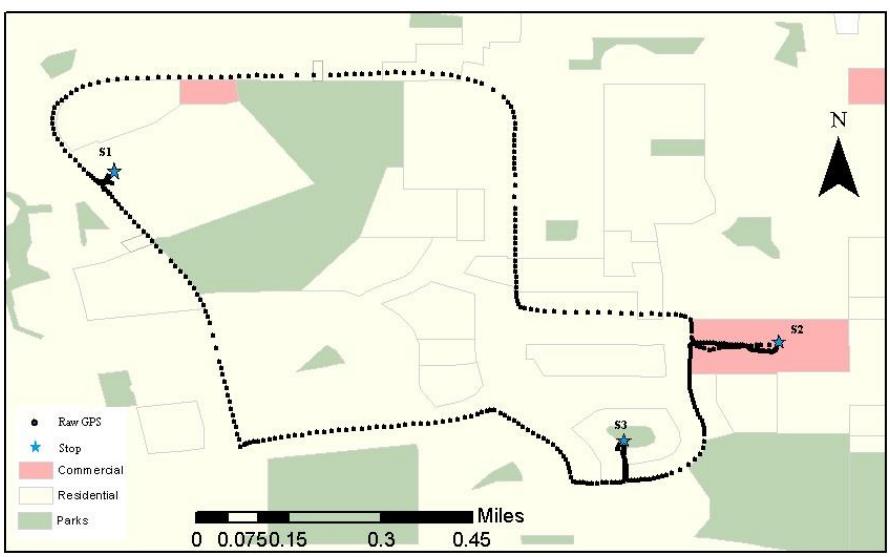


Figure 3-22 An example of a stop annotation with land use types

In this case, the three semantic regions had spatial intersections with the three stop episodes from the trajectory. Therefore, semantic trajectory is a sequence of three land use types with a temporal duration.

(2) Annotation with probable visited POI category types

This sub-section enables the annotation of the stops with different POI category types. Table 3-2 provides a list of POI and their category types. The POIs were divided into 9 category types. For instance, dining restaurants and pubs were associated with the "Food" category type; libraries, schools, and universities were associated with the "Education" category type.

Table 3-2 POIs and their category types

Category type	POIs
Food	bar, café, pub, dining restaurant, bakery, fast food restaurant, food court
Recreation	park, sports center, cinema, concert hall, gym, museum, night club, spa, stadium, zoo, bar, casino, theater
Religious	church, mosque
Education	school, college, university, library
Shopping	shopping mall, strip mall, plaza, book store, clothing store, electronics' stores, furniture store, pet store
Daily shopping	grocery store, wholesale store, departmental store, supermarket, bakery, butcher's shop
Business services	post office, car rental, gas station, ATM, industrial place, personal business
Health services	dental office, pharmacy, clinic, hospital
Accommodation	hotel, hostel

The pseudocode in Algorithm 3 shows the detailed procedure of retrieving probable visited POI category types for a given stop. The inputs of the algorithm are:

- A set of stops. Each stop S is a triple: $S_i = \{x_i, y_i, t_i, t_{i+1}, \Delta_{duration_i}\}$

where (x_i, y_i) represents the coordinates of the stop, t_i is the start time, t_{i+1} is the end time, and $duration_i$ represents the duration of the stop.

- A set of POIs. Each POI is associated with predefined categories, as displayed in Table 3.2. A POI includes:

$$POI_i = \{x_i, y_i, C_i, H_i, MST_i\}$$

where (x_i, y_i) represents the geographical position of the POI, C_i is the category type of POI, H_i is the opening hour of the POI, and MST_i is the minimum service time that each POI would provide.

- A set of characteristics of users. Maximum Walking Distance (MWD) and User Walking Speed on a road network (UWS).

Algorithm 3. POI Type Annotation

Input:
Stops $T_{stops} = \{s_1, s_2, \dots, s_m\}$,
POI layer $A_{poi} = \{a_1, a_2, \dots, a_u\}$
MWD
UWS

Output:
POI category type probability for each stop $Prob_{cat} = \{c_1, c_2, \dots, c_n\}$

```
generalPOIs=[]
probablePOIs=[]
prob=[]
begin
    for all s_i = (x_i, y_i, t_i, t_{i+1}, Δ_{duration_i}) do
        if distance (s_i, A_poi) ≤ MWD and
        s_i.time ⊂ A_poi.H_i then
        generalPOI ← A_poi

    for all POI_i in generalPOIs
        if TE ≤ duration_i then
        probablePOIs ← poi

    for all probablePOIs do
        POI_cat = {p ϵ probablePOIs: μ(p) = cat}
        dist = distance(s,p) for each p ϵ probablePOIs
        mass = length(POI_cat)
        prob ← (POI_cat, mass/dist^2)
    return POI_cat
end
```

In order to extract the POI category type for each stop, three steps are required. First, general POIs are selected; second, probable POIs are extracted, and finally gravity model is used to calculate the probability for each POI category type. To detect the general POIs two conditions are taken into account, as seen in lines 5-8. First, each POI has to be within a certain spatial range, which is defined the MWD (the distance likely to be accepted for a walk from a stop to a POI). This means that POIs should not be too far away from the stop. The Dijkstra algorithm is used to compute the distance between each stop and the POIs on a road network. Second, the time period of each stop has to be compatible with the opening hours of the POIs. A stop during the closure of a POI cannot be matched with that POI, so for example a stop at 11 pm can be matched with a restaurant or a pub but not with a museum. Therefore, general POIs are selected for a stop if they can be reached by walking and their opening time intersects with the stop time duration.

To find the probable POIs, users need to have enough time to go and visit the POI based on the Minimum Service Time (MST), and return to the stop (lines 9-11). Therefore, TE (3-3) is the time that a person needed to reach the POI, visit the POI and come back to the stop again.

$$TE = 2 * TP + MST \quad (3\text{-}3)$$

TP (3-4) is the time a person needs to cover the distance, where d is the distance between the stop and the POI and UWS is the user's speed on a road network.

$$TP = \frac{d}{UWS} \quad (3\text{-}4)$$

Once the probable visited POIs are selected, the algorithm measures probability for each POI category type (lines 12-16). A method based on the gravity model is considered for this purpose. The gravity model (3-5) is a model derived from Newton's Law of Gravitation and used to predict the degree of interaction between a stop and each POI.

$$Gravity\ law = \frac{mass_1 * mass_2}{distance^2} \quad (3\text{-}5)$$

The definition of the gravity model is represented using the principle of bodies' attraction where $mass_1$ represents a stop and $mass_2$ represents the number of the probable visited POIs in each category, and the distance is the sum of all the distances of POIs associated to the same category type. This means that the POIs associated to the same category type are assigned the same probability of being visited. More formally, for every stop s the probability p of a category type is determined as:

$$P(s_i, c_j) = \sum \frac{|\{p_k \epsilon probablePOIs(s_i) | \mu(p_k) = c_j\}|}{(d(s_i, p_k)^2)} \qquad (3\text{-}6)$$

In this formula, s_i is the stop, c_j indicates the category of POI p_k and d is a function returning the distance between each stop and the POIs associated to the same category type. Thus, using equation (3-6), a probability is associated to each possible category type relative to the stops. Therefore, stops are annotated with the POI's category type rather than the POI itself.

To do this not only the distance of the POIs from the stops are taken into account, but also the characteristics of the location where the user stopped. For example, a stop in an area with many restaurants and few stores, the gravity model gives more mass to restaurants respect to stores, then making a better distinction between the two possible POI category types (Food or Shopping). For example, as shown in Figure 3-23, there is stop and several probable visited POIs around it. The POIs, located at different distances from the stop, belong to different category types. According to the gravity model definition above, the probabilities for the two activities are the following:

Figure 3-23 A stop and the POI category types of the probable visited POIs

$$P(stop, food) = \frac{3}{3^2 + 4^2 + 23^2}$$

$$P(stop, business\ service) = \frac{3}{5^2 + 6^2 + 20^2}$$

The normalized values for the category types; food and business services are 0.45 and 0.55, respectively. This result means that we have a higher probability to have a business service

category type since the POIs mapped to business service are closer compared to the POIs mapped to Food.

3.3.2.2.3 Semantic features extraction

The last step of the semantic enrichment process is to extract semantic features from the annotated stops with the land use types. As shown in Algorithm 4, different semantic features such as stop frequency, average duration and begin time for each stop are extracted.

Algorithm 4. Semantic Features Extraction

Input:
Selected land use type $L_i = (x_i, y_i, t_i, t_{i+1}, \Delta_{duration}, r_i)$

Output:
Stop frequency S_f, average duration of a stop S_d, and begin time T_e

```
begin
    for all L_i do
        compute stop frequency per week (S_f)
        compute average duration of stop per week (S_d)
        extract begin time of a stop (T_e)
    return S_f, S_d, and T_e
end
```

3.3.2.3 Ontology based activity model

In this section an ontology based activity model is introduced, which performs an inference on the most probable activities performed by users during their trip. Given the extracted features from the previous sub-sections, the ontology model is populated and integrated in a formalism that is capable of reasoning different activity types. The activity types are defined as axioms using domain knowledge. The model (STOM) is composed of four ontologies (see Figure 3-24).

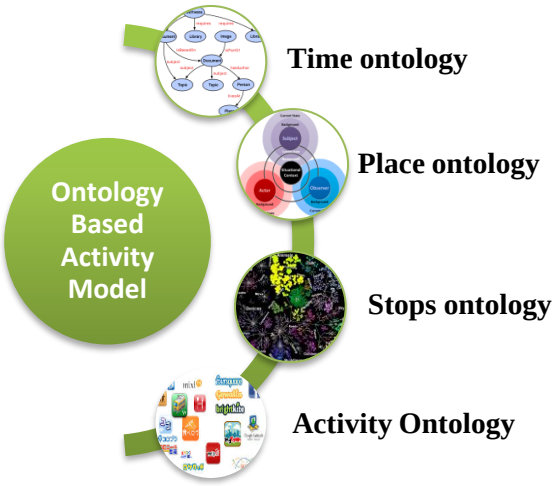


Figure 3-24 Ontology based activity model components

Activity Ontology contains user activity type classes (see Figure 3-25). Description logic is used to formalize activity types with axioms.

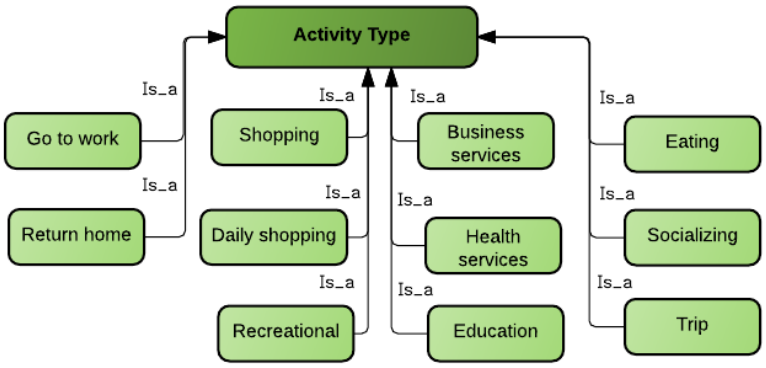


Figure 3-25 Activity ontology

Place Ontology contains various classes of the POIs and the land use types. Each of them denotes a geo-referenced object, such as a restaurant, a shop, a lake or other objects. For example, the classes in the place ontology are shown in Figure 3-26.

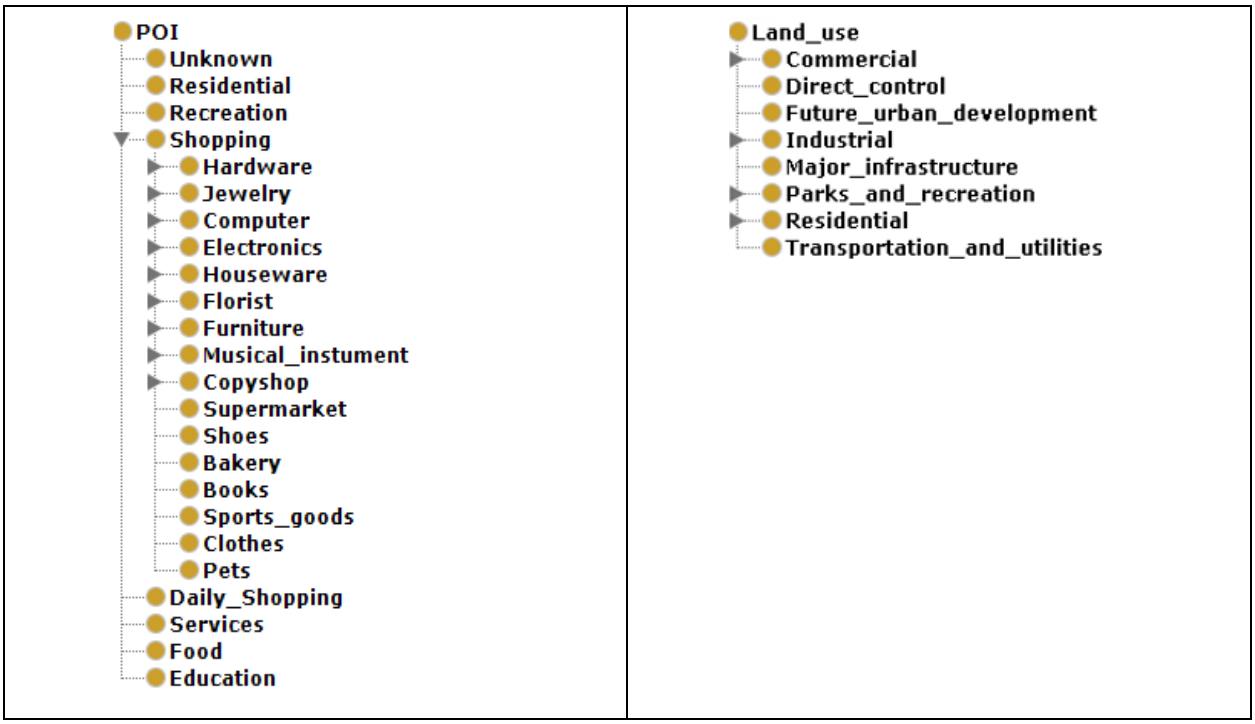


Figure 3-26 Place ontology for POI and land use classes

Time Ontology contains temporal references in which the activity types can occur. It is designed for modelling of time in qualitative terms (e.g., morning, evening). A representation of how times/days are hierarchically organized is given by the following axioms in description logic:

$$morning \sqsubseteq time$$

$$weekday \sqsubseteq day$$

$$holiday \sqsubseteq day$$

$$Saturday \sqsubseteq holiday$$

Stop ontology contains places that a user would stay for a period of time including semantic features such as stop frequency, average duration, and begin time. In the following sub-section, the ontology rules and relations among ontologies are explained.

3.3.2.3.1 User semantic trajectory ontology rules

As explained in Section 3.2, the activity types are defined using axioms based on different semantic features included in the ontology model to express relations between the ontologies. Table 3-3 shows some semantic rules between different ontology concepts. For instance, in rule number one, if the land use type is residential, the POI type is null, the begin time is evening, the stop frequency

per week is more than 5 and the average duration is more than 10 hours per week then the moving object is ‘spending time at home’, i.e., AT= Return home.

Table 3-3 Domain rules associated to activity types

No	Land use Type	POI Category Type	Features			Activity Type
			T_b	S_f	S_d	
1	Residential	-	Evening or night	≥ 5	≥ 9 hours	Return home
2	Residential	-	Evening or night	≥ 1	≥ 30 minutes	Socializing
3	Commercial	Shopping	Evening or night	≥ 1	≥ 30 minutes	Shopping
4	Commercial	Daily Shopping	Evening or night	≥ 1	≥ 30 minutes	Daily Shopping
5	Any Type	Any Type	Morning	≥ 5	≥ 8 hours	Work Full-Time
6	Any Type	Any Type	Any Time	≥ 3	≥ 5 hours	Work Part-Time
7	Commercial	Food	Any Time	≥ 1	≥ 30 minutes	Eating
8	Institutional	Education	Any Time	≥ 1	≥ 1 hour	Education
9	Park	-	Evening or night	≥ 1	≥ 30 minutes	Recreational
10	Commercial	Business Services	Evening or night	≥ 1	≥ 1 hour	Business Services
11	Any Type	Recreation	Evening or night	≥ 1	≥ 1 hour	Recreational
12	Unknown	Unknown	Any Time	≥ 1	≥ 1 hour	Trip
..	….	…	…	…	…	…
n	….	…	…	…	…	…

Another example is rule number 2: if the moving object stops once within a residential land use type per week, with an average stop duration of more than 30 min and the begin time is evening or night, then the moving object is ‘visiting a friend’, i.e., AT= Socializing. Therefore, applying the criteria outlined above, different activity types can be inferred. In this step, the ontology represents the concepts, rules, and assumptions presented in the considered application domain. The added value of having such an ontology based approach, allows one to define axioms

in terms of high-level semantic concepts, abstracting away from the geometry coordinates of the geographical features. Indeed, in this approach, each stop is treated as a semantic concept instead of using spatial coordinates. The assertion of these relationships is an existential restriction, which is specified using the following example axiom expressed in the Web Ontology Language (OWL) syntax as semantic rule number one specifies a typical home activity (Table 3-4):

Table 3-4 Definition of home activity ontology rule

Concept	Definition in Description Logic
Return home Activity	$\equiv Stop \sqcap HasLanduseType.Residential \sqcap HasBeginTime.Evening \sqcap HasPoiType.null \sqcap \exists \geq 5\ Frequency \sqcap \exists \geq 600\ Duration$

The above axiom defines the “return home” activity as a stop where the land use type is residential, has a frequency of six times a week, the duration is equal or above 600 minutes, the stop starts in the evenings and the stop has no POI type. Similarly, one can exploit this definition for characterizing individual trajectories, thus defining the concept of different activity types. Finally, the ontology model is populated with trajectories, the ontology inference engine is executed, and the defined axioms are interpreted to classify the ontology instances using the appropriate concepts on user activity.

3.3.3 Semantic behavior modelling

Once the data has been semantically annotated with the activity types in the previous step, data mining is applied to the data in order to extract behavior patterns with respect to the semantics generated in the previous step. Behavior type as described in definition 11, it indicates the regularities between users’ activity types. For instance, if a user goes shopping, what is the probability the user might do other activity types such as going for shopping again, going to visit a friend or return home afterward (Figure 3-27). The figure shows a schematic representations of the activity types and their connections.

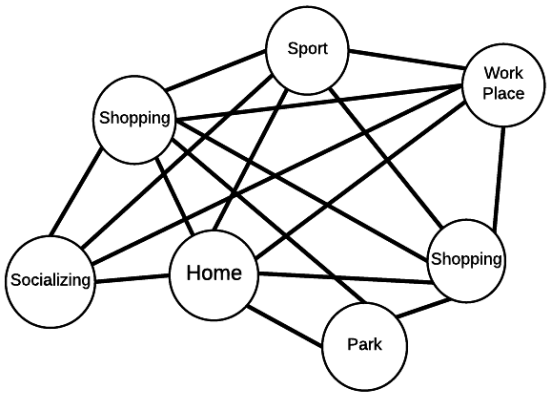


Figure 3-27 Association between different activity types

Figure 3-28 shows the procedure to model the behavior of users. It includes data preprocessing and rule mining. The input data for the modelling process are semantic trajectories, which were inferred in the activity recognition step and their corresponding properties such as start time, activity duration, and other properties. The input data has to be preprocessed and transformed into a proper dataset that is suitable as an input for rule mining. The objective of the proposed approach is to support the analyst in understanding the connection between thematic attributes and the patterns. The generated rules constitute the behavior model, which are imported into the ontology model. Each of the steps is explained further in the following sub-sections.

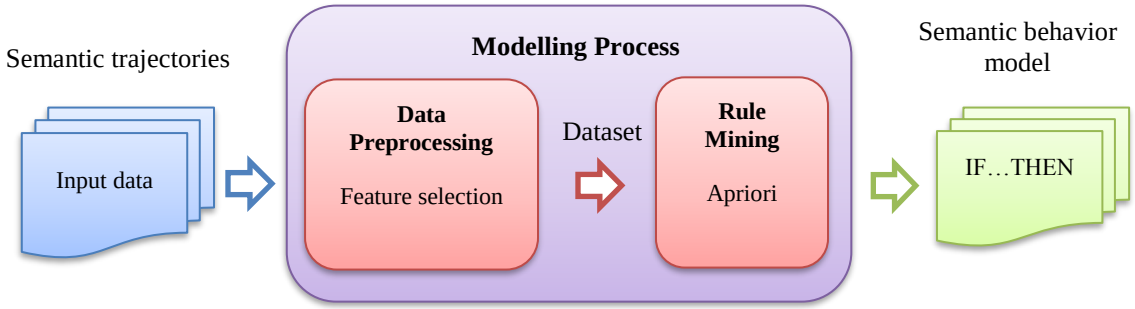


Figure 3-28 Semantic behavior model procedure

3.3.3.1 Semantic trajectories

Trajectories, which are annotated with the activity types will create semantic trajectories. As shown in Table 3-5, the output is a semantic trajectory of a user, which not only has the activity types but also has some other semantic features such as time, duration, and date. For instance, On

June 8th, the user has left home in the morning and has gone to work and stayed there for 9 hours and 13 minutes.

Table 3-5 Semantic trajectory of a user

Date	Start	Time_1	End	Time_2	Day	Stayed
8-Jun	Home	Morning	Work	Morning	Monday	9:13
8-Jun	Work	Afternoon	Shopping	Evening	Monday	2:01
8-Jun	Shopping	Evening	Home	Evening	Monday	10:12

3.3.3.2 Data preprocessing

Data preprocessing is necessary to obtain an appropriate dataset. This process involves several steps such as feature selection, transformation, and aggregation of the data. The attributes for the dataset are defined and thus finally the set of all used itemsets is determined. The ontology model is used for representing terms and relationships of the activity types. Humans regularly perform their activities at different times. Thus, it is very improbable to find regular relationships in a specific timestamp. This paper defines different ranges for date and time to extract some common behavior patterns. The outcome of several rule mining processes can be combined to obtain a more detailed model.

3.3.3.3 Association rules mining

Association rules mining methods are used to extract relationships from the observed activity types. The idea of mining association rules was introduced by Agrawal et al. (1993). The original and most common field of application is market basket analysis providing rules like "Consumers who bought product A also bought product B". This research has adapted this concept and use rules to describe relationships between different activity types of users as behavior types. The dataset D contains all the preprocessed activities that is used as input for the rule mining process. The information inside a dataset is represented by attribute-value-pairs called items. Let $A = \{a_1, \dots, a \quad \}$ be a finite set of all n task relevant items. It is determined during the preprocessing, which was explained in the previous sub-section. A subset $X \subseteq A$ is named itemset. Association Rule Mining is based on finding and analyzing frequent itemsets in the dataset. Let an instance I be a non-empty subset of A: $I \subseteq A$. All instances together form the dataset $D = \{I_1, \dots, I_n\}$. An

association rule is an implication of the form $X \Rightarrow Y$ where $X \subset A, Y \subset A, X \neq \emptyset, Y \neq \emptyset$ and $X \cap Y \neq \emptyset$. An instance satisfies a rule if it contains the itemsets X and Y. An instance I contains an itemset X where $X \subseteq I$. Two important measures of significance and interestingness are support and confidence (Agrawal and Srikant, 1994). The support (3-7) is the percentage of instances that contain X and Y, i.e. it indicates how often X and Y occurred together in the data.

$$sup\ (X \Rightarrow Y) = \frac{|\{I_i | X \cup Y \subseteq I_i, I_i \in D\}|}{|D|} \qquad (3\text{-}7)$$

The confidence of a rule (3-8) is a relative measure for the number of instances containing Y that also contains X.

$$conf\ (X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} \qquad (3\text{-}8)$$

For the behavior modelling only association rules are used that satisfy a minimum support and a minimum confidence. This process is generally performed in two stages. The first generates a set of frequencies for each activity. Minimum support criteria is used to control which activities are carried through to stage two. The process is exhaustive in that it identifies all itemsets that have frequencies greater than that defined by the minimum support. The second stage generates relevant association rules from the frequent itemsets by estimating a confidence level for each frequent itemset and removing all itemsets that do not meet the minimum confidence criteria. The minimum support and confidence criteria generally vary according to the types of patterns that the analyst is interested in identifying.

Apriori algorithm

Algorithm 5 shows the pseudocode and procedures of the *apriori* algorithm. At first step, it finds L_1. At steps 2 to 10, set of large k-itemset (L_{k-1}) generates set of candidate k-itemset (C_k). Procedure apriori-gen at step 3 generates candidates and uses prune method to remove the items are not repeatable (Agrawal and Srikant, 1994).

Algorithm 5. Apriori algorithm (Agrawal and Srikant, 1994)

Input:
Dataset (D)

Minimum support $(\min\_sup)$

Output:
Frequent itemset (L)

```
L_1 = find − frequent − 1itemsets(D);
for (k = 2; L_{k−1} ≠ ∅; k++){
    C_k = apriori − gen(L_{k−1},min-sup);
    for each transaction t∈D
        C_t = subset(C_k, t);
        for each candiate c ∈C_t
            c.count++;
    }
    L_k = {c∈C_k | c.count ≥ min − sup}
}
return L = ∪_k L_k;
```

```
procedure apriori − gen(L_{k−1}: frequent (k − 1) − itemsets;
min − sup: minimum support threshold)
for each itemset l_1∈L_{k−1}
    for each itemset l_2∈L_{k−1}
        if (l_1[1] = l_2[1])∧(l_1[2] = l_2[2]) ∧ … ∧ ( l_1[k − 2] = l_2[k − 2]) ∧ (l_1[k − 1] <
        l_2[k − 1])then{
            c = l_1 ⋈ l_2;
            if has − infrequent − subset(c, L_{k−1}) then
                delete c;
            else add c to C_k;
        }
return C_k;
```

```
procedure has − infrequent − subset(c: candidate k − itemset; L_(k − 1): frequent (k − 1) − itemsets);
for each (k − 1) − subset s of c
    if s ∉ L_{k−1} then
        return TRUE;
return FALSE;
```

3.3.3.4 Ontology based behavior model

As mentioned in Section 3.3.2, different types of behavior patterns such as semantic, semantic and time, semantic and space, and finally, semantic and space-time are considered as it can be seen in Figure 3-29.

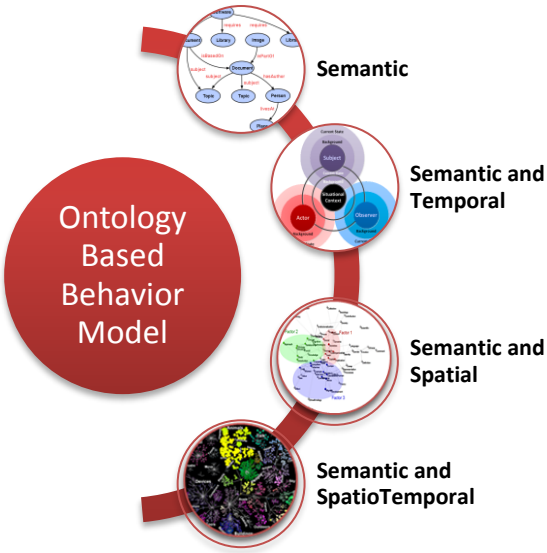


Figure 3-29 Different semantic behavior models

3.4 Summary

Firstly, in this chapter the proposed semantic conceptual data model that has been used in this research is presented. Some definitions regarding the important concepts were provided. Secondly, an ontology based semantic knowledge discovery framework was described for inferring human behavior by interpreting movement patterns. The framework consisted of three different steps, namely: semantic trajectory ontology modelling, activity recognition, and semantic behavior modelling. In the first step, ontology model was built based on the proposed semantic conceptual model. Next, user activity types were extracted and semantic trajectories were built. In the third step, association rule mining was used to extract the user semantic behavior model. In the next chapter, the process of implementing the described framework into a system as a prototype is described.

CHAPTER FOUR: PROTOTYPE IMPLEMENTATION

4.1 Introduction

Given the basic components illustrated in the previous chapter, the ontology based semantic knowledge discovery framework is implemented into a prototype system. This chapter shows how the system handles each step of the process: from the ontology construction to the activity recognition and semantic behavior modelling. A detailed explanation of each essential components of the system is presented.

4.2 Prototype Implementation

The final objective of this research is to develop a prototype to evaluate the applicability and usefulness of generated information using the proposed methodology. In order to develop the prototype, a number of components are required (Mousavi and Hunter, 2012b). As shown in Figure 4-1, the overall system has several essential components such as a database server, an ontology constructor, a mining module, a reasoning engine, and a user interface. The database server receives the current position data and stores it. It also executes the activity recognition step, which was described in Chapter 3. In the ontology constructor, the ontology model, STOM is built. The extracted semantic features from the activity recognition step are used to populate STOM and then the reasoning engine is executed in order to classify the activity types. In the mining module, association rule methods are applied on the semantic trajectories to extract semantic behavior rules. The results are employed in the reasoning engine. The process needs the current position of the user, which is transferred to the reasoning engine to find out relevant services. A more detailed description of the overall system components is provided in the following sub-sections.

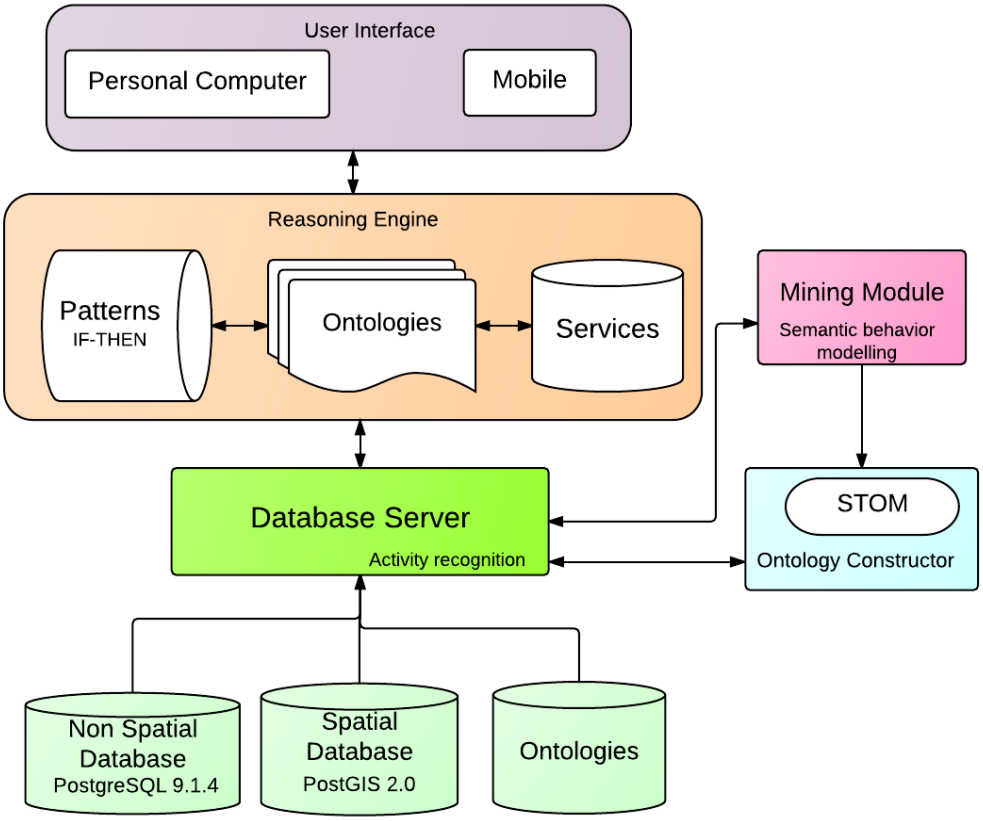


Figure 4-1 Main components of the system prototype

4.2.1 Database server

In this research, PostgreSQL 9.1.4 and the spatial database extension PostGIS 2.0 are used to manage trajectory data, maps/layers, mined semantic patterns, and ontologies. The database server runs the activity recognition step. Therefore, in this component, data are cleaned and trajectories are reconstructed. Later, stops are computed and using the required algorithms the stops are annotated and semantic features are extracted. Next, the extracted features are sent to the ontology constructor component to populate the STOM to be reasoned in order to extract activity types. Later, semantic trajectories are created. After that, the semantic trajectories are sent to the mining module to extract semantic behavior rules. This component is connected to the reasoning engine and the mining module.

4.2.2 Ontology constructor

This component represents the main concepts in the movement behavior domain and it is equipped with a reasoning engine for activity and behavior type inference. To create the ontology model four steps were considered. The first is specification of activities, which describes why the ontology is being constructed, and what its intended uses are. In this research, the specification will include activity identification as the domain ontology. Secondly, conceptualization of activities converts an informally perceived view of the domain ontology into a conceptual model, typically represented as a graph and/or tables. Third, formalization transforms the conceptual model into a formal machine-readable model, and last, implementation codes machine-readable models in a computational ontology language using an ontology editor. There are several kinds of ontology editors and browsers to perform the formalization and implementation processes. Protégé and Swoop, which are the most popular ontology editors were used in this research. These ontology editors first convert the conceptual model to a formal conceptualization and then transfer that to a machine-readable ontology OWL. Protégé is based on a logical model which makes it possible for concepts to be defined as well as described. It is composed of individuals, properties, and classes. Classes represent concrete concepts. Individuals are nominated as instances. A property generally indicates a relationship between two individuals.

The STOM (Sub-section 4.2.2.1) and service ontology (Sub-section 4.2.2.2) were built using Protégé software in this component. The mapping between ontology concepts to the data and patterns was defined using an ontology mapping process (Sub-section 4.2.2.3). Next, the Eclipse 4.2.2, which is one the most popular tools for programming in Java, was utilized to employ this mapping and enable applications to gain a data ontology view on a database via the Jena API as shown in Figure 4-2. The STOM was used in the activity recognition step, and the service ontology was used to model the semantic behavior of a user and also to match available services based on the model using match processing (Sub-section 4.2.2.4).

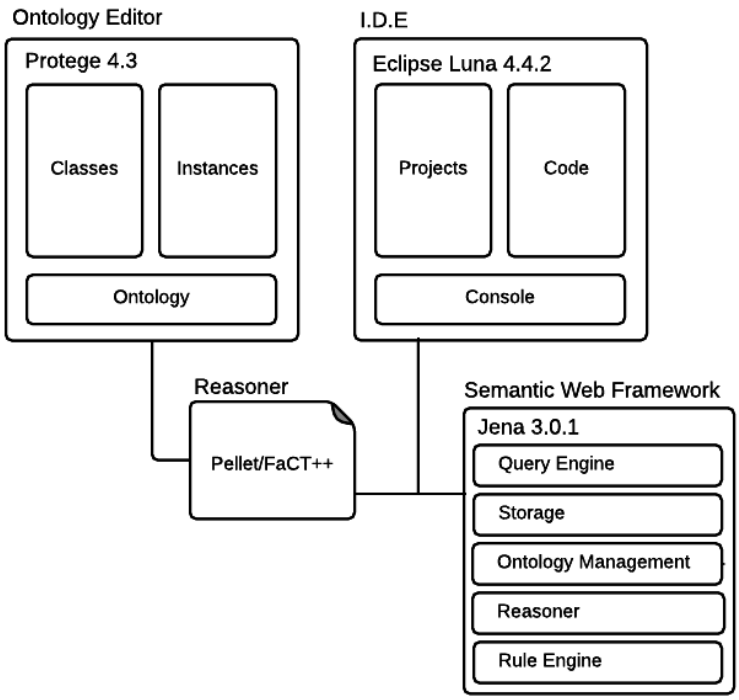


Figure 4-2 Ontology construction architecture

4.2.2.1 STOM

As seen in Figure 4-3, the ontology model consists of different classes. Figure 4-4 illustrates the structure of the STOM. The highest node is a *Thing* class. A subclass of *Thing* class is the *moving object* class. Furthermore, the *moving object* class has several subclasses: *trajectory*, *stop*, *move*, *place*, *activity*, and *behavior*. In Protégé, if a *trajectory* is a subclass of *moving object* then all instances of *trajectory* are instances of *moving object*, without any exceptions. The property (-has subclass) is shown by the line which connects the *Thing* and *moving object* classes together to designate that the *Thing* class has a subclass, i.e., *moving object* class. Likewise, the *moving object* class is connected to other classes to emphasize that these classes are subclasses of the *moving object*.

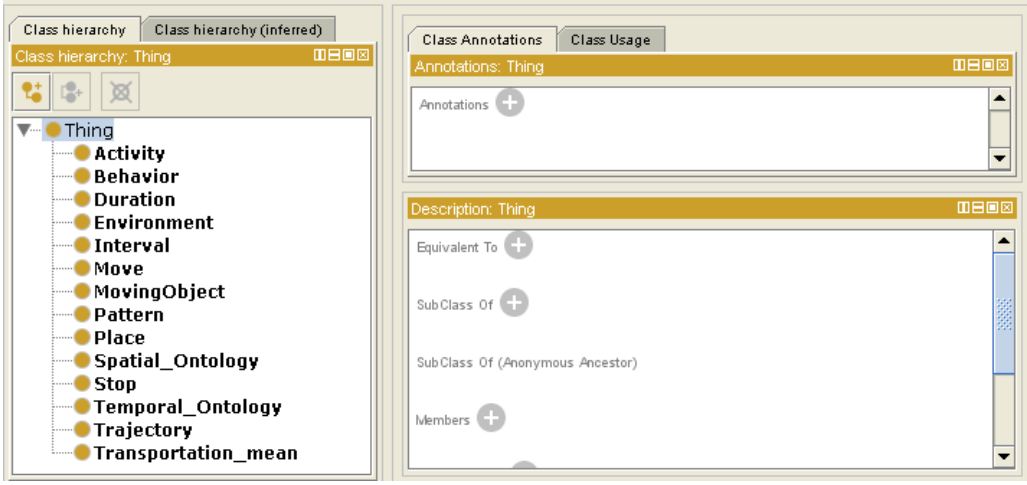


Figure 4-3 Classes of the STOM

By default, Protégé classes are assumed to overlap with one another. In order to separate a group of classes from one another, the concept of "disjoint" is used. This ensures that an individual who has been stated to be a member of one of the classes in the group cannot be a member of any other classes in that group. With this in mind, in the ontology model, all classes have been made disjointed from each other. This means that it is not possible for an individual to be a member of multiple classes. Protégé properties state relationships between two individuals. There are two main types of properties, object properties and datatype properties. Object properties link one individual to another. Datatype properties link an individual to an XML schema datatype value or an OWL.

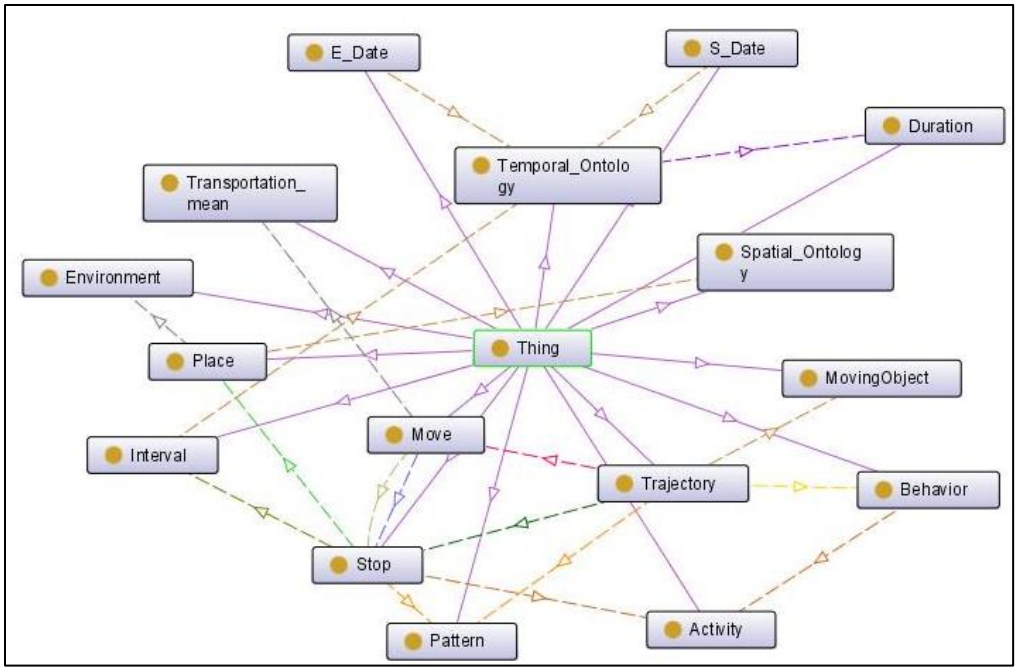


Figure 4-4 Hierarchical structure of the STOM

All relationships between the above noted classes and their properties have been shown in Figure 4-5. Arc types for different classes indicating the types of relationships between classes can be seen in Figure 4-6. For instance, the trajectory class has the property of "IsComposedOf Stop" with the stop class.

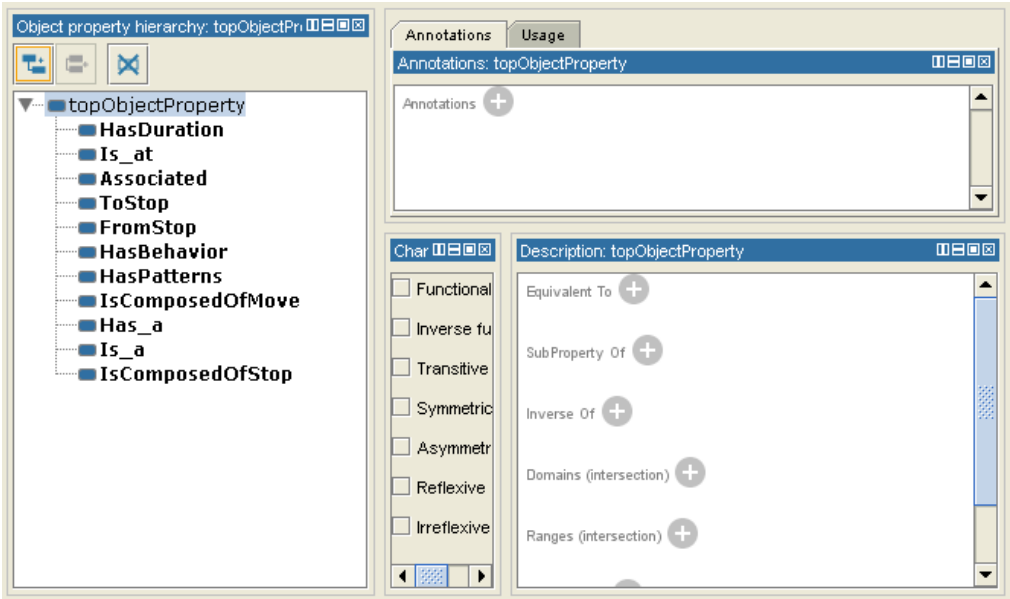


Figure 4-5 Object properties to define relationships between different concepts

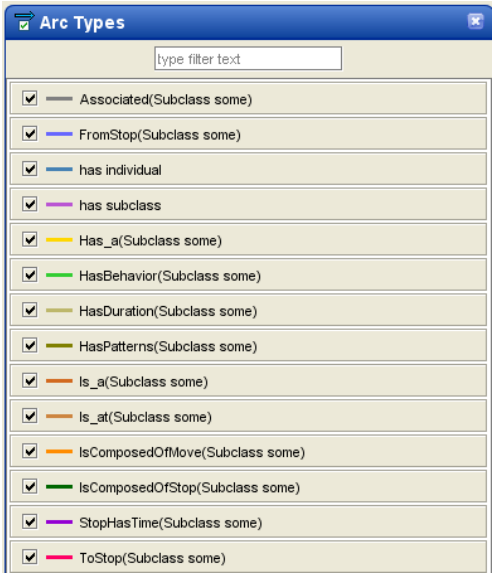


Figure 4-6 Arc types for different classes indicating the type of relationship between the classes

4.2.2.2 Service ontology

The ontology formalization and implementation processes of the service ontology were developed in Protégé. To create a service ontology, the STOM was called inside this ontology. Therefore, all classes and subclasses including all properties of the ontology model transferred to the service ontology. It has three subclasses, namely: spatial services, temporal services and spatiotemporal services. Later, the object and datatype properties between the ontology model and the service classes were created. Figure 4-7 shows the classes and subclasses of the service ontology.

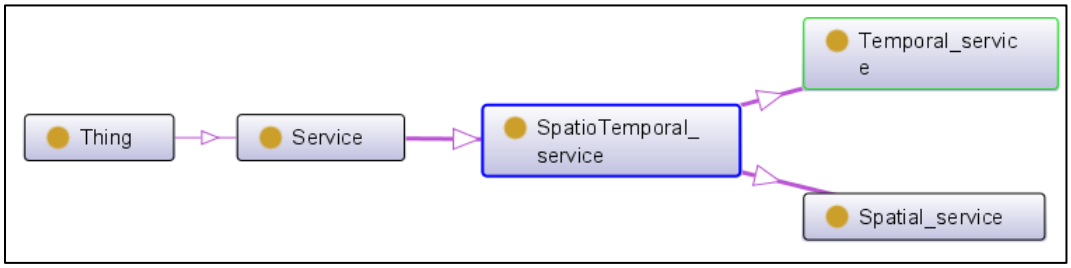


Figure 4-7 Service ontology classes

4.2.2.3 Mapping process

In this research D2RQ mapping, which is a declarative language is used to specify rules for translating relational database schemas into their corresponding OWL ontological structures. The

data repository includes different tables such as land use, POI, and stop in a database. Therefore, the mapping process was used to map and convert the data to a data ontology format (e.g., Resource Description Framework (RDF) or OWL). This process is achieved by performing mapping among semantically similar schema components. Figure 4-8 illustrates the mapping process.

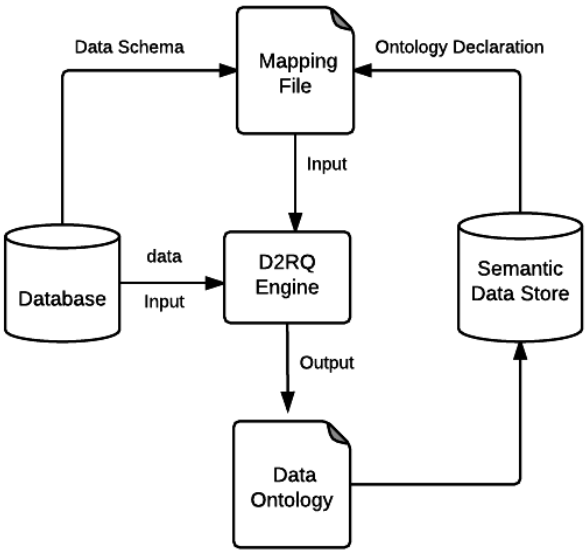


Figure 4-8 Relational database to RDF mapping process

For instance, to import a simple table named "Restaurant" representing standard information such as ID, name, and address, it is defined as below to map the "Restaurant" concept to the "Restaurant" table.

"Restaurant" ← *"SELECT ID, name, address FROM Restaurant*

In addition, the Swoop software was utilized to convert RDF to OWL ontologies and the resulting OWL ontologies are transferred to the service ontology component. In fact, the data ontology is an instance of the geometry ontology, geography ontology, place ontology, and ontology model. As it can be seen in Figure 4-9, different activity types and stops are defined in an ontology. Each activity type has different features such as duration, previous stop, time, and day of the week.

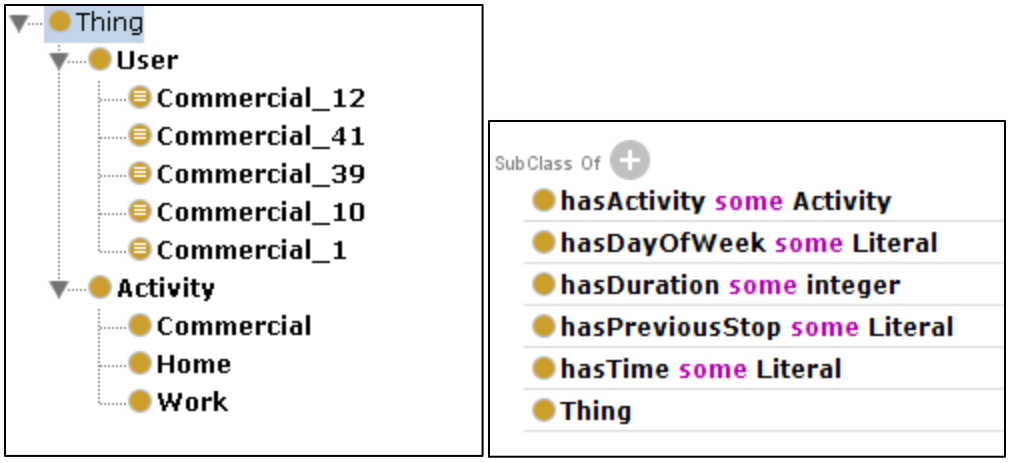


Figure 4-9 Ontology classes in the user and activity definition

4.2.2.4 Matching process

Once the data ontology is constructed from the relational database, the matching process is required. The matching process creates connections between various variables of the data ontology outputs and the service matching component inputs. In this research, the matching process loads the service matching ontology into the data ontology. The Swoop software was employed for this matching process.

4.2.3 Mining module

The Weka software was used for the mining module, which runs association rule algorithm on semantic trajectories stored in the data repository in order to generate semantic behavior patterns. These patterns are then translated into data ontology instances to populate ontology classes. After that, they are parsed to the service ontology, where the reasoning engine interprets them and newly generated knowledge is stored into the repository. As can be seen in Figure 4-10 the patterns are set as axioms in the STOM for reasoning.

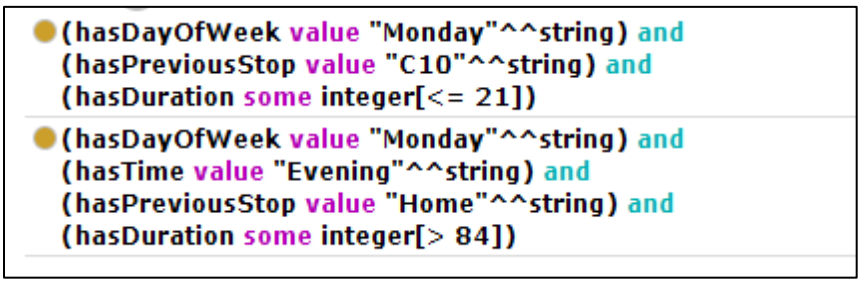


Figure 4-10 Semantic behavior model of a user

4.2.4 Reasoning engine

The reasoning engine is executed by FaCT++ reasoner. It is used to interpret the predefined axioms and it classifies ontology instances. The aim of the reasoning engine is to match and deduce useful context using the service ontology. As can be seen in Figure 4-11, the "Home" activity type is defined as an axiom by considering different semantic features.

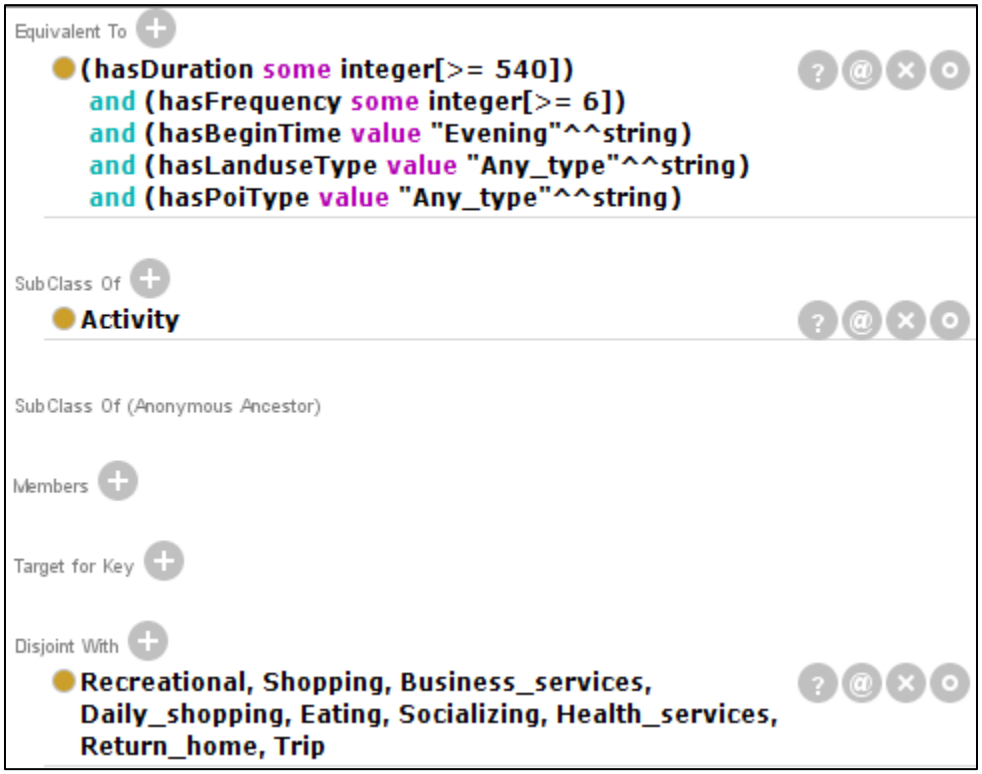


Figure 4-11 An axiom to define home activity type

As an example of semantic behavior type reasoning, as seen in Figure 4-12, the current status of a user is presented as the day of the week (Monday), the previous stop of the user ("home") and the duration for which the user stayed there (20 minutes). As a result, the reasoner has classified the current position as space_Commercial_1, which is a type of a shopping service. This decision is based on the user semantic behavior model and his current status.

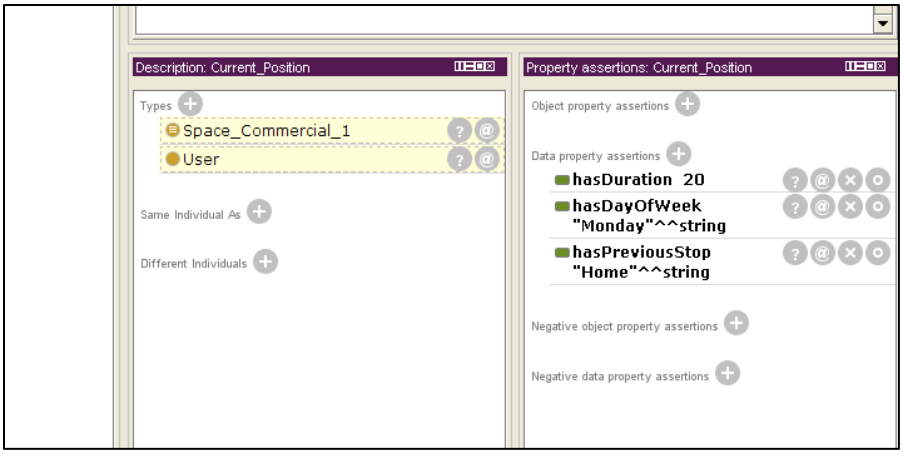


Figure 4-12 The current status of a user and the reasoned shopping service

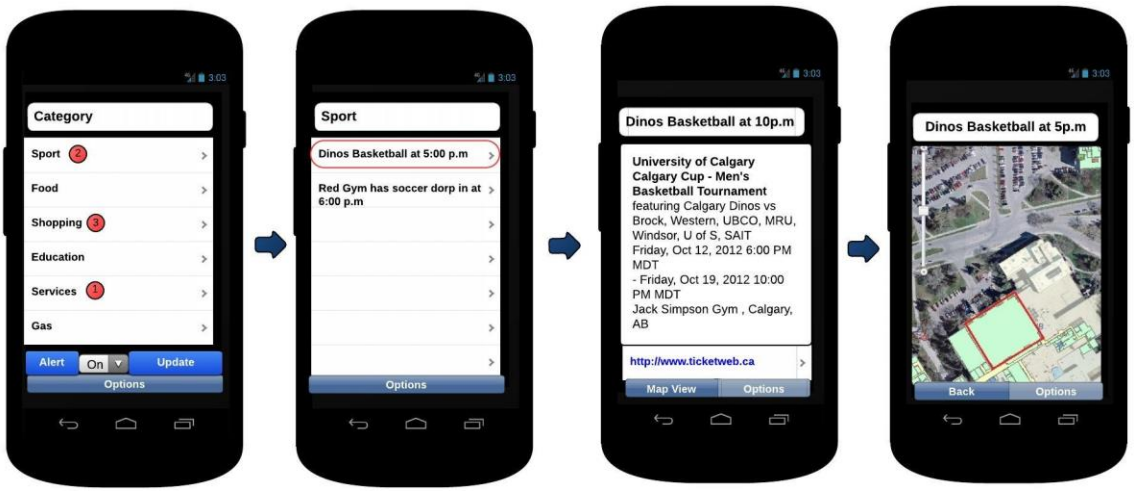
4.2.5 User interface

The user interface has capabilities for data visualization and it also provides an interactive interface for service providers to register their services to the system. Users are to interact with a mobile phone or a desktop computer. The user interface was developed using PhoneGap, which is an open source mobile development framework and it enables software programmers to build applications for mobile devices using JavaScript, HTML5, and CSS3. It uses standards-based web technologies to bridge web applications and mobile devices. To display maps and feature rendering Leaflet was used as a web mapping API. The application server Apache 2.2.12 is used to manage user interaction and to send a list of relevant services, as a result from the reasoning engine, to be displayed in the user interface.

Figure 4-13 presents a schematic of the main page of the prototype, which has different functionalities such as monitoring, searching and, providing services. The “monitoring” provides different maps and displays the current position of users. “Searching” help users to find nearby places of interest.

Figure 4-13 The main page of the user interface

The "services" functionality delivers relevant available services to the users based on their semantic behavior model. As it can be seen in Figure 4-14-(a), a user has received several services in different categories such as sport, shopping, and business services. Once the user chooses one of the categories, the available services will be shown to the user (Figure 4-14-(b)). The user can click on the services and see the details (see Figure 4-14-(c)). In addition, if the user clicks on the "Map view" button, they can view the location of the recommended service on the map (Figure 4-14-(d)).



(a) (b) (c) (d)

Figure 4-14 The services page of the prototype

If users set the alert option to “on” (see Figure 4-15), then they will leave any new services pop up on the main screen of the device. Once they have a new service, they can click on the “check it now” button to see the details or click on the “Later” button to visit it at a later time.

Figure 4-15 The popped up service message when the alert is turned on

4.3 Summary

In this chapter the entire process of the prototype implementation is explained. The prototype had several components such as a database server, an ontology constructor, a mining module, a reasoning engine, and a user interface. Protégé and Swoop were used as ontology editors in this research due to their popular tools for ontology creation. In the next chapter, the conducted experiment in this research and also the results are described.

CHAPTER FIVE: EXPERIMENTS AND RESULTS

5.1 Introduction

This chapter describes in detail the experimental analysis on two different datasets. First, the activity recognition and the semantic behavior modelling steps are executed. Moreover, the prototype testing is elaborated on. Finally, the results of each step are discussed and evaluated.

5.2 Experiments

The performance of the framework was evaluated using a simulated dataset and two different real datasets. The GPS data used for this experiment were captured by two users who have installed an application (developed based on the prototype for this research) on their phone and carried the phone with them while driving a car in the city for a few months. The users have signed a disclosure agreement to collect their movement data and provide it for analysis. In this regard, some techniques (Chen and Liu, 2011; Monreale et al., 2011) have been suggested to make the semantic trajectories anonymous, which is out of the scope of this research. As seen in Table 5-1, there are two different datasets: Calgary and Tehran datasets, which are explained in the following sections.

Table 5-1 Different datasets used for the experiment

Dataset	# Objects	# GPS records	Tracking time	# Land use	# POI
Calgary	1	862,046	1 year	8,050 polygons	17,307 points
Tehran	1	325,152	4 months	12,850 polygons	268,122 points

First, the simulated dataset was used to investigate the effect of the UWD in algorithm 3 on the POI category types annotation. The result of this investigation was applied in the two real datasets. Second, the experiment took into account the raw GPS data as input to the activity recognition step. The data were cleaned and stops were identified. Then each stop was annotated with land use type and POI category type. After that, all the necessary features were used to populate the STOM to infer activity types. Once the activity types were inferred, they were used to create semantic trajectories, which were input for the semantic behavior modelling step.

5.2.1 Simulated dataset

To validate the activity recognition approach, the activity sequence for one of the study participants was simulated for a period of two months. It was chosen that the user should have 2 to 5 stops in a day to perform different activity types. The cumulative number of stops for each day of the week are shown in Table 5-2. For instance, 26 stops were generated for all Thursdays, whereas only 19 stops were generated for all Mondays.

Table 5-2 Cumulative number of stops for each day of the week

Day	Stops
Monday	19
Tuesday	24
Wednesday	20
Thursday	26
Friday	25
Saturday	19
Sunday	19

Table 5-3 shows the number of different land use types used in the simulation. For instance, for residential areas, six different places were used, including the place where the user lives. The rest of the residential areas are ones that the user would stop to visit friends.

Table 5-3 Number of different land use types

Stop Type	Place type
Residential	6
Commercial	10
Institutional	1
Parks	3
Industrial	2

Figure 5-1 shows all the stops annotated with different land use types. The majority of the places were located in the west of the city.

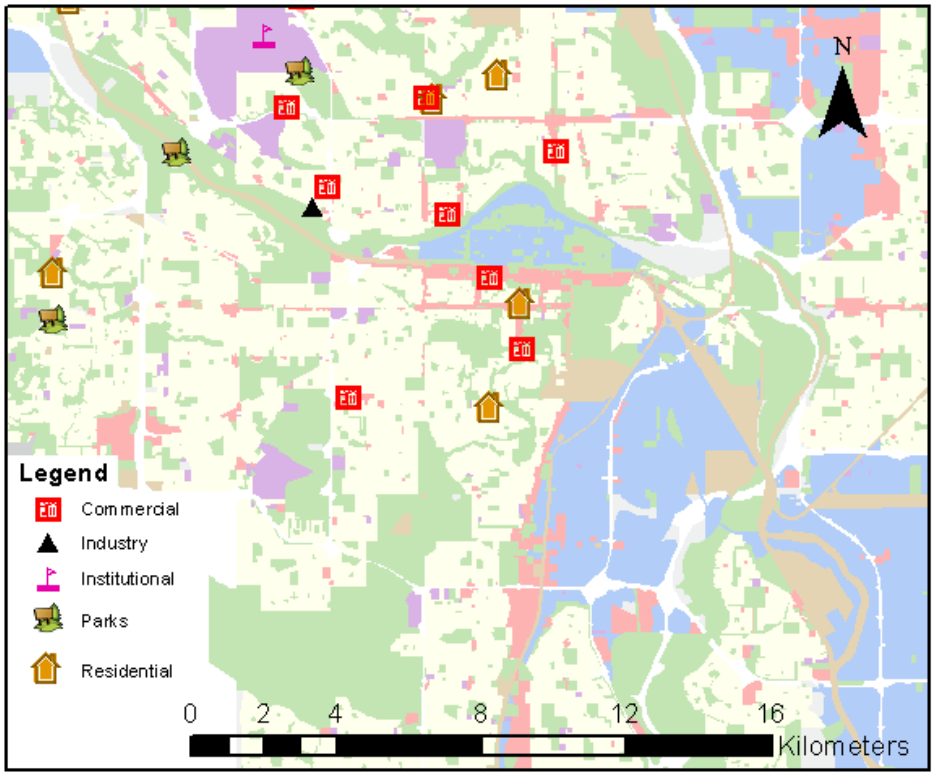


Figure 5-1 Annotated stops with different land use types

Table 3 shows the land use distribution for the generated stops. Most of the stops were annotated with commercial (47.7 %) and residential (43.7 %) areas.

Table 3: Land use type distribution

Land use Type	Percent
Residential	43.3 %
Commercial	47.7 %
Industrial	2.1 %
Institutional	2.6 %
Parks	5.3 %

Algorithm 3 was used in order to annotate the stops with the POI category types. The UWS was considered as 5 km/h since this is an average speed for users while walking. Sensitivity analysis was used to observe the effect of distance on the POI category type. For example, it was applied to guess what is the expected distance to walk from a stall in a parking lot to a specific

place where the activity takes place. Therefore, four different values 50, 75, 100, and 125 meters were assigned for the UWD parameter. As shown in Figure 5-2, different number of POI category types were assigned to stops by using different distance values. For instance, by using 50 meters in the algorithm, only 8 food category types were assigned to the stops, whereas by using 125 meters, 17 of them were assigned to the stops. It is noticeable that the longer the distance used in the algorithm, the less unknown stops were obtained. The reason is that once the user has stopped at an area such as a residential area, usually there are no POIs nearby by using 50 meters as a UWD. But, by using longer distances, there is a higher chance to assign POIs nearby the stops.

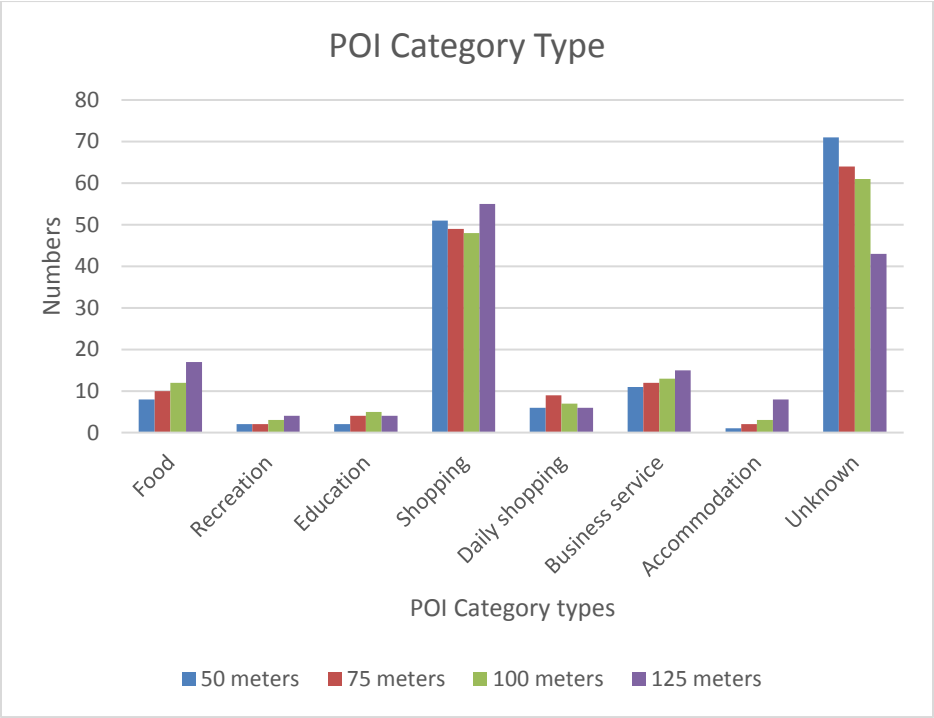


Figure 5-2 The number of stops assigned with different POI category types for various distances

Figure 5-3, depicts the accuracy of the POI category type annotation using different distances. By using 50 meters, the highest accuracy was obtained for the business services and the accuracy for the rest of the POI category types was low. By using 75 meters, the highest accuracy was obtained for the education POI category type and also the best accuracy was achieved for the daily shopping POI category type compared to the other distance values. By using 100 meters, the

accuracy was almost equally high for all the POI category types except for the daily shopping one. By using 125 meters, the accuracy was only high for the education POI category type.

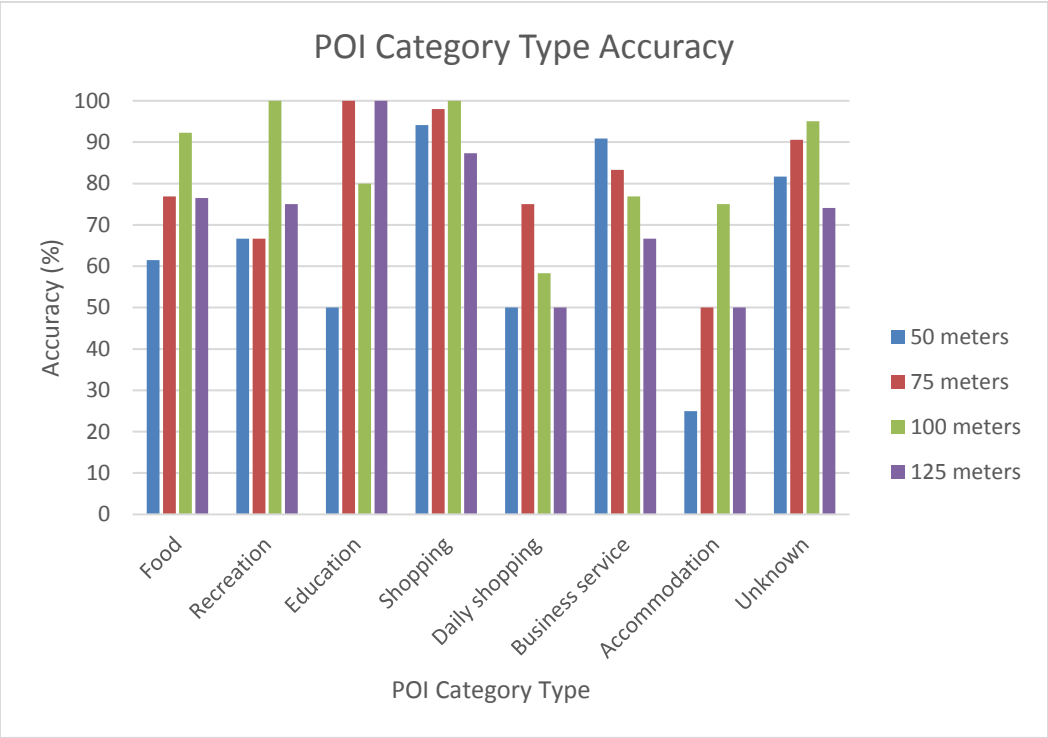


Figure 5-3 Sensitivity analysis of the POI category type for different distance values

The overall accuracies of the POI category type annotation using different distance values are shown in Figure 5-4. The highest accuracy was obtained when 100 meters was used in the algorithm and the lowest belonged to the 50-meter distance.

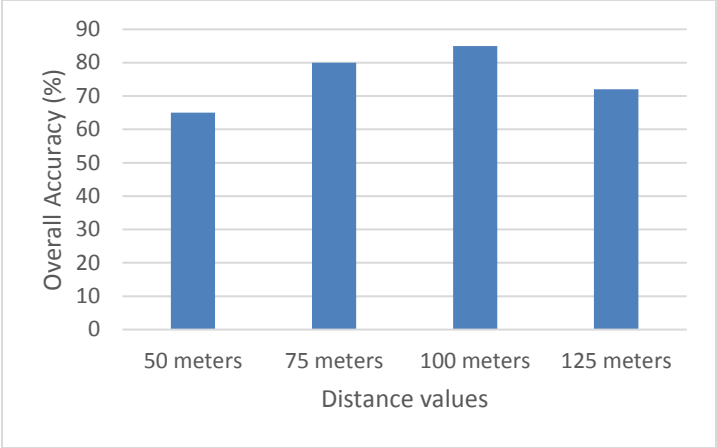


Figure 5-4 The overall accuracy of the POI category annotation using different distance values

Moreover, sensitivity analysis was used to investigate the effect of POI category types annotated using different distance values on the activity reasoning approach. Figure 5-5 shows the accuracy of inferred activity types based on different distance values. By considering the POI category types, which were annotated using 50-meter distance, the highest accuracies were obtained for the business services, socializing, and return home. The reason for the first activity type is that the accuracy of the POI category type for the business services was high by using 50-meter distance. The reason for the second two activity types is that the domain rules, which were defined to infer activity types, were not related to the POI category types and only the land use types were considered in this regard. By considering the POI category types, which were annotated using 75-meter distance, the best accuracy was achieved for the daily shopping activity type compared to the others POI category types, which were annotated using different distance values. By considering the POI category types, which were annotated using 100-meter distance, the best accuracy was obtained for the recreational activity type and the accuracy was almost high for all the activity types except for the daily shopping activity type. By considering the POI category types, which were annotated using 125-meter distance, the best accuracy was obtained for the education activity type and the accuracy for the rest of the activity types was low. It was noticeable that the accuracy of go to work activity type was the same for all different types. The reason is that the domain rule for this type of activity was not based on the POI category type and the land use type. It was based on other semantic features such as stop begin time, stop frequency, and stop duration.

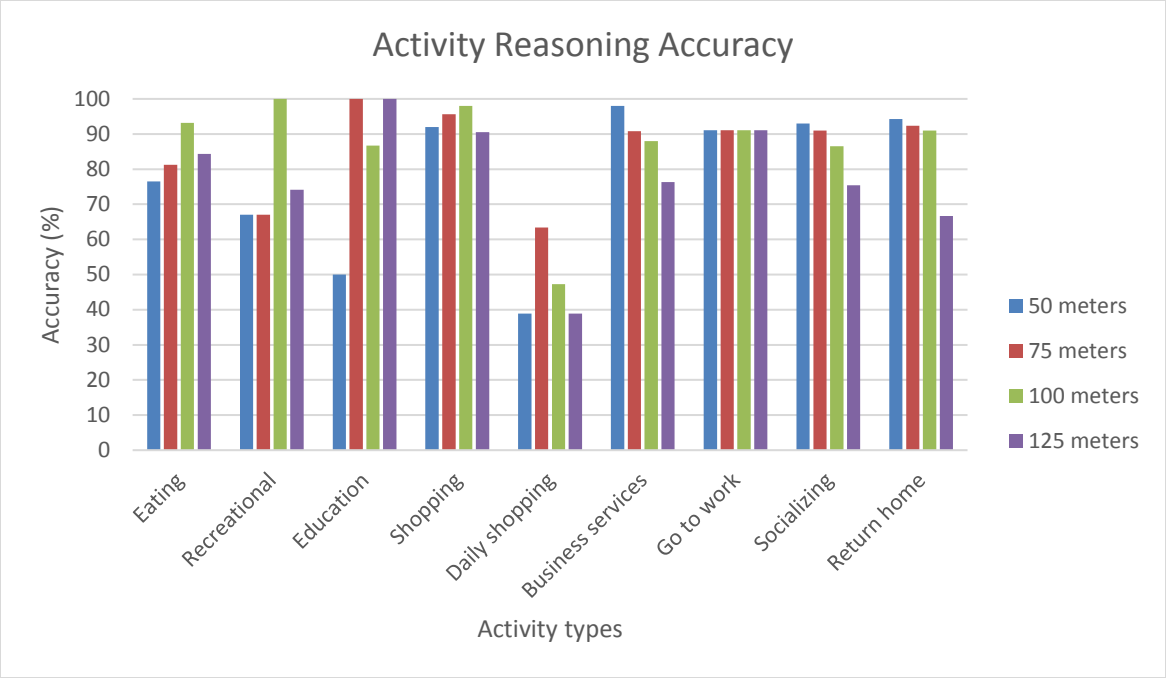


Figure 5-5 Sensitivity analysis of the activity reasoning for different distance values

The overall accuracies of the activity reasoning using different distance values are shown in Figure 5-6. The highest accuracy was obtained when the POI category types, which were annotated using 100-meter distance was considered in the activity reasoning and the lowest belonged to the 50- and 125-meter distances.

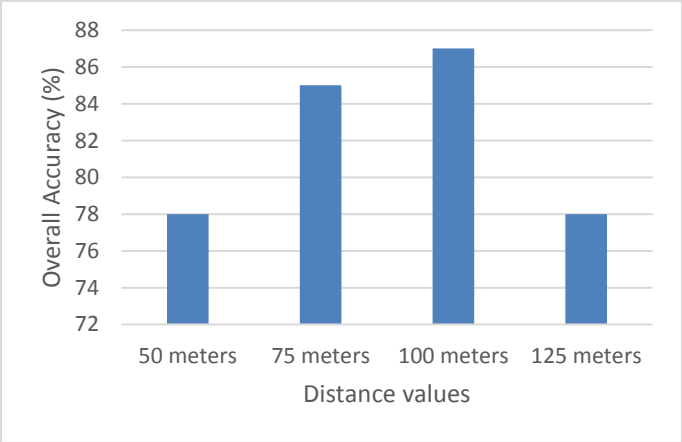


Figure 5-6 The overall accuracy of the activity type reasoning using different distance values

Therefore, the value of the UWD parameter in algorithm 3, which is used to annotate POI category type to the stops is considered as 100-meter in the experiments.

5.2.2 Calgary dataset

The Calgary's dataset was acquired over a year within the urban area in the City of Calgary in 2010. As it can be seen in Figure 5-7, the last six months of the year had more data recorded than the first six months.

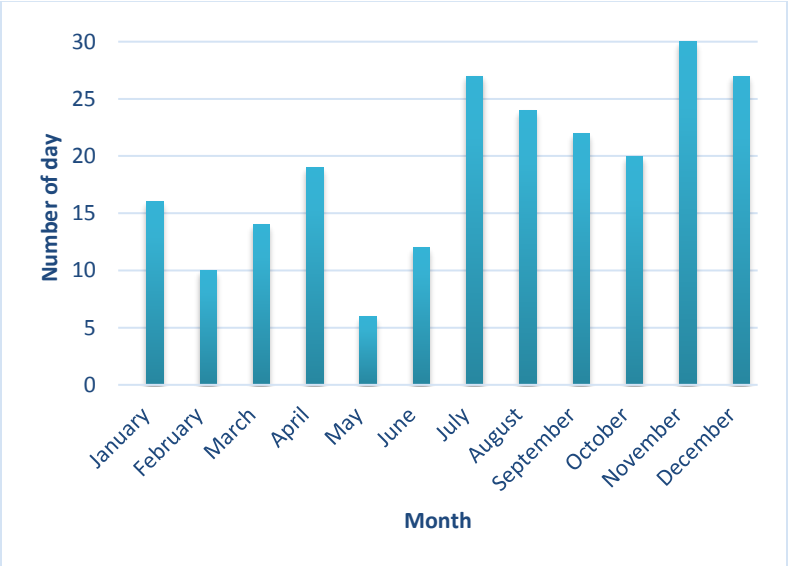


Figure 5-7 Calgary's raw GPS data acquired over a year

It has a total of 862,046 GPS records. The attributes collected include: user id, date, speed, heading, mode, and location of the user as shown in Table 5-4. The x and y coordinates were transformed to an attribute called geometry (geometry column).

Table 5-4 Attributes of the collected data

User Id	Date	Speed	Heading	Mode	Geometry
1	2010-07-06 07:51:18	25.74	24.3	Car	"POINT(-114.133044 50.940534)"
1	2010-07-06 07:51:19	30.19	22.7	Car	"POINT(-114.132991 50.940454)"

Figure 5-8 visualizes the Calgary's dataset and shows the routes that the user is used to complete various activity types.

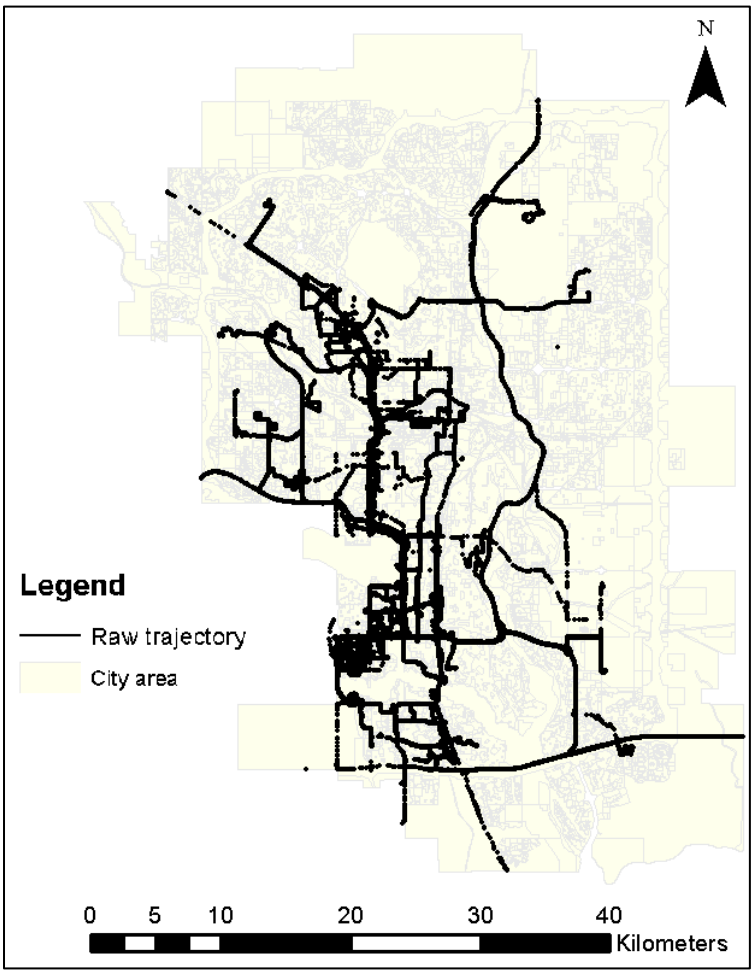


Figure 5-8 Visualization of the Calgary's dataset

5.2.2.1 Data sources

(1) Land use data of Calgary

As shown in Figure 5-9, the land use data includes different types such as commercial, urban development, residential, institutional, industrial, parks, major infrastructure, and transportation. The land use is predominantly residential, with most industrial uses in the eastern half of the city. Commercial and green space land use types are spread throughout the city. Each polygon has semantic annotations.

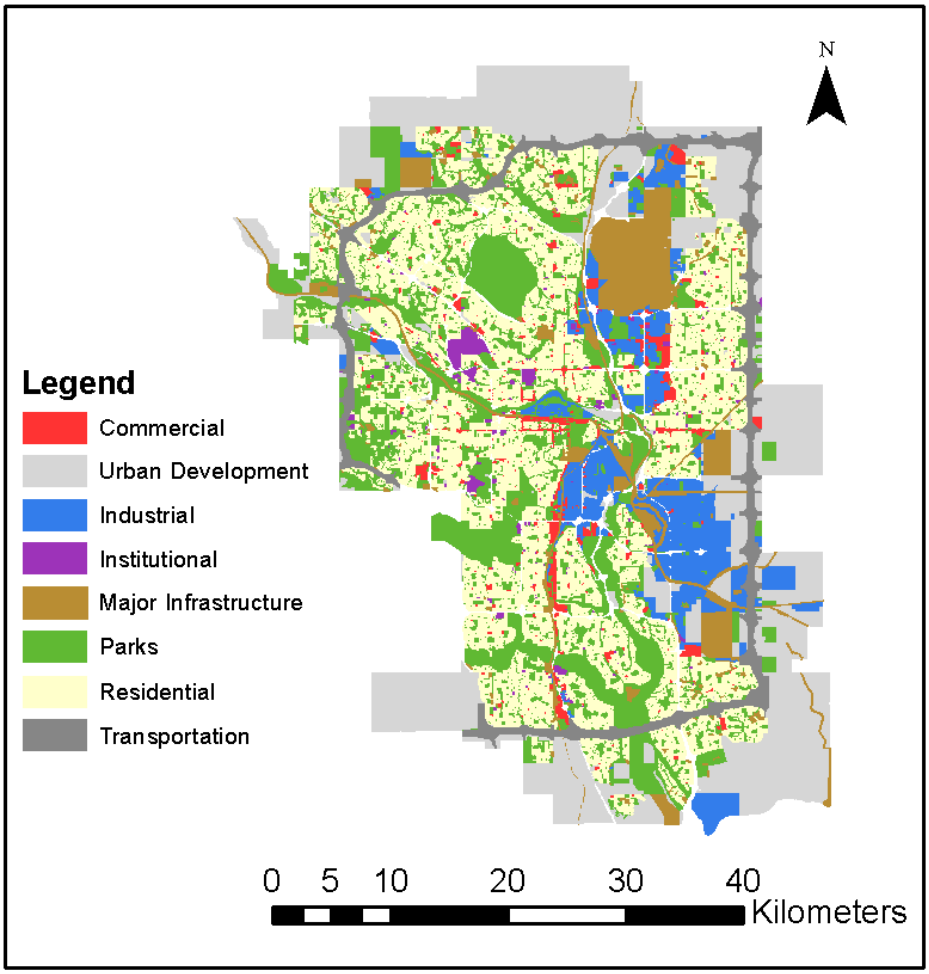


Figure 5-9 Land use of the city of Calgary

Table 5-5 shows the distribution of different land use types in the city of Calgary. Around 35 % of the city is covered by residential types, 30 % is covered by parks, and 25 % is covered by commercial types.

Table 5-5 Land use types distribution in the city of Calgary

Land use type	Percent
Residential	34.29 %
Commercial	25.09 %
Urban development	2.47 %
Industrial	2.45 %
Institutional	2.41 %
Major infrastructure	2.94 %
Parks	30.35 %

(2) POI data of Calgary

The POIs were downloaded from the OSM. There are 17,307 POIs, which were divided into nine category types: 2,304 food, 2,213 recreation, 407 religious, 281 education, 1,072 shopping, 996 daily shopping, 1,721 business services, 312 health services, and 7,749 accommodation. Table 5-6 provides a list of POI category types and the number of POIs in each category.

Table 5-6 POIs and their category types

Category type	Number of POIs
Food	2,304
Recreation	2,213
Religious	407
Education	281
Shopping	1,072
Daily shopping	996
Business services	1,973
Health services	312
Accommodation	7,749

Since some of the opening hours were not currently included in the POIs' data, a time table (Table 5-7) was manually created based on the typical openings of different POIs. Note that the days of the week were divided into Monday-Friday, Saturday, and Sunday. Moreover, MST, which was the minimum service time, was added to the table. This research does not intend to find an optimum value for the MST since it might be varying based on different application domains.

Table 5-7 Opening hours of the POIs in different days

POI	Mon-Fri Opening Hours	Sat Opening Hours	Sun Opening Hours	Category	MST
Bank	9:30 – 17:00	9:00 – 16:00	Closed	Business service	15 min
Shopping mall	9:30 – 21:00	9:30 – 20:00	11:00 – 18:00	Shopping	30 min
Restaurant	11:00 – 23:00	11:00 – midnight	11:00 - midnight	Food	20 min
Post office	9:00 – 18:00	10:00 – 17:00	Closed	Business service	20 min
….	….	….	….	….	….

5.2.2.2 Activity recognition

As described in Section 3.3.2, the activity recognition step is composed of three different steps: data preparation, semantic enrichment process, and activity based ontology model. First, the data were cleaned. Then, stops were identified from the cleansed data and were annotated with probable visited places. Finally, most probable activity performed by the user during each stop was inferred.

5.2.2.2.1 Data preparation

The data preparation procedure was applied to the described collected data. First, trajectories, which were not in the monitored area were removed. Next, the dataset was cleaned from the inconsistencies such as empty values, duplicates and outliers. Below is a summary of data in terms of number of error records were identified.

- Full duplication identical value in all attributes: 251 records.
- Empty value in speed, geometry, and date: 5 records.
- Outliers: 122 pairs of records.

In the trajectory identification, 239 daily basis trajectories were extracted. Unrealistic attributes such as trip durations that are too short were removed. Moreover, 28 Weekly basis trajectories were extracted.

5.2.2.2.2 Semantic enrichment process

The first step in the semantic enrichment process was to extract stops from the cleansed raw trajectory data. The stops were computed via Algorithm 1. Next, the stops were annotated with the land use type (Algorithm 2) and the POI category type (Algorithm 3).

Stop detection

The inputs of this algorithm were cleansed raw trajectory and two parameters Δ_{speed}, and $\Delta_{duration}$. As described in Chapter 3, these parameters play important roles in determining the number of stop episodes. Different ranges for the parameters were considered $\Delta_{duration}$ (from two to ten minutes) and Δ_{speed} (from zero to 20 km/h). Figure 5-10 shows the number of stops for the dataset. With higher $\Delta_{duration}$, the number of stops decreased when given a low Δ_{speed}; whilst with a higher Δ_{speed}, the stop number goes up and saturates, because stops computed with a higher coefficient

Δ_{speed} usually have longer duration. Empirical evaluations suggested to use $\Delta_{speed} < 8$ km/h and $\Delta_{duration} \geq 4$ min to obtain the best accuracy. By visualizing the extracted stops on a map, it turned out that some of the extracted stops were located in the middle of streets, which shows that they have been extracted by mistake.

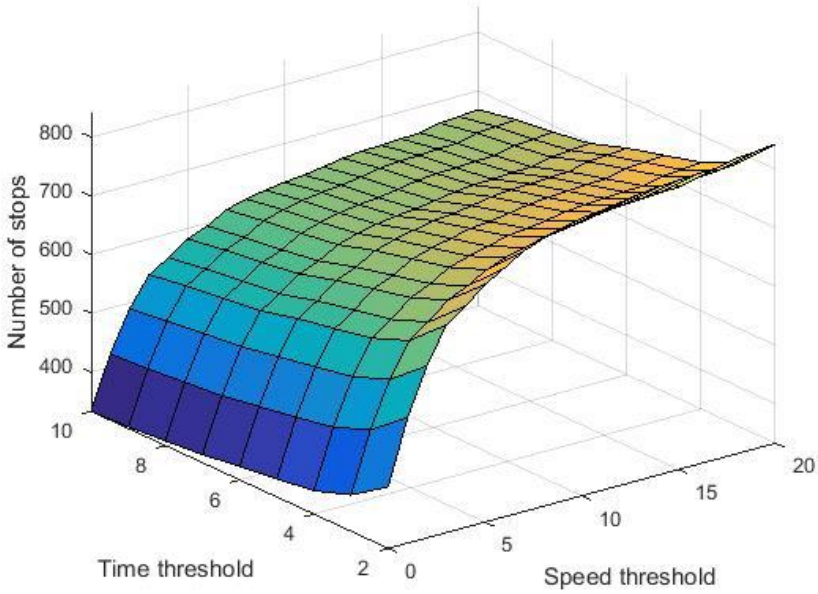


Figure 5-10 Number of stops based on different time durations

After the stops were computed, the moves were implicitly represented as a relationship between two consecutive stops. As a result of the stop detection, 1,237 sub-trajectories with 832 moves and 801 stops over the dataset were produced. As shown in Table 5-8, the number of days recorded in the raw GPS data and also the number of stops in each day is demonstrated. For instance, there were 35 days recorded for Wednesdays and 161 stops were extracted from those days whereas there were 35 days recorded for Tuesdays and only 87 stops were extracted. Although the number of recorded days are the same but the user had more stops on Wednesdays than Tuesdays. This shows that people might have different number of stops on different days.

Table 5-8 Number of days recorded and number of stops in each day

Day of Week	Days	Stops
Monday	39	131
Tuesday	35	87
Wednesday	35	161
Thursday	38	105
Friday	41	123
Saturday	27	113
Sunday	24	81
Sum	239	801

Generally, users might have different number of stops in different days and months. For instance, Figure 5-11 shows the number of stops in two different months, namely, November and December. The user had 12 stops on Nov 3, whereas there were only three stops on Nov 5. As can be seen in the figure, some of the days presented patterns in terms of number of stops.

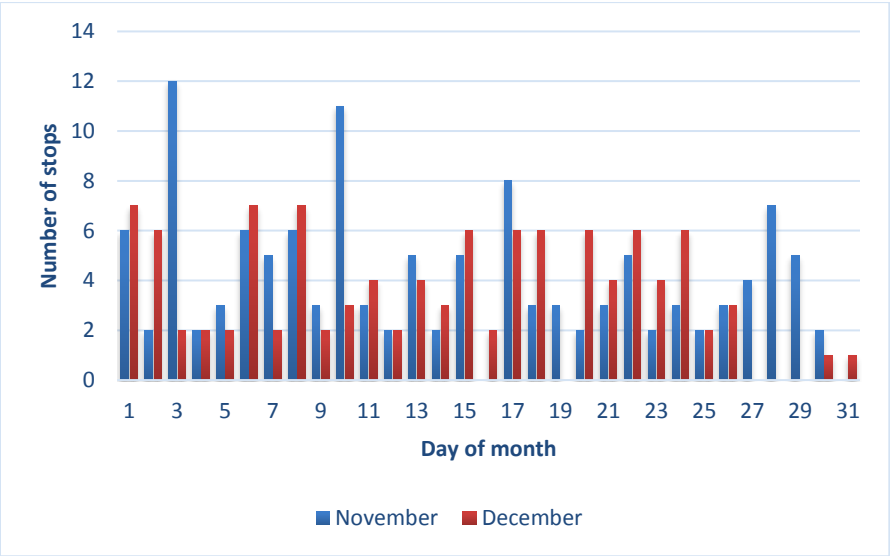


Figure 5-11 Number of stops in different days in November and December

Land use type annotation

Algorithm 2 was used to extract land use types. Figure 5-12 shows the detailed land use type distribution over the trajectories.

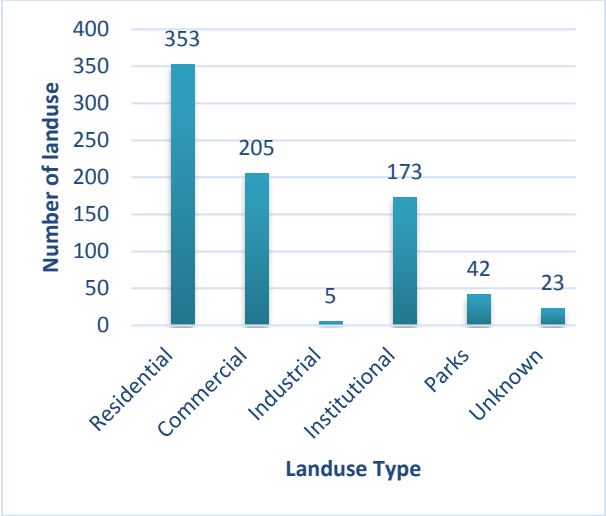


Figure 5-12 Land use type distribution for the user trajectory

Most of the stops were observed in residential areas (44.1%), commercial areas (25.6%), and institutional areas (21.6 %) (Table 5-9).

Table 5-9 Land use type distribution

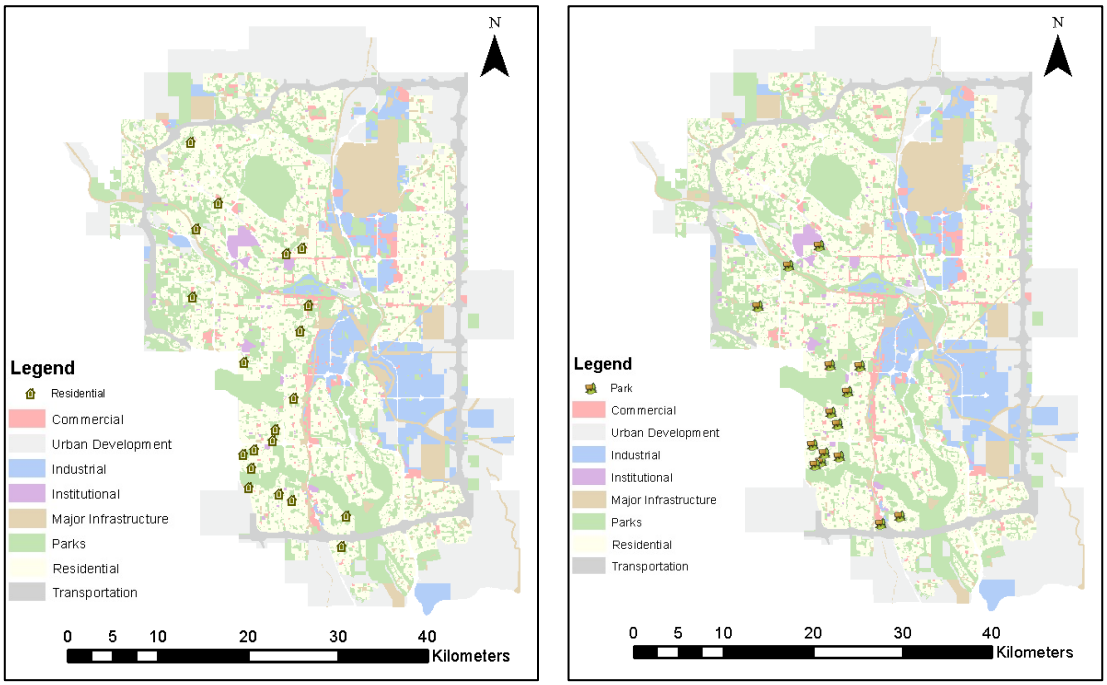
Land use Type	Percent
Residential	44.1
Commercial	25.6
Industrial	0.6
Institutional	21.6
Parks	5.2
Unknown	2.9

Table 5-10 shows the number of different land use types that the user had stopped at. For instance, the user had stopped in three different institutional types whereas he had stopped in 42 different commercial types. As it can be seen in the table, the stops in the commercial land use type have a wide variety.

Table 5-10 Number of different land use types that the user has stopped

Stop Type	Place type
Residential	20
Commercial	42
Institutional	3
Out of town	6
Park	15
Industrial	6

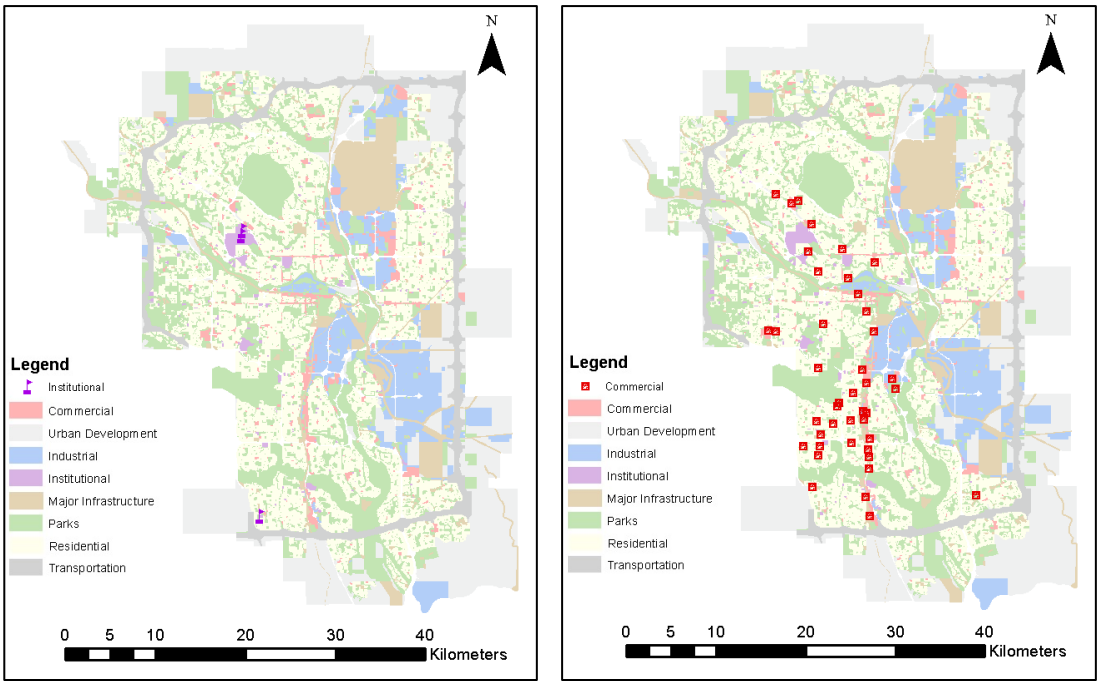
Figure 5-13 shows the annotated stops with different land use types; (a) residential and (b) parks. The residential type was annotated in 20 different places, whereas the parks type was annotated in 15 different places. As seen in Figure 5-13, the majority of the places were located in the west of the city.



(a) (b)

Figure 5-13 Annotated stops with different land use types; (a) Residential and (b) Parks

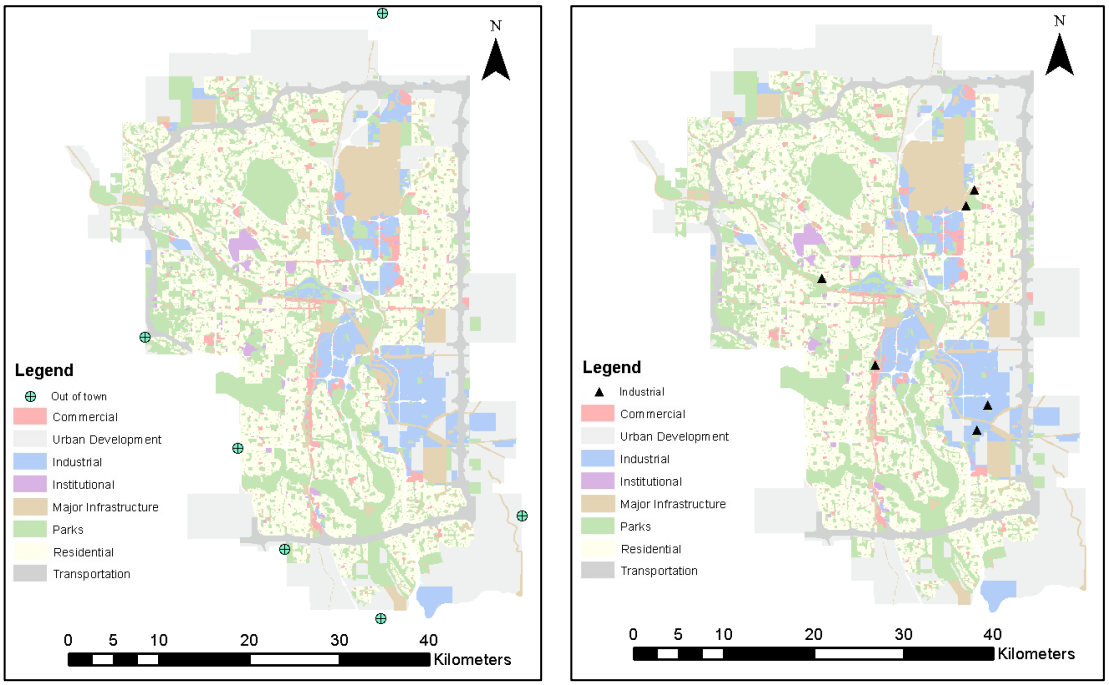
Figure 5-14 shows the annotated stops with different land use types; (a) institutional and (b) commercial. The institutional type was annotated in three different places, whereas the commercial type was also annotated in 42 different places.



(a) (b)

Figure 5-14 Annotated stops with different land use types; (a) Institutional and (b) Commercial

Figure 5-15 shows the annotated stops with different land use types; (a) out of town and (b) industrial. The out of town type was annotated in six different places, whereas the industrial type was also annotated in six different places.



(a) (b)

Figure 5-15 Annotated stops with different land use types; (a) Out of town and (b) Industrial

Figure 5-16 shows all the stops annotated with different land use types. As it can be seen, most of the stops have happened in the area where the user lives.

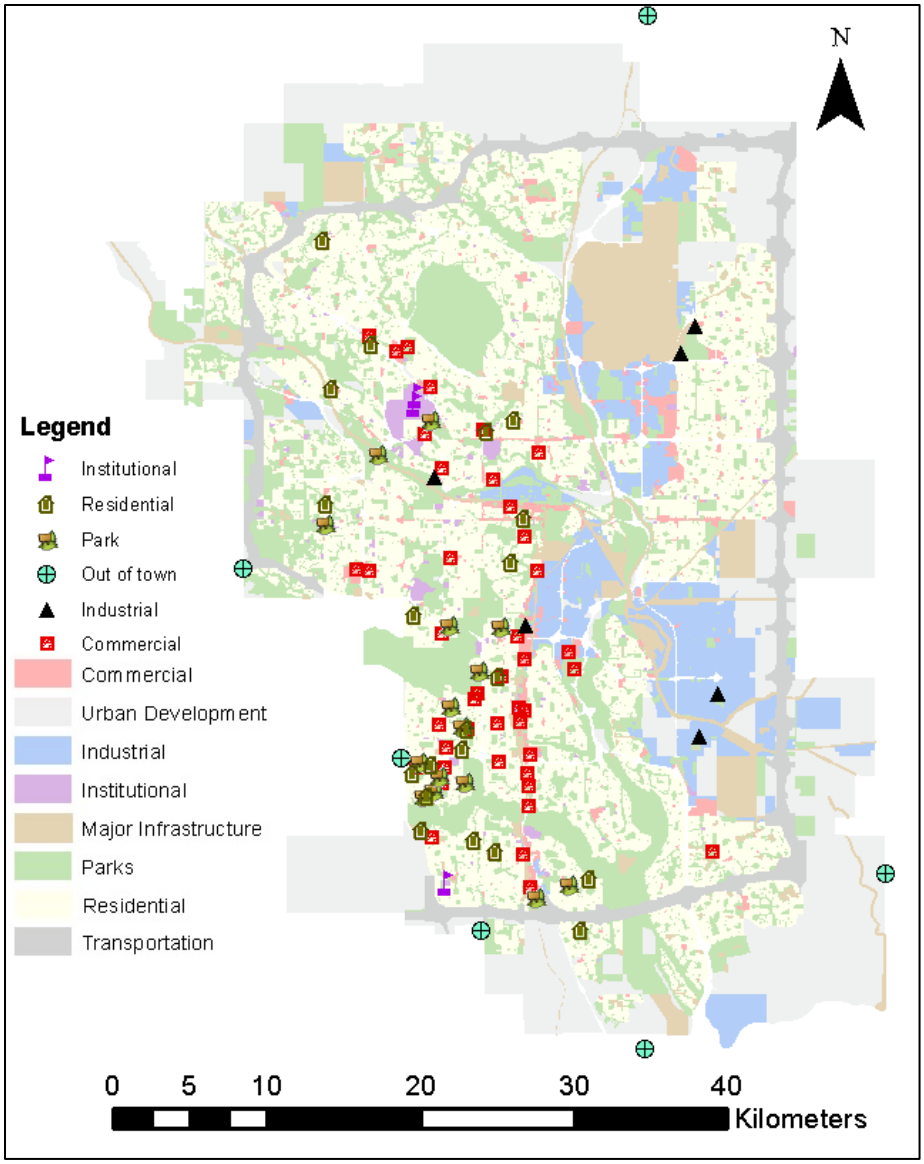


Figure 5-16 All detected stops in different land use types

POI category type annotation

As explained in Sub-section 3.3.4.1, Algorithm 3 was used to extract the most probable POI category type for each stop. UWS was considered 5 km/h and UWD was considered 100 metres as discussed earlier in the simulated dataset. Therefore, each stop was buffered with a distance of 100 m (distance which is likely to be accepted for a walk from car's parking spot to a place) in order to extract general POIs. As shown in Figure 5-17, most of the stops belonged to the shopping (30.1%), business services (18.9%) and food (17.7%) categories.

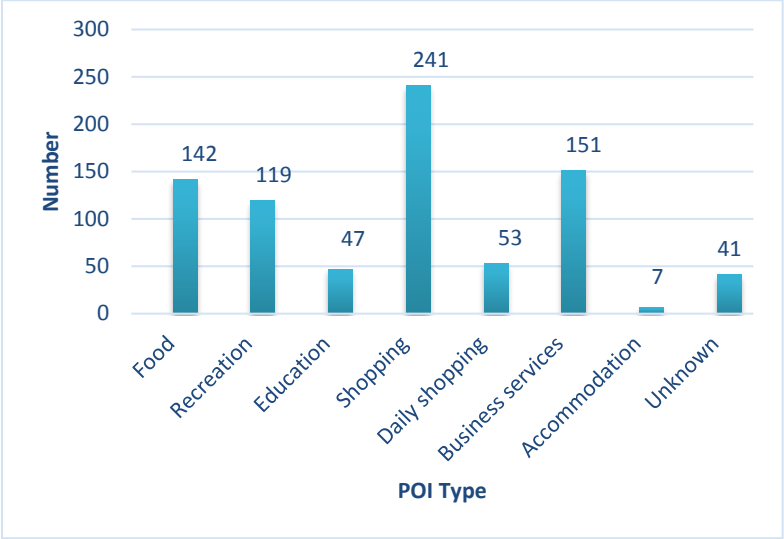


Figure 5-17 Number of POI types assigned to the stop trajectories

Table 5-11 represents the POI category type annotation in which probabilities are returned using the gravity formula (3-5). For instance, for the stop S_1 two different category types were computed; food category type with the probability of 0.65 and the business service category type with the probability of 0.35.

Table 5-11 The most probable POI category type

Stop	POI Category Type	Probability
S_1	Food	0.65
S_1	Business service	0.35
S_2	-	-
S_3	Recreation	1

5.2.2.2.3 Ontology based activity model

The ontology based activity model consists of different ontologies such as stop, time, place, and activity ontologies. The time ontology contained the temporal discretization such as absolute intervals as shown in Table 5-12, start time, end time, and duration were also included.

Table 5-12 Temporal discretization of time ontology

Time Definition	
Morning	4:00 AM - 11:59 AM
Afternoon	12:00 PM - 4:59 PM
Evening	5:00 PM - 8:59 PM
Night	9:00 PM - 3:59 AM

The stop ontology contained meaningful stops of the user, which means the user might have an activity to do. Each stop has some features such as frequency, begin time, and average duration in the place as shown in Table 5-13. For instance, the user had stopped six times a week in a residential land use type named "Residential1" in the evenings with the average of 614 minutes per week.

Table 5-13 Stops, stop frequency and average stayed time in the stop ontology

Tuesday			
Stops	**Frequency**	**Begin time**	**Average (min)**
Commercial1	1	Evening	221
Commercial29	2	Night	14
Institutional1	**5**	**Morning**	**441**
Residential1	**6**	**Evening**	**614**
Residential14	1	Evening	22

The place ontology contained the POIs and the land use types as shown in Table 5-14. For instance, for the stop S_1, commercial was assigned as land use type and two different POI category types were computed; food category type with the probability of 0.65 and the business service category type with the probability of 0.35.

Table 5-14 POI and land use in the place ontology

Stop	Land use Type	POI Category Type	Probability
S_1	Commercial	Food	0.65
S_1	Commercial	Business service	0.35
S_2	Residential	-	-
S_3	Parks	Recreation	1

Activity type's inference

The ontology model was populated with the above ontologies and the reasoning step executed the reasoner using the axioms that had been defined in Chapter 3 for each activity type. Table 5-15 shows some inferred activity types using the axioms on the ontology model.

Table 5-15 Some of the inferred activity types

Land use Type	POI Category Type	Features			Activity Type
		T_b	S_f	S_d	
Residential	-	Evening	6 days per week	10.2 hours	Return Home
Residential	-	Evening	1 day per week	45 min	Visiting
Commercial	Shopping	Afternoon	2 days per week	41 min	Shopping
Institutional	-	Morning	5 days per week	8.2 hours	Go to work

5.2.3 Tehran dataset

The second experiment was performed over the Tehran's dataset, which was acquired in four months within the urban area in the City of Tehran in 2015. As it can be seen in Figure 5-18 the November month had the most data recorded. It is a total of around 325,152 GPS records.

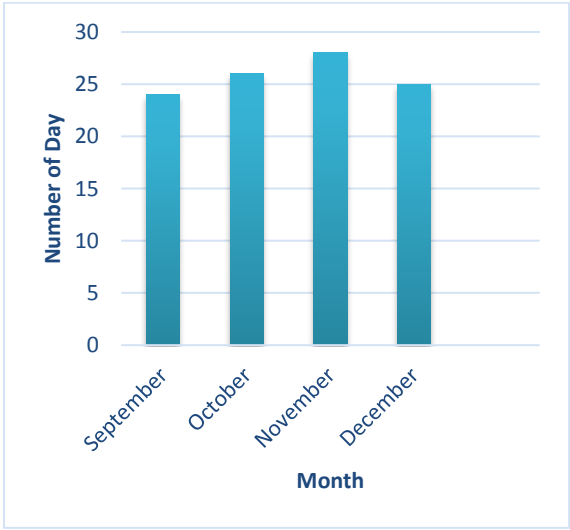


Figure 5-18 Tehran's raw GPS data acquired in four months

Figure 5-19 visualizes the Tehran's dataset and shows the routes that the user is used to complete various activity types.

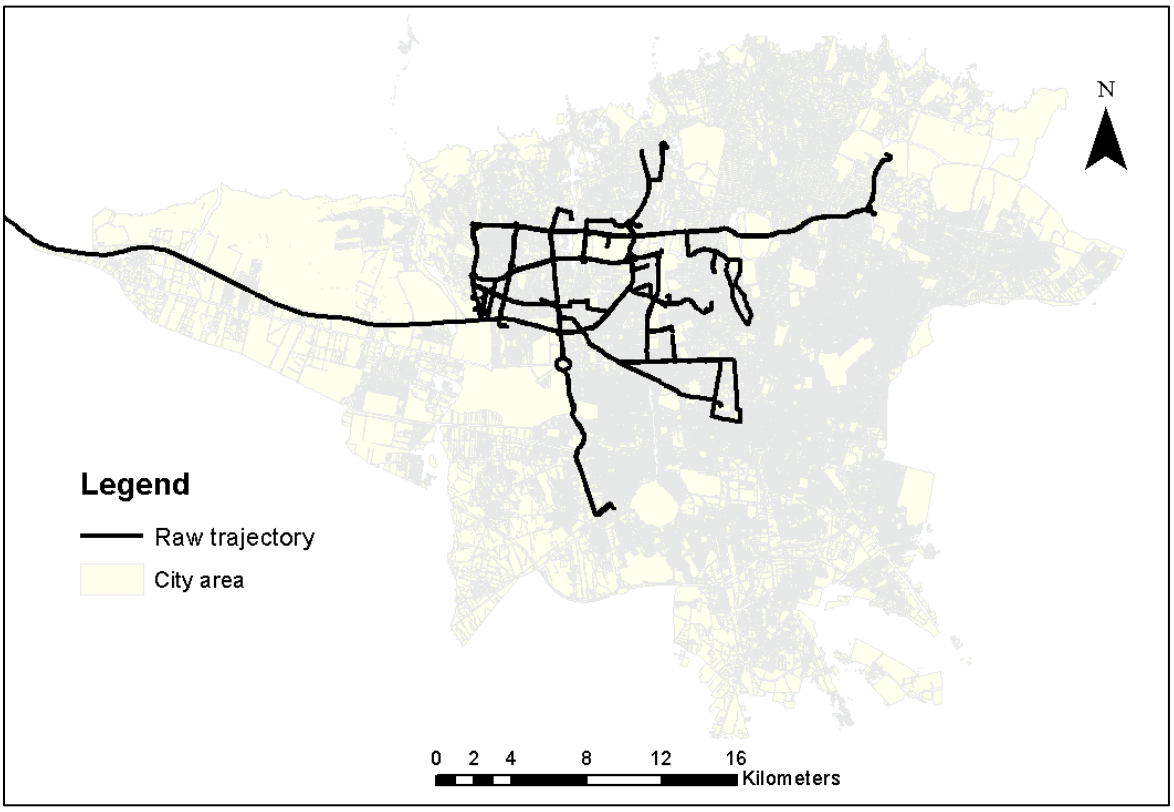


Figure 5-19 Visualization of the Tehran's dataset

5.2.3.1 Data sources

(1) Land use data of Tehran;

As shown in Figure 5-20, the land use data includes different types such as commercial, urban development, residential, institutional, industrial, parks, major infrastructure, and transportation. The land use is predominantly residential, with most industrial uses in the southern half of the city. Commercial and green space types are spread throughout the city.

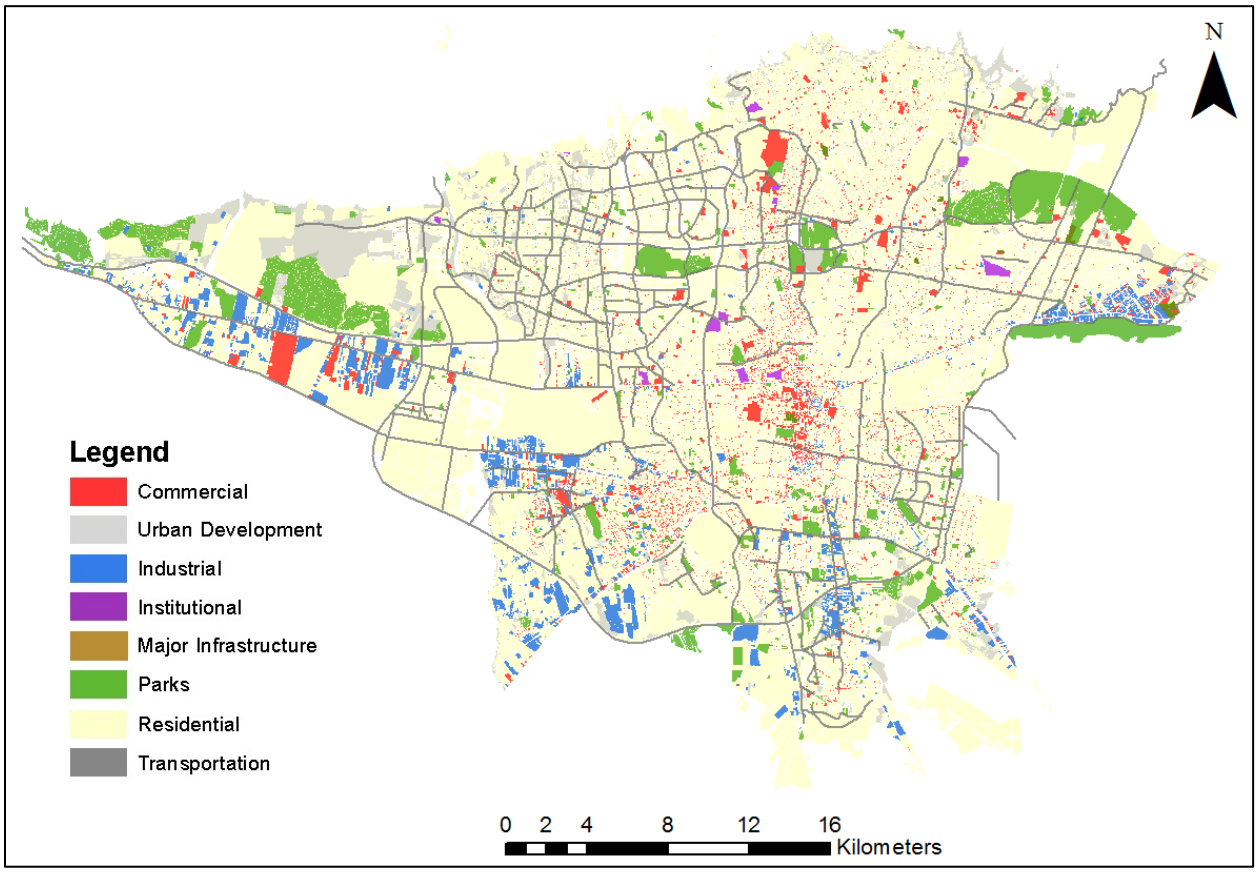


Figure 5-20 Land use of the city of Tehran

Table 5-16 shows the distribution of different land use types in the city of Tehran. Around 29 % of the city is covered by residential types, 20 % is covered by commercial, and 12 % is covered by parks types.

Table 5-16 Land use types distribution in the city of Tehran

Land use type	Percent
Residential	28.8 %
Commercial	19.6 %
Urban development	6.1 %
Industrial	4.4 %
Institutional	4.3 %
Major infrastructure	2.94 %
Parks	11.7 %

(2) POI data of Tehran

There were 268,122 POIs, which were divided into 9 category types: 42,161 food, 22,101 recreation, 7,015 religious, 16,158 education, 12,269 shopping, 11,562 daily shopping, 84,015 business services, 2,603 health services, and 61,532 residential area. Table 5-17 provides a list of POI category types and the number of POIs in each category.

Table 5-17 POIs and their category types

Category type	Number of POIs
Food	42,161
Recreation	22,101
Religious	7,015
Education	16,158
Shopping	12,269
Daily shopping	11,562
Business services	92,721
Health services	2,603
Accommodation	61,532

5.2.3.2 Activity recognition

Data preparation, semantic enrichment process, and activity based ontology model was also applied for the Tehran dataset. First, the data was cleaned. Then, stops were identified from the cleansed data and were annotated with probable visited places. Finally, most probable activity performed by the user during each stop was inferred.

5.2.3.2.1 Data preparation

The data preparation procedure was also applied to the Tehran data. Below is a summary of data in terms of number of error records that have been identified.

- Full duplication identical value in all attributes: 71 records.
- Empty value in speed, geometry, and date: 2 records.
- Outliers: 19 pairs of records.

In the trajectory identification, 103 daily basis trajectories were extracted.

5.2.3.2.2 Semantic enrichment process

The first step in the semantic enrichment process was to extract stops from the cleansed raw trajectory data. The stops were computed via Algorithm 1. Next, the stops were annotated with the land use type (Algorithm 2) and the POI category type (Algorithm 3).

Stop detection

Different ranges for the parameters were considered $\Delta_{duration}$ (from two to ten minutes) and Δ_{speed} (from zero to 20 km/h). Figure 5-21 shows the number of stops for the dataset. Empirical evaluations suggested to use $\Delta_{speed} < 8$ km/h and $\Delta_{duration} \geq 7$ min to obtain the best accuracy. The value of the duration threshold was higher since the traffic flow of city was higher than the Calgary's dataset.

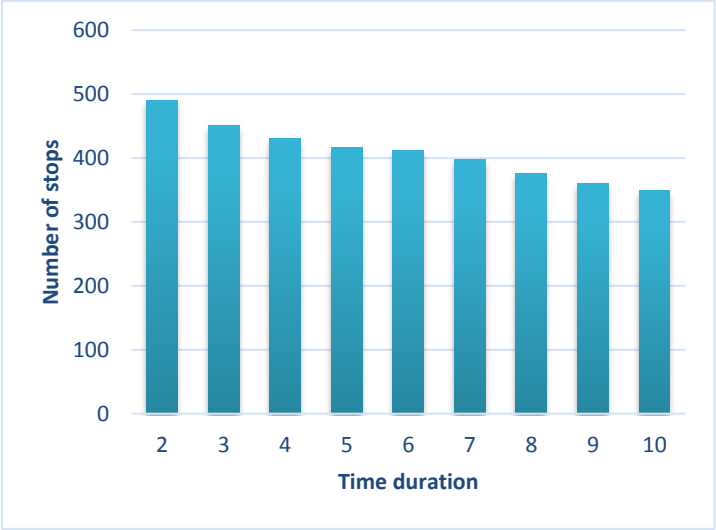


Figure 5-21 Number of stops based on different time durations

As a result of the stop detection, 421 sub-trajectories with 409 moves and 398 stops over the dataset were produced. As shown in Table 5-18, the number of days recorded in the raw GPS data and also the number of stops in each day is demonstrated. For instance, there were 16 days recorded for Monday and 69 stops were extracted from those days whereas there were 16 days recorded for Tuesdays and only 49 stops were extracted. Although the number of recorded days are the same but the user had more stops on Monday than Tuesdays.

Table 5-18 Number of days recorded and number of stops in each day

Day of Week	Days	Stops
Monday	16	69
Tuesday	16	49
Wednesday	15	55
Thursday	14	60
Friday	10	57
Saturday	16	56
Sunday	16	52
Sum	103	398

Land use type annotation

Algorithm 2 was used to extract the land use types. Figure 5-22 shows the detailed land use type distribution over the trajectories.

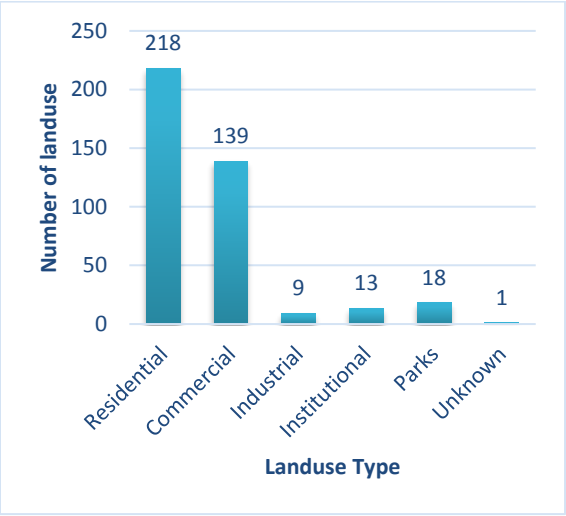


Figure 5-22 Land use type distribution for the user trajectory

Most of the stops were observed in residential areas (55 %), commercial areas (35 %), and park areas (4.5 %) (Table 5-19).

Table 5-19 Land use type distribution

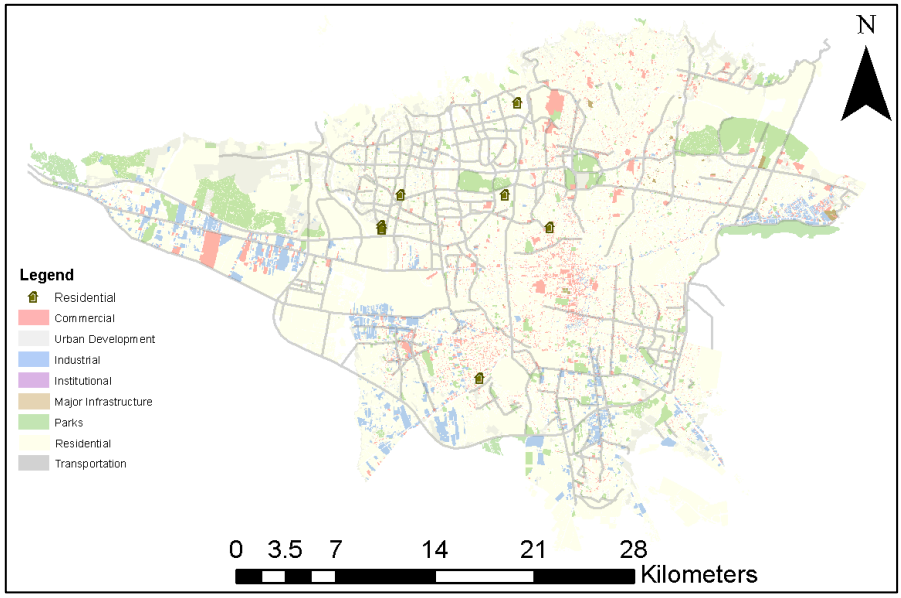
Land use Type	Percent
Residential	54.8
Commercial	34.9
Industrial	2.3
Institutional	3.3
Parks	4.5
Unknown	0.3

Table 5-20 shows the number of different land use types that the user had stopped at. For instance, the user had stopped in 8 different residential types whereas he had stopped in 29 different commercial types. As it can be seen in the table, the commercial type has a wide variety.

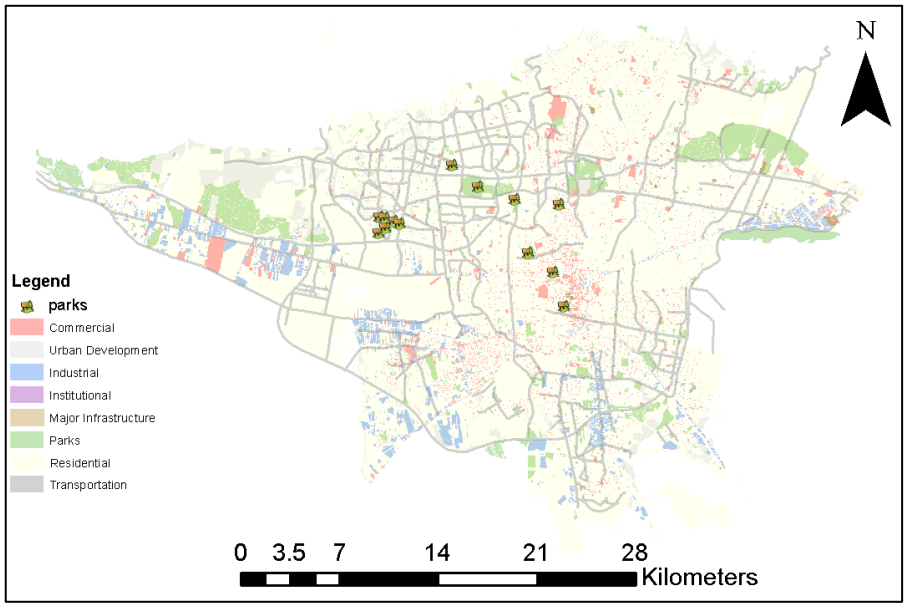
Table 5-20 Number of different land use types that the user was stopped

Stop Type	Place type
Residential	8
Commercial	29
Institutional	6
Out of town	1
Park	13
Industrial	5

Figure 5-23 shows the annotated stops with different land use types; (a) residential and (b) parks. The residential type was annotated in eight different places, whereas the parks type was annotated in 13 different places. As seen in Figure 5-23, the majority of the places were located in the center of the city. This could be inferred that the user either lives or works in the center of the city.



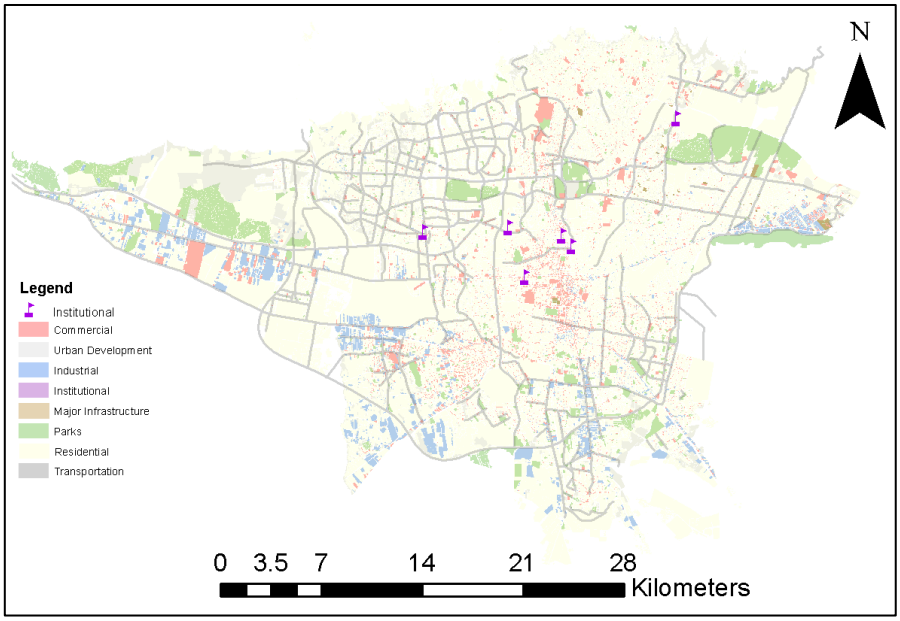
(a)



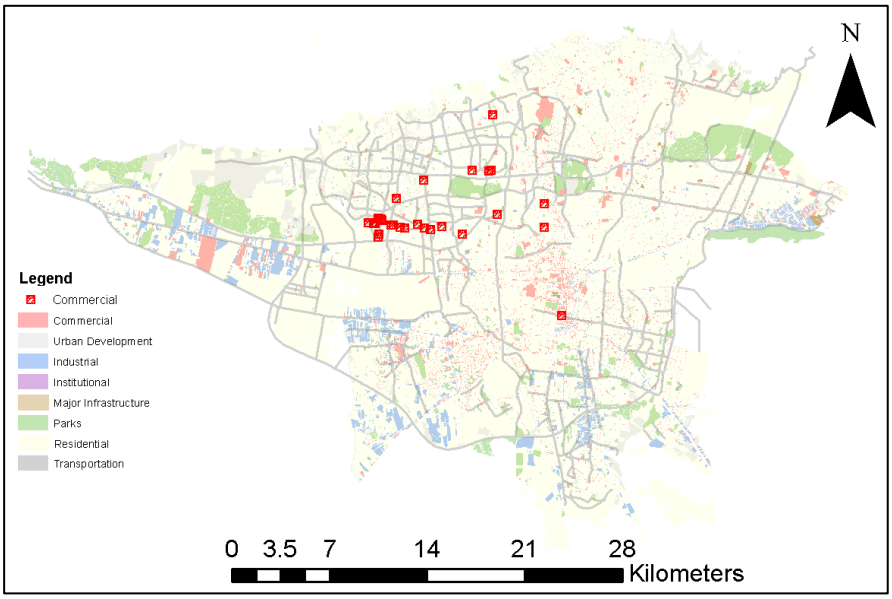
(b)

Figure 5-23 Annotated stops with different land use types; (a) Residential and (b) Parks

Figure 5-24 shows the annotated stops with different land use types; (a) institutional and (b) commercial. The institutional type was annotated in six different places, whereas the commercial type was also annotated in 29 different places.



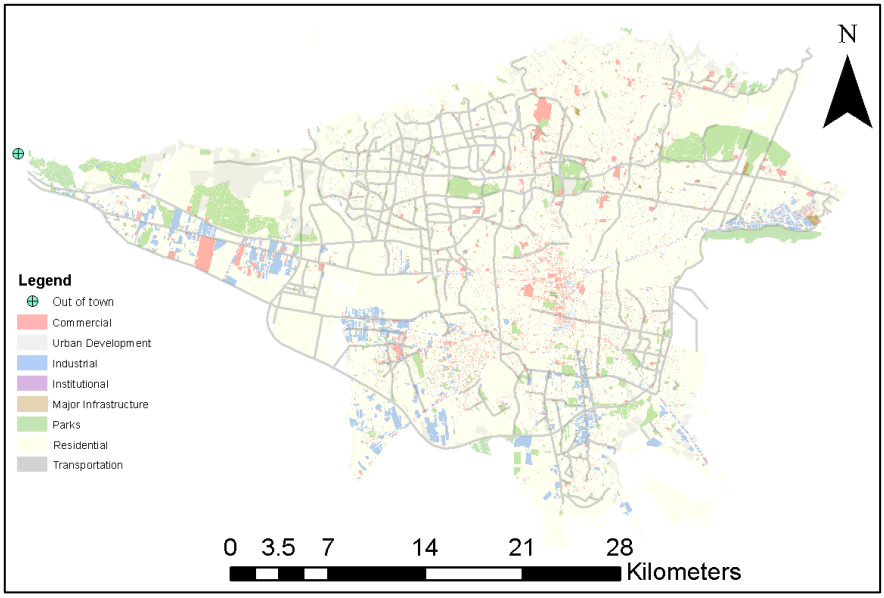
(a)



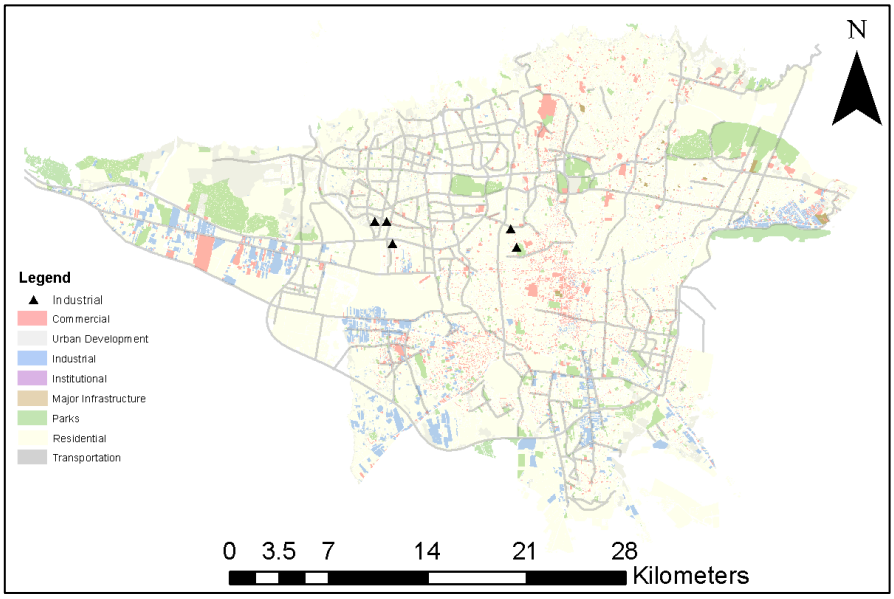
(b)

Figure 5-24 Annotated stops with different land use types; (a) Institutional and (b) Commercial

Figure 5-25 shows the annotated stops with different land use types; (a) out of town and (b) industrial. The out of town type was annotated in one place, whereas the industrial type was also annotated in five different places.



(a)



(b)

Figure 5-25 Annotated stops with different land use types; (a) Out of town and (b) Industrial

Figure 5-26 shows all the stops annotated with different land use types. As it can be seen, most of the stops have happened in the area where the user lives.

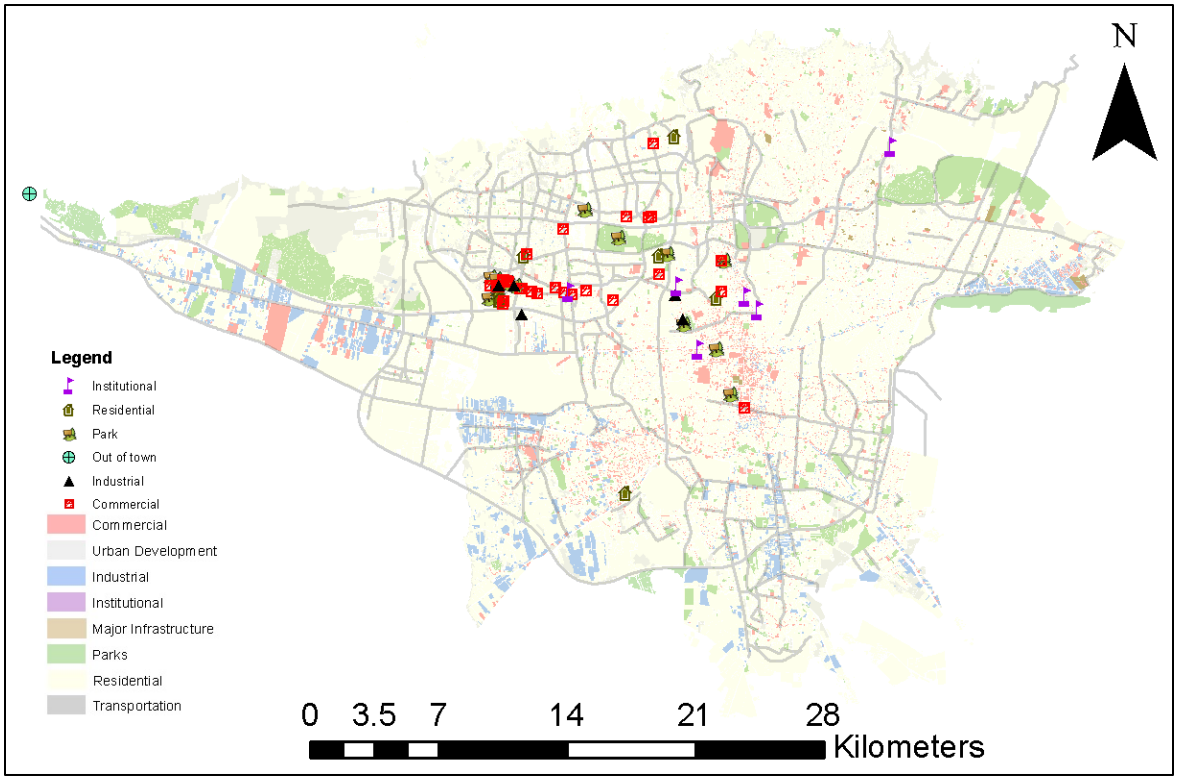


Figure 5-26 All detected stops in different land use types

POI category type annotation

As explained in Sub-section 3.3.4.1, Algorithm 3 was used to extract the most probable POI category type for each stop. UWS was considered 5 km/h and UWD was considered 100 metres. As shown in Figure 5-27, most of the stops belonged to the food (26.4%), shopping (25.4%) and business services (20.4%) categories.

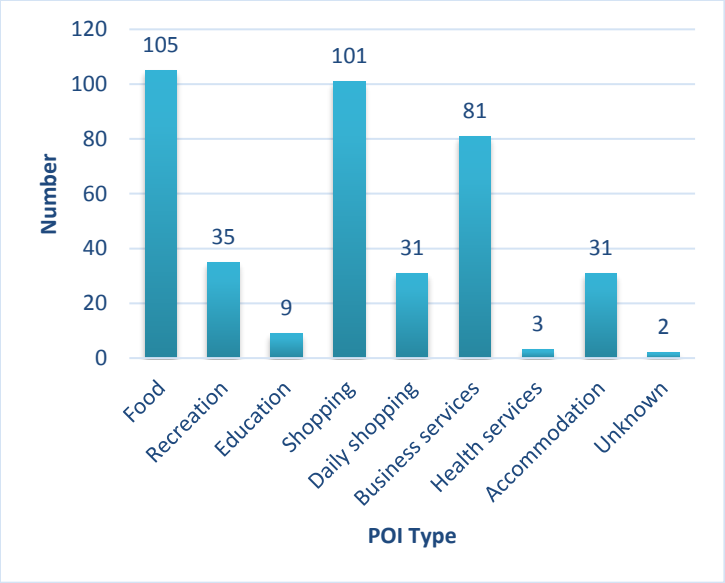


Figure 5-27 Number of POI types assigned to the stop trajectories

5.2.3.2.3 Ontology based activity model

Activity type's inference

The ontology model was populated with the above ontologies and the reasoning step executed the reasoner using the axioms that had been defined in Chapter 3 for each activity type. Table 5-21 shows some inferred activity types using the axioms on the ontology model.

Table 5-21 Some of the inferred activity types

Land use Type	POI Category Type	Features			Activity Type
		T_b	S_f	S_d	
Residential	-	Evening	1 day per week	215 min	Recreational
Commercial	Business service	Afternoon	2 days per week	181 min	Business services
Commercial	Shopping	Morning	6 days per week	9.1 hours	Go to work
Residential	-	Evening	6 days per week	11.4 hours	Return Home

5.2.4 Semantic behavior modelling

As shown in Table 5-22, the activities were detected using the ontology based activity model.

Table 5-22 Semantic trajectory of the user

Date	Start	Time_1	End	Time_2	Day	Activity duration [hh:mm]
8-Jun	Home	Morning	Work	Morning	Monday	9:13
8-Jun	Work	Afternoon	Shopping	Evening	Monday	2:01
8-Jun	Shopping	Evening	Home	Evening	Monday	11:05
22-Jun	Home	Morning	Work	Morning	Monday	8:26
22-Jun	Work	Evening	Home	Evening	Monday	0:15
22-Jun	Home	Evening	Shopping	Evening	Monday	0:17
22-Jun	Shopping	Evening	Business services	Evening	Monday	3:02

Next step is to find the association between the user's activity types. Different behaviors were defined based on different attributes such as spatial, temporal, and semantic. The next step is to use apriori to find rulesets. These rulesets are defined as axioms in the ontology model.

5.2.4.1 Data preprocessing

In the data preprocessing step, temporal information was generalized to the semantics of time (see Table 5-23) to aid identification of overlaps between activity types, and to simplify/reduce the number of patterns extracted. For instance, the period time of 4:00 AM to 11:59 AM was labeled as morning.

Table 5-23 Temporal discretization of time ontology

Semantic time	Time period
Morning	4:00 AM - 11:59 AM
Afternoon	12:00 PM - 4:59 PM
Evening	5:00 PM - 8:59 PM
Night	9:00 PM - 3:59 AM

Moreover, the activity types were codified to a machine-readable format suitable for further analysis (see Table 5-24), i.e., the activity type named, "Eating" was coded "Ea", "Return home" was coded "Rh", etc.

Table 5-24 List of codified activity types

Activity type	Code
Eating	Ea
Recreational	Re
Education	Ed
Shopping	S
Daily shopping	Ds
Business services	Bs
Go to work	G
Trip	T
Socializing	So
Return home	Rh

5.2.4.2 Association rule mining

Apriori algorithm was applied to determine support for, and confidence in a particular pattern. Minimum support and confidence criteria generally vary according to the types of patterns to be identified. A number of association rules found in the Calgary's semantic trajectories using minimum support (3-7) of 2, and a minimum confidence (3-8) of 60%. All the extracted rulesets were stored into the data repository and imported into the ontology model. Different rulesets are shown below from different behavior types. Table 5-25 lists a number of rules extracted from semantic attributes. For instance, if the user performed a recreational activity, there was 2.3 % support and 66.7 % confidence that the user will return home or if the user performed a work activity, there was 26 % support and 78 % confidence that the user will return home.

Table 5-25 A number of extracted association rules developed from semantic attributes

Rules	Support	Confidence
If Recreational THEN Return home	2.3%	66.7%
If Shopping THEN Return home	22.1%	62.1%
If Work THEN Return home	26%	78%

Table 5-26 lists a number of rules extracted from semantic and time attributes. For instance, if the user performed a work activity in the afternoon, there was 3.1 % support and 88.1 % confidence that the user will return home or if the user was at home in the evening, there was 12.2 % support and 75 % confidence that the user will go shopping.

Table 5-26 A number of extracted association rules developed from semantic and time attributes

Rules	Support	Confidence
If Work – Afternoon THEN Return home	3.1%	88.1%
If Home - Morning THEN Work	26%	94.1%
If Home - Evening THEN Shopping	12.2%	75%

Table 5-27 lists a number of rules extracted from semantic and space attributes. For instance, if the user performed a shopping activity at location number 39, there was 6.1 % support and 100 % confidence that the user will return home or if the user was performed a socializing activity at location number 3, there was 2.3 % support and 100 % confidence that the user will return home.

Table 5-27 A number of extracted association rules developed from semantic and space attributes

Rules	Support	Confidence
If Shopping_39 THEN Return home	6.1%	100%
If Shopping_10 THEN Shopping_39	6.9%	66.7%
If Socializing_3 THEN Return home	2.3%	100%

Table 5-28 lists a number of rules extracted from semantic and space-time attributes. For instance, if the user performed a shopping activity at location number 39 in the evening, there was 5.3 % support and 88.9 % confidence that the user will return home or if the user was at home in the evening, there was 11.45 % support and 68.5 % confidence that the user will go shopping at location number 10.

Table 5-28 A number of extracted association rules developed from semantic and space-time attributes

Rules	Support	Confidence
If Shopping_39 – Evening THEN Return home	5.3%	88.9%
If Return home - Evening THEN Shopping_10	11.45%	68.5%
If Socializing_3 – Evening THEN Return home	2.3%	100%

5.3 Research Evaluation

This sub-section presents the evaluation of the activity recognition and semantic behavior modelling steps.

5.3.1 Activity recognition evaluation

Figure 5-28 illustrates a web interface application to visualize daily semantic trajectories and collecting users' feedback, which give the ground truth to validate the proposed methodology for activity recognition in this research. The users were asked to provide their feedback weekly via the website. The interface displays three parts: (a) the visualization with Leaflet map API[4], (b) the date that allows users to select a day to see the extracted activity types related to the day, and (c) the list of activity types associated with the trajectory, which consist of time, type of place, duration and activity type. In the bottom of the box, the user is asked to verify if the inferred activity type is correct or not. If it is not correct and users click on the "No" button, then they are asked to enter the real activity type that they have done. If they select "No stop", then this indicates that no stop was occurred at all.

[4] Application Programming Interface

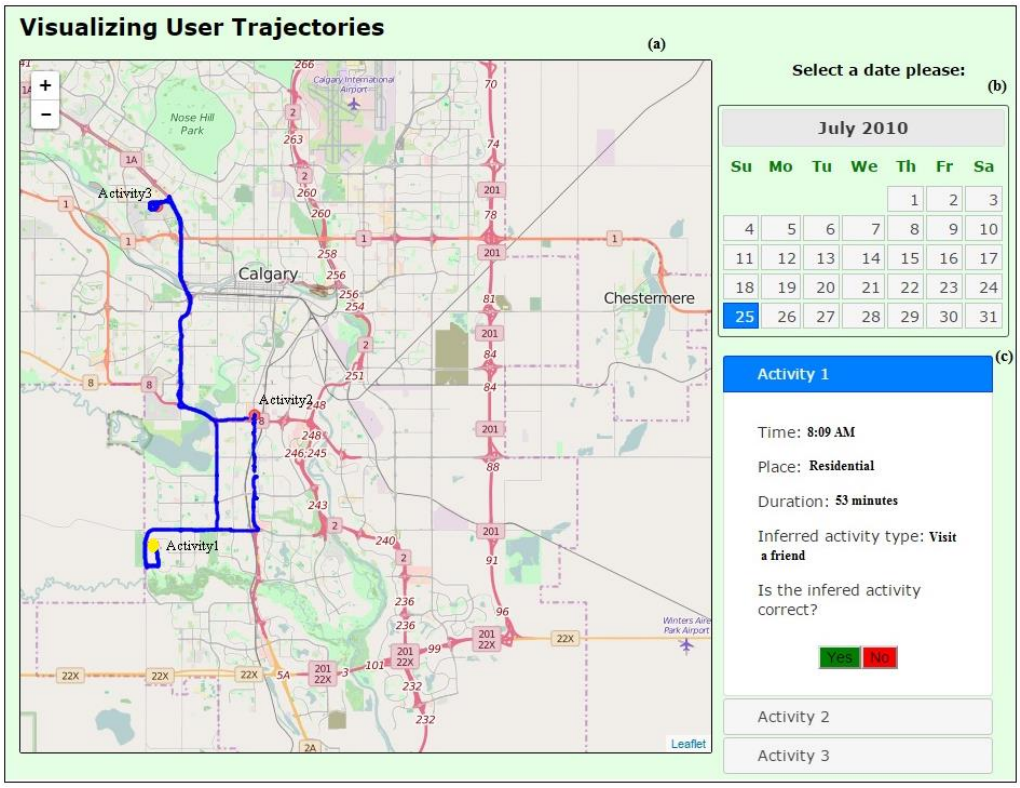


Figure 5-28 User interface to visualize user's trajectories in order; to get his/her feedback

As shown in Figure 5-28, for instance, a user on July 25th, 2010 had 3 different activity types displayed with red circles on the map. The selected activity type (activity1) is shown in yellow circle. As shown in Figure 5-29, the first stop occurred in a residential area.

Figure 5-29 The first activity of the user shows the stop in a residential area on March 4th 2010 displayed on the map

Therefore, the inferred activity types using the proposed method are compared to the collected feedback of the users. The accuracy is the number of correctly inferred activity types over the number of total inferred activity types from the dataset.

Calgary's dataset

The experimental outcome and the evaluation results of the Calgary's dataset are depicted in Table 5-29. It shows the accuracies per activities i.e. the percentage of activities correctly identified w.r.t. the number of declared activities (of the same type). For example, good results for activities of type "business services" (the method recognized 97.3 % of them) were obtained, while the method was unable to identify "daily shopping" (the method recognized 35.9 % of them). The average accuracy of the proposed method was 83 %. In this dataset, the accuracy of 80 % of the activity types were above the average accuracy.

Table 5-29 Accuracy of extracted activities using user's feedback for Calgary dataset

Activity Type	Accuracy (%)
Eating	88.4
Recreational	86.1
Education	69.5
Shopping	91.3
Daily shopping	35.9
Business services	97.3
Go to work	93.8
Trip	88.6
Socializing	90.6
Return home	89.1

Tehran's dataset

Table 5-30 depicts the accuracies per activities for Tehran's dataset. Good results were obtained for the activity type "eating" (the method recognized 92.4 % of them), while the method was unable to identify "daily shopping" (the method recognized 55.9 % of them) the same as the Calgary's dataset. The overall accuracy of the proposed method was 82 %. In this dataset, the accuracy of 60 % of the activity types were above the average accuracy.

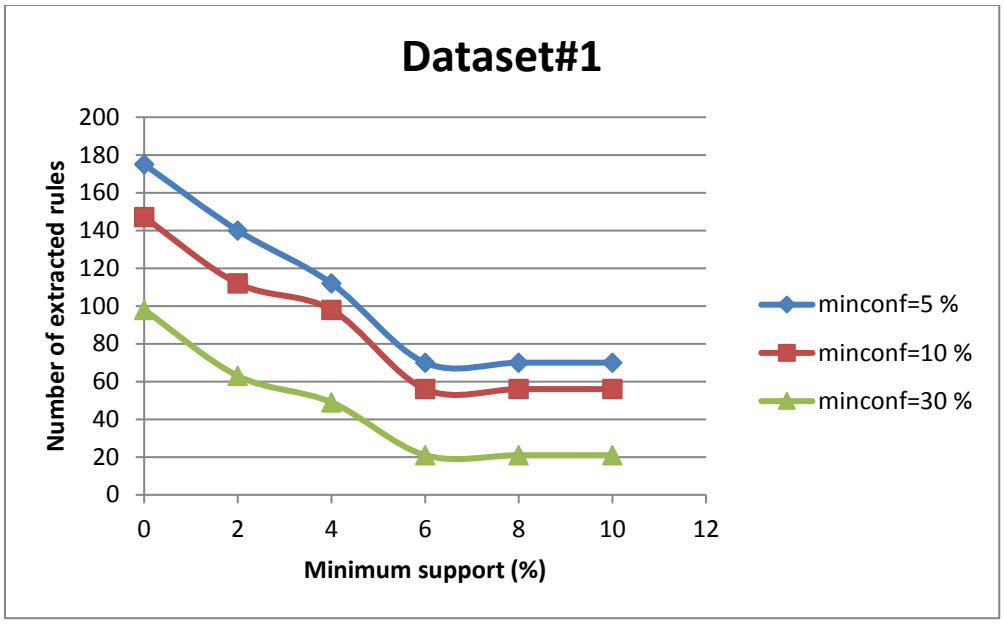
Table 5-30 Accuracy of extracted activities using user's feedback for Tehran dataset

Activity Type	Accuracy (%)
Eating	92.4
Recreational	88.2
Education	92.3
Shopping	71.3
Daily shopping	55.9
Business services	84.2
Go to work	92.1
Trip	66.6
Socializing	88.1
Return home	80.6

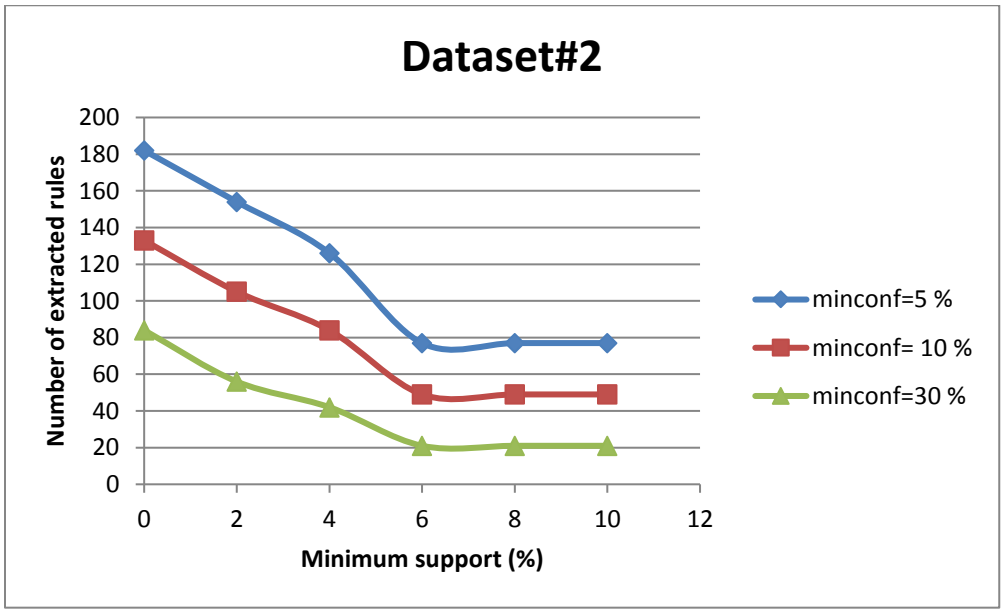
It was observed that the accuracy of the inferred activity types was related to the availability of the POIs around the stops. The more POIs were available around the stops, the better accuracy were obtained for the activity types, which were related to the POI category types. For instance, since there were few POIs for the daily shopping category type in both datasets, the minimum accuracy was obtained in the proposed activity recognition approach.

5.3.2 Behavior modelling evaluation

In order to evaluate the extracted semantic behavior rules, the dataset was divided into two datasets. Every even row and odd row in the original dataset were assigned as dataset#1 and dataset#2, respectively. In this regard, each dataset was used in order to generate rules by considering different thresholds for the minimum support (0 %, 2 %, 4 %, 6 %, 8 %, and 10 %) and the minimum confidence (5 %, 10 %, and 30 %). As shown in Figure 5-30, semantic and time behavior rules were extracted from dataset#1 (Figure 5-30-(a)) and dataset#2 (Figure 5-30-(b)). As it can be observed, the impact of the support and confidence thresholds on the number of extracted rules were pretty much the same for both datasets.



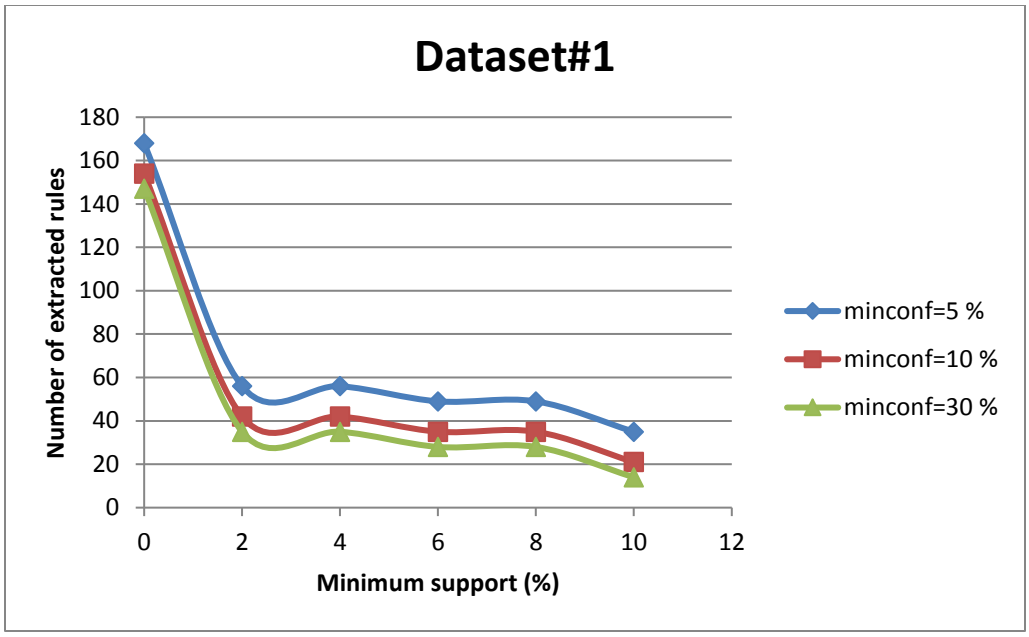
(a)



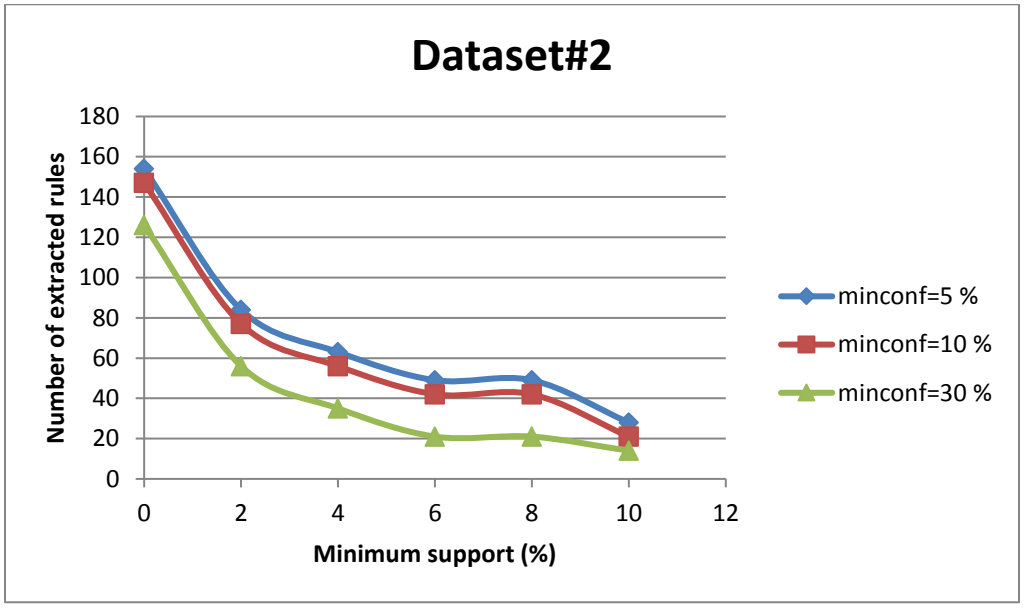
(b)

Figure 5-30 Number of extracted semantic and time rules by considering different support and confidence threshold

As shown in Figure 5-31, semantic and space behavior rules were extracted from dataset#1 (Figure 5-31-(a)) and dataset#2 (Figure 5-31-(b)). the impact of the support and confidence thresholds on the number of extracted rules were pretty much the same for both datasets except when the minimum support was between 2 and 4 %.



(a)



(b)

Figure 5-31 Number of extracted semantic and space rules by considering different support and confidence threshold

Moreover, to evaluate the extracted behavior models, different cases are elaborated on below. Based on the user's semantic and time behavior ruleset, in the evening, the user usually

performs a business activity (fuels with gas). Therefore, once the user passes by any gas station in the evening, he would get a notification to fuel with gas (Figure 5-32).

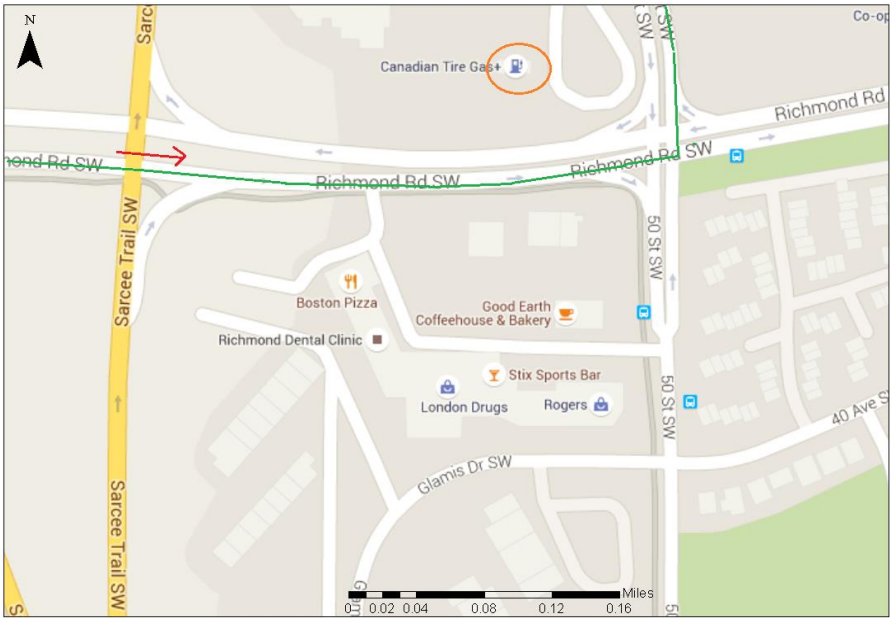


Figure 5-32 Testing the system using time and semantic behavior type

Based on the user's semantic behavior ruleset, since the user had some activities such as recreational, once he passes by a park, he would get a notification about the park (Figure 5-33).

Figure 5-33 Testing the system using semantic behavior type

In this case, the user passed by two different places such as a park and a mall. Based on his semantic and space behavior ruleset, he would get two notifications, one for each location (Figure 5-34).

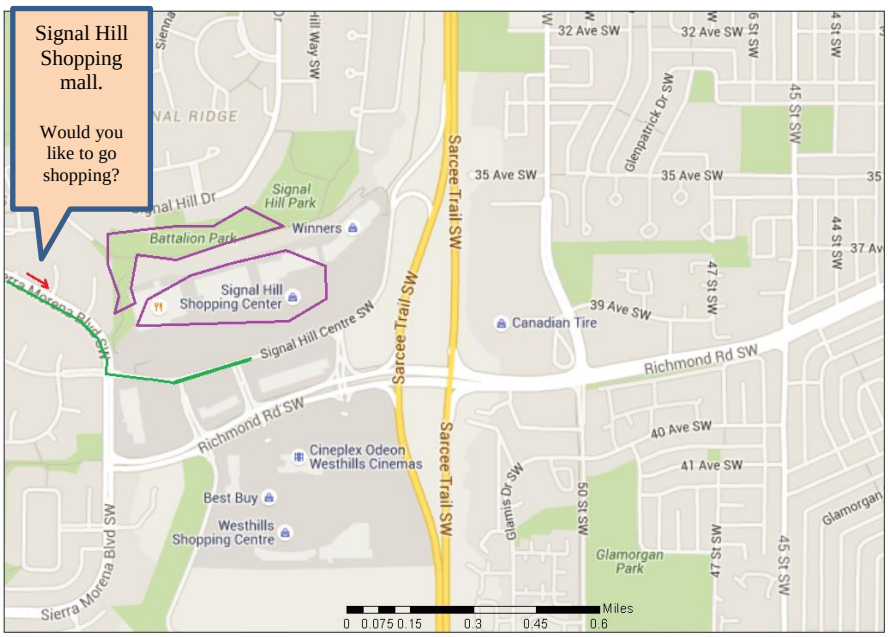


Figure 5-34 Testing the system using space and semantic behavior type

Based on the user's semantic and space-time behavior ruleset, if the user performs a recreational activity at location number 5 in the evening, then he would go shopping afterward. Therefore, based on this ruleset, the user has received a notification about a shopping activity while leaving the park and passing by a mall (Figure 5-35).

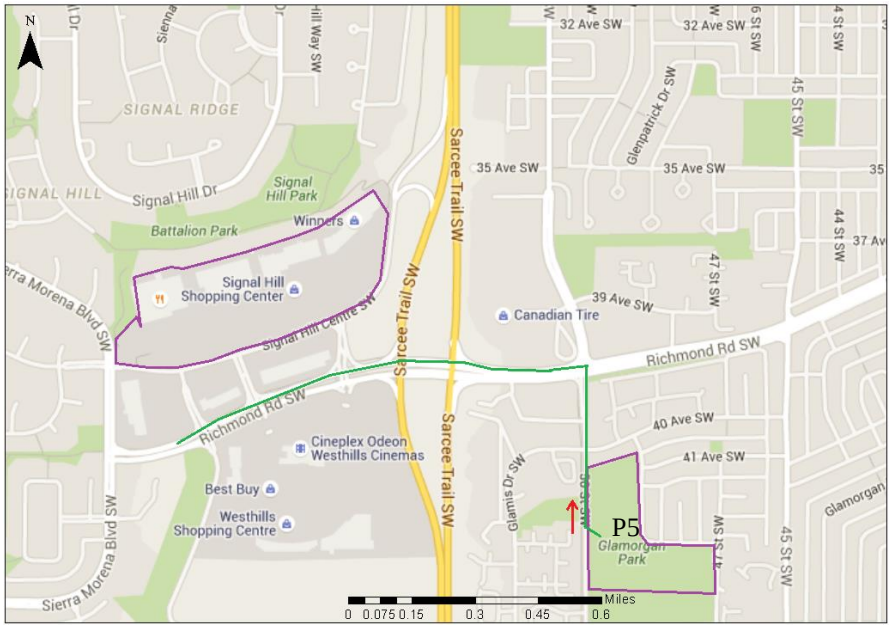


Figure 5-35 Testing the system using semantic and space-time behavior type

Moreover, location based advertisement was used to test the extracted behavior models. Therefore, several runs of the user based on different scenarios were tested. The origin and destination points of the user was selected randomly from the user's stops, which he had performed different activities at. The Dijkstra algorithm was used to compute the shortest path between the points. As mentioned earlier, the service ontology itself consists of different service types, namely: temporal, spatial and, spatiotemporal (Figure 5-36). The temporal services are the ones with a limited time period. For instance, there's 30% discount on everything at Tommy this week. The spatial services are the ones that will be sent in specific distance of the user. For instance, if the user's distance to Tommy store is less than or equal to 200 meters, a notification will be sent to the user. The spatiotemporal services are the ones that not even the user has to be in specific distance of the store, but also it has to match with the time period. For instance, if the user's distance to Tommy store is less than or equal to 200 meters in this week, a notification will be sent to the user.

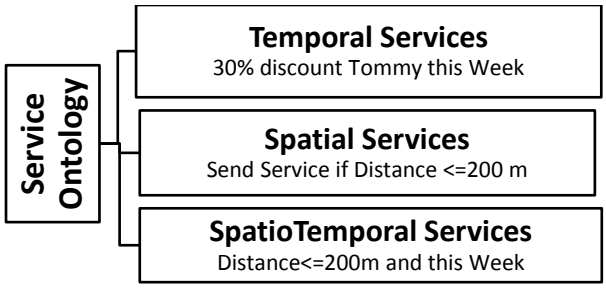


Figure 5-36 service ontology, which consists of different types

As can be seen in Table 5-31, different service layers were downloaded from open data catalogue to be used in this research to implement the prototype such as sport, amenity, community center and business. Each one had variety of services to provide. For instance, amenity layer had some services related to parks and recreation.

Table 5-31 Different service categories of the city of Calgary

Sport	Amenity	Community center	Business
Education School	Parks Recreation	Attraction Community center Court Hospital Library School	Retail dealer Contractor Cleaning services Food services Personal services

As it can be seen in Table 5-32, the above mentioned services are related to the extracted activities. For instance, for recreational activities, two different service categories named sport and amenity were considered.

Table 5-32 Different service categories related to different activity types

Activity type	Service category
Socializing	Sport/Amenity Services
Shopping	Business Services
Daily shopping	Business Services
Eating	Business Services
Education	Sport/Community Services
Recreational	Sport/Amenity Services

Business services	Business Services
Health services	Community Services

Four different behavior models: semantic, semantic and space, semantic and time, and semantic and space-time were considered and 800 regular services were defined in this testing. As it can be seen in Table 5-33, by considering different behavior models, the number of regular services has significantly reduced for each behavior model. For example, by considering semantic and temporal behavior model, there was 61.4 % service reduction. This indicates that considering the user's behavior model can customize the available services for the user by filtering out some irrelevant services.

Table 5-33 Comparing different behavior types with the number of delivered services

Behavior type	Percentage of service reduction
Semantic	48.5 %
Semantic and space	73.1 %
Semantic and time	61.4 %
Semantic and space-time	93.3 %

5.4 Summary

In this chapter the experimental analysis of the proposed framework using two different datasets was presented. First, the activity recognition step was executed to infer different activity types using the activity based ontology model. Second, the semantic behavior modelling was used to model different users' behavior types. In this step, association rule mining was used to extract rulesets. For the evaluation, the results of the activity recognition method were evaluated using the users' feedback, which was collected through the designed website. Moreover, for the semantic behavior modelling, location based advertisement was used to test the extracted behavior models. The results showed that applying the extracted rulesets could significantly reduce the number of available services and customize the services based on the rules. In the next chapter the results are concluded and some future works are recommended.

CHAPTER SIX: CONCLUSIONS

In this chapter, first the proposed ontology based semantic knowledge discovery framework in this research is summarized. The results of the main parts of the framework such as the activity type recognition and the semantic behavior modelling are discussed. After that, the contributions of the current research are presented. Finally, some suggestions for future research are provided.

6.1 Summary

In recent decades, the large availability of mobility data has raised many research interests and challenges for analyzing trajectories. In addition, smartphones with several embedded sensors expose new opportunities for identifying the lifestyle of an individual, including variety of activity types. Existing studies on such mobility data are mainly limited to analyzing the raw trajectories without enough consideration of associated higher level semantics. This is the case for many sophisticated techniques for trajectory data querying and mining that have been proposed in the trajectory database and data mining fields during the last decade. Recently, studies from the semantic web community and the activity recognition domain have started to build some meaningful concepts for representing mobility. However, the gap between low level sensing data and high level mobility concept is still unclear.

Therefore, this research proposed an ontology based semantic knowledge discovery framework for understanding the behavior of movement data and semantically interpreting trajectory patterns. It consisted of three different steps, namely: semantic trajectory ontology modelling, activity recognition, and semantic behavior modelling. First, a semantic conceptual data model was proposed, which was based on several semantic features such as POI type, land use type, stop begin time, stop frequency, and stop duration. In this model, some definitions regarding the important concepts such as stop, move, activity type, and behavior type were provided. Moreover, an ontology model was built based on the proposed semantic conceptual model, which covered four different dimensions including geometry, geography, theme, and service.

The activity recognition approach was used to infer users' activity types and consequently create semantic trajectories. It consisted of several steps. The first step was data preparation, where the raw data was cleaned and daily and weekly basis trajectories were identified. The second step was the semantic enrichment process, which included stop detection, finding probable visited

places, and extracting semantic features. Once stops were detected, they were annotated with the POI and land use types, which were downloaded from the OSM. After that, several semantic features such as stop begin time, stop frequency and average duration were extracted. The final step was ontology based activity modeling. The model consisted of different ontologies such as time, place, stop, and activity type. The retrieved information from the previous steps were used to populate the STOM for classifying different activity types by reasoning. The inferred activity types were evaluated by users through a web interface, which asked the users about the correctness of the inferred results. Since the previous research works in the activity recognition have considered only few features in order to extract different activity types, therefore, this research was unable to compare the proposed approach with another works.

Next, the semantic behavior modelling was used to model different users' behavior types. Different steps such as data preprocessing and association rule mining were used for this modelling. The model considered four different behavior types such as semantic, semantic and space, semantic and time, and semantic and space-time. Finally, a prototype was developed in order to test the performance of the framework. This was accomplished by customizing the location based advertisement based on the extracted behavior models.

The performance of the framework was evaluated using a simulated dataset and two different real datasets. The simulated dataset was used to investigate the effect of distance on the POI category type annotation. Different distance values, 50, 75, 100, and 125 were used in the testing. The best result was obtained by considering 100-meter distance. It was observed that the accuracy of the activity recognition method was related to the availability of the POIs around the stops. The more POIs were available around the stops, the better results were obtained for the acitivity types, which were related to the POI category type. In comparison with other methods, the proposed method for activity recognition considered several constraints to annotate stops with the POI category types and the land use types. One constraint was added to relate the duration of the stop to the typical duration of the visits. Therefore, for each POI a minimum service time was defined to express the minimum amount of time that a person needs to spend to visit the place. Another constraint was that the amount of time a person could spent in a place is not the complete stop duration, but the time needed to cover the distance between the POI and the stop must be taken into account. There are in fact some critical cases where the activities performed at a certain place may not be uniquely identified. It is needed to point out that a number of assumptions are at

the basis of the proposed approach. For example, there is a general assumption that during a stop a person performs one activity while sometimes more than one activity can be done.

The proposed semantic behavior modelling was applied to discover interesting knowledge about different behavior types of users as rulesets. In order to evaluate the extracted rules, the dataset was divided into two different datasets. Each dataset was used in order to generate rules by considering different thresholds for the minimum support (0 %, 2 %, 4 %, 6 %, 8 %, and 10 %) and the minimum confidence (5 %, 10 %, and 30 %). The results showed that the impact of the thresholds on the number of extracted rules were pretty much the same for both datasets. It can be concluded that the knowledge discovery process is application dependent and it is needed to integrate geographic information into the analysis of trajectories in order to extract clearer and more meaningful patterns. The results showed that applying the extracted rulesets could significantly reduce the number of available services and customize the services based on the rules. Therefore, these rulesets could be exploited to make intelligent predictions about user's future behavior given the time and location of the user.

6.2 Contributions

Towards the motivation and research challenges to establish an ontology based semantic knowledge discovery framework for computing and understanding mobility data, this thesis formulates three major contributions.

- The first contribution of this research was developing a semantic conceptual data model in order to reason different activity types and also to understand movement behavior. In this regard, several semantic features such as POI type, land use type, stop begin time, stop frequency, and stop duration were added to the model to describe different activity and behavior types of users. The conceptual data model helped in developing an ontology model that covered four different dimensions including geometry, geography, theme, and service to address the interactions between them. The purpose of this contribution was to address research questions Q1 and Q2.
- The second contribution was proposing an ontology based activity model to infer different activity types. The semantic conceptual data model was used to develop the model. This study investigated various extracted features and background information based on the

ontology model to extract the activity types. The proposed method considered several constraints to annotate land use and POI types to the stops. Moreover, different axioms were defined using common sense rules to reason activity types. The purposed of this contribution was to address research question Q3.

- The third contribution was semantic behavior modelling. Different semantic behavior types were defined and developed to extract rulesets. The association rule technique, *apriori* was used to extract semantic behavior models. The purposed of this contribution was to address research question Q4. Moreover, a prototype was developed to evaluate the proposed framework using one of the LBSs called location based advertisement. This was accomplished by customizing the service on the extracted behavior model. The purposed of this development was to address research question Q5.

6.3 Future Research

For better understanding mobility data, the research work described in this thesis could be continued in the future. Listed are several recommendations that should be investigated further:

- Extracting more activity types from multiple sensors

Computing semantic trajectories from multiple sensors embedded in smartphones, namely the study of combining GPS and accelerometer. Integrating GPS with accelerometer data can provide further semantics for better understanding mobility data. Besides GPS and accelerometer, extra sensors such Wi-Fi, sound, Bluetooth, and gyroscopes can be explored.

- Developing online trajectory computing algorithms for streaming movement data

Data in real time systems is usually coming from different sources in a decentralized way, and cannot be preprocessed and merged together in advance. The systems should be able to perform online trajectory computing in a distributed setting that is often encountered in some large scale application scenarios.

- Investigating the usefulness of geo-social network data in semantic trajectory computation

Recently, there are a lot of geo-social networks, which are supporting, collecting, and annotating trajectories such as Google Latitude, Foursquare, Facebook Place, and Gowalla. As geo-social

networks enable users to conveniently tag their movement and social activities with user defined tags. These tagging information can be used to annotate and generate semantic trajectories.

- Using socio-demographic information in human activity recognition

Users with different social demographic characteristics would perform different activity types. Therefore, it is helpful to investigate the effects of this kind of information in the activity type recognition approaches.

- Using fuzzy ontology in the proposed framework

Since human behavior is a complex study, using fuzzy ontology in the proposed framework can help to model and reason about vague and uncertain knowledge in different application domains. Therefore, the affect of spatial, temporal, and semantic uncertainty of the discovered patterns can be investigated.

- Addressing privacy and security issues

With more mobile application deployed and more data collected, the problem of privacy and security is bound to increase. Therefore, privacy-aware analysis methods have to be applied to avoid the disclosing of personal information such as the visit to special places. This thesis did not cover the issues of privacy-preserving semantic trajectory computing, but this definitely will be an important research topic in the future.

REFERENCES

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: ACM SIGMOD Record. pp. 207–216.

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules, in: Proc 20th Int Conf Very Large Data Bases VLDB. pp. 487–499.

Alvares, L.O., Bogorny, V., Kuijpers, B., de Macedo, J.A.F., Moelans, B., Vaisman, A., 2007. A model for enriching trajectories with semantic geographical information, in: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems. p. 22.

Alvares, L.O., Loy, A.M., Renso, C., Bogorny, V., 2011. An algorithm to identify avoidance behavior in moving object trajectories. Journal of the Brazilian Computer Society 17, 193–203.

Amine, A., Mohamed, O.A., Bellatreche, L., 2013. Modeling Approaches and Algorithms for Advanced Computer Applications. Springer.

An, A., Khan, S., Huang, X., 2003. Objective and subjective algorithms for grouping association rules, in: Proceedings of the Third IEEE International Conference on Data Mining. p. 477.

Andrienko, G., Andrienko, N., Wrobel, S., 2007. Visual analytics tools for analysis of movement data. ACM SIGKDD Explorations Newsletter 9, 38–46.

Andrienko, N., Andrienko, G., 2007. Designing visual analytics methods for massive collections of movement data. Cartographica: The International Journal for Geographic Information and Geovisualization 42, 117–138.

Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999. OPTICS: ordering points to identify the clustering structure, in: ACM Sigmod Record. ACM, pp. 49–60.

Appice, A., Berardi, M., Ceci, M., Malerba, D., 2005. Mining and filtering multi-level spatial association rules with ARES. Foundations of Intelligent Systems 13–13.

Ashbrook, D., Starner, T., 2003. Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing 7, 275–286.

Baglioni, M., Fernandes de Macêdo, J.A., Renso, C., Trasarti, R., Wachowicz, M., 2009a. Towards semantic interpretation of movement behavior, in: Advances in GIScience. pp. 271–288.

Baglioni, M., Fernandes de Macêdo, J.A., Renso, C., Trasarti, R., Wachowicz, M., 2009b. Towards semantic interpretation of movement behavior. Advances in GIScience 271–288.

Baglioni, M., Macedo, J., Renso, C., Wachowicz, M., 2008. An ontology-based approach for the semantic modelling and reasoning on trajectories, in: Advances in Conceptual Modeling–Challenges and Opportunities. Springer, pp. 344–353.

Baldauf, M., Fröhlich, P., Masuch, K., Grechenig, T., 2011. Comparing viewing and filtering techniques for mobile urban exploration. Journal of Location Based Services 5, 38–57.

Bashir, F.I., Khokhar, A.A., Schonfeld, D., 2007. Object trajectory-based activity classification and recognition using hidden Markov models. Image Processing, IEEE Transactions on 16, 1912–1919.

Bauer, M., Baldes, S., 2005. An ontology-based interface for machine learning, in: Proceedings of the 10th International Conference on Intelligent User Interfaces. pp. 314–316.

Bay, S.D., Pazzani, M.J., 2001. Detecting group differences: Mining contrast sets. Data Mining and Knowledge Discovery 5, 213–246.

Bellandi, A., Furletti, B., Grossi, V., Romei, A., 2007. Ontology-driven association rule extraction: A case study. Contexts and Ontologies Representation and Reasoning 10.

Bernstein, A., Hill, S., Provost, F., 2002. Intelligent assistance for the data mining process: An ontology-based approach. NYU Working Paper No. 2451/14146.

Bhattacharya, T., Kulik, L., Bailey, J., 2012. Extracting significant places from mobile user GPS trajectories: a bearing change based approach, in: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, pp. 398–401.

Binh, D.T.T., Duong, N.V., Cuong, H.M., 2011. Applying Gravity model to analyze trade activities of Vietnam.

Bittner, T., Winter, S., 1999. On Ontology in Image Analysis, in: In Integrated Spatial Databases: Digital Images and GIS, P. Agouris and. Springer-Verlag, pp. 168–191.

Blythe, P.W., Miller, G.F., Todd, P.M., 1996. Human simulation of adaptive behavior: Interactive studies of pursuit, evasion, courtship, fighting, and play, in: From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior. pp. 13–22.

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

Bogorny, V., Heuser, C.A., Alvares, L.O., 2010. A conceptual data model for trajectory data mining, in: Geographic Information Science. Springer, pp. 1–15.

Bogorny, V., Kuijpers, B., Alvares, L.O., 2009. ST-DMQL: A Semantic Trajectory Data Mining Query Language. International Journal of Geographical Information Science 23, 1245–1276.

Bogorny, V., Renso, C., Aquino, A.R., Lucca Siqueira, F., Alvares, L.O., 2014. Constant–a conceptual data model for semantic trajectories of moving objects. Transactions in GIS 18, 66–88.

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. Transportation Research Part C: Emerging Technologies 17, 285–297.

Boulicaut, J.F., Masson, C., 2005. Data mining query languages. Data Mining and Knowledge Discovery Handbook 715–726.

Boulmakoul, A., Karim, L., Lbath, A., 2012. Moving object trajectories meta-model and spatio-temporal queries. arXiv preprint arXiv:1205.1796.

Brisson, L., 2007. Knowledge extraction using a conceptual information system (ExCIS). Ontologies-Based Databases and Information Systems 119–134.

Buchin, M., Driemel, A., van Kreveld, M., Sacristán, V., 2010. An algorithmic framework for segmenting trajectories based on spatio-temporal criteria, in: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 202–211.

Cannataro, M., Comito, C., 2003. A data mining ontology for grid programming. Proceedings of (SemPGrid2003) 113–134.

Cao, H., Mamoulis, N., Cheung, D.W., 2007. Discovery of periodic patterns in spatiotemporal sequences. Knowledge and Data Engineering, IEEE Transactions on 19, 453–467.

Cao, X., Cong, G., Jensen, C.S., 2010. Mining significant semantic locations from GPS data. Proceedings of the VLDB Endowment 3, 1009–1020.

Charest, M., Delisle, S., 2006. Ontology-guided intelligent data mining assistance: Combining declarative and procedural knowledge, in: Artificial Intelligence and Soft Computing. pp. 9–14.

Chaves, M., Silva, M., Martins, B., 2005. A geographic knowledge base for semantic web applications.
Chen, H., Rakha, H.A., McGhee, C.C., 2013. Dynamic Travel Time Prediction using Pattern Recognition, in: 20th World Congress on Intelligent Transportation Systems, Tokyo, Japan.
Chen, K., Liu, L., 2011. Geometric data perturbation for privacy preserving outsourced data mining. Knowledge and Information Systems 29, 657–695.
de Lucca Siqueira, F., Bogorny, V., 2011. Discovering chasing behavior in moving object trajectories. Transactions in GIS 15, 667–688.
Dellis, E., Paliouras, G., 2007. Management of large spatial ontology bases, in: Ontologies-Based Databases and Information Systems. Springer, pp. 102–118.
Dodge, S., 2011. Exploring Movement Using Similarity Analysis (PhD thesis). Universitaet Zuerich.
Dodge, S., Weibel, R., Lautenschütz, A.K., 2008. Towards a taxonomy of movement patterns. Information Visualization 7, 240–252.
Elsner, J.B., Kara, A.B., 1999. Hurricanes of the North Atlantic: Climate and society. Oxford University Press, USA.
EU Coordination Action, FET OPEN, 2009. Mobility, Data Mining, and Privacy (MODAP).
EU Marie Curie Project N 295179, 2009. SEmantic Enrichment of trajectory Knowledge discovery (SEEK).
European Science Foundation, 2009. Knowledge Discovery from Moving Objects (MOVE).
Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI magazine 17, 37–54.
Fonseca, F.T., Egenhofer, M.J., Agouris, P., Câmara, G., 2002. Using ontologies for integrated geographic information systems. Transactions in GIS 6, 231–257.
Fox, J., Daly, A., Hess, S., Miller, E., 2014. Temporal transferability of models of mode-destination choice for the Greater Toronto and Hamilton Area. Journal of Transport and Land Use 7, 41–62.
Fraile, R., Maybank, S.J., 1998. Vehicle Trajectory Approximation and Classification., in: BMVC. pp. 832–840.
Frank, A., 1997. Spatial ontology: A geographical information point of view. Spatial and Temporal Reasoning 135–153.
Fu, Z., Hu, W., Tan, T., 2005. Similarity based vehicle trajectory clustering and anomaly detection, in: Image Processing, 2005. ICIP 2005. IEEE International Conference on. IEEE, p. II–602.
Furletti, B., Cintia, P., Renso, C., Spinsanti, L., 2013. Inferring human activities from GPS tracks, in: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. ACM, p. 5.
Galton, A., 2005. Dynamic collectives and their collective dynamics. Spatial Information Theory 300–315.
Giannotti, F., 2011. Mobility, Data Mining and Privacy Understanding Human Movement Patterns from Trajectory Data. IEEE, pp. 4–5. doi:10.1109/MDM.2011.103
Giannotti, F., Giannotti, G., Pedreschi, D., 2008. Mobility, data mining, and privacy: geographic knowledge discovery. Springer.
Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., Trasarti, R., 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory

data. The VLDB Journal—The International Journal on Very Large Data Bases 20, 695–719.
Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D., 2007. Trajectory pattern mining, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 330–339.
Giannotti, F., Pedreschi, D., 2008. Mobility, data mining and privacy: Geographic knowledge discovery. Springer Science & Business Media.
Gidófalvi, G., Pedersen, T.B., 2009. Mining long, sharable patterns in trajectories of moving objects. Geoinformatica 13, 27–55.
Gómez, L.I., Kuijpers, B., Vaisman, A.A., 2008. Aggregation languages for moving object and places of interest, in: Proceedings of the 2008 ACM Symposium on Applied Computing. ACM, pp. 857–862.
Gonzalez, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782.
Grenon, P., Smith, B., 2004. SNAP and SPAN: Towards dynamic spatial ontology. Spatial cognition and computation 4, 69–104.
Griffin, T., Huang, Y., 2008. A Decision Tree Classification Model to Automate Trip Purpose Derivation, in: Proceedings of the 18th International Conference on Computer Applications in Industry and Engineering. USA.
Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing? International journal of human-computer studies 43, 907–928.
Gruber, T.R., 1992. Ontolingua: A mechanism to support portable ontologies. Stanford University, Knowledge Systems Laboratory.
Gruber, T.R., others, 1995. Toward principles for the design of ontologies used for knowledge sharing. International journal of human computer studies 43, 907–928.
Guarino, N., 1998. Formal ontology in information systems: proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. Ios Pr Inc.
Guarino, N., 1997. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. Information Extraction A Multidisciplinary Approach to an Emerging Information Technology 139–170.
Guarino, N., Oberle, D., Staab, S., 2009. What is an Ontology?, in: Handbook on Ontologies. Springer, pp. 1–17.
Guc, B., May, M., Saygin, Y., Körner, C., 2008. Semantic annotation of gps trajectories, in: 11th AGILE International Conference on Geographic Information Science.
Guting, R.H., Almeida, V., Ansorge, D., Behr, T., Ding, Z., Hose, T., Hoffmann, F., Spiekermann, M., Telle, U., 2005. Secondo: An extensible DBMS platform for research prototyping and teaching, in: Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on. pp. 1115–1116.
Güting, R.H., Böhlen, M.H., Erwig, M., Jensen, C.S., Lorentzos, N.A., Schneider, M., Vazirgiannis, M., 2000. A foundation for representing and querying moving objects. ACM Transactions on Database Systems (TODS) 25, 1–42.
Güting, R.H., Schneider, M., 2005. Moving objects databases. Elsevier.
Güting, R.H., Schneider, M., 2005. Moving objects databases. Morgan Kaufmann Pub.
Han, B., Liu, L., Omiecinski, E., 2012. Neat: Road network aware trajectory clustering, in: Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on. IEEE, pp. 142–151.

Han, J., Dong, G., Yin, Y., 1999. Efficient mining of partial periodic patterns in time series database, in: Data Engineering, 1999. Proceedings., 15th International Conference on. IEEE, pp. 106–115.

Hipp, J., Mangold, C., Güntzer, U., Nakhaeizadeh, G., 2002. Efficient rule retrieval and postponed restrict operations for association rule mining. Advances in Knowledge Discovery and Data Mining 52–65.

Hobbs, J.R., Pan, F., 2004. An ontology of time for the semantic web. ACM Transactions on Asian Language Information Processing (TALIP) 3, 66–85.

Huang, L., Li, Q., 2010. Automatic activity identification from raw GPS vehicle tracking data, 2010, in: Proceedings of the Canadian Geomatics Conference.

Huang, L., Li, Q., Yue, Y., 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. ACM, pp. 27–30.

Hunter, A., 2007. Sensor-based animal tracking.

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

Jasper, R., Uschold, M., 1999. A framework for understanding and classifying ontology applications, in: Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling, and Management KAW. pp. 16–21.

Jensen, C.S., 2002. Research Challenges in Location-enabled M-services, in: Mobile Data Management, 2002. Proceedings. Third International Conference on. pp. 3–7.

Jeung, H., Shen, H.T., Zhou, X., 2007. Mining trajectory patterns using hidden Markov models, in: Data Warehousing and Knowledge Discovery. Springer, pp. 470–480.

Jeung, H., Yiu, M.L., Jensen, C.S., 2011. Trajectory pattern mining, in: Computing with Spatial Trajectories. Springer, pp. 143–177.

Jeung, H., Yiu, M.L., Zhou, X., Jensen, C.S., Shen, H.T., 2008. Discovery of convoys in trajectory databases. Proceedings of the VLDB Endowment 1, 1068–1080.

Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., Mascolo, C., 2013. Geo-spotting: mining online location-based services for optimal retail store placement, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 793–801.

Kincses, Á., Tóth, G., 2014. The Application of Gravity Model in the Investigation of Spatial Structure. Acta Polytechnica Hungarica 11, 5–19.

Kotiyal, B., Kumar, A., Pant, B., Goudar, R.H., Chauhan, S., Junee, S., 2013. User behavior analysis in web log through comparative study of Eclat and Apriori, in: Intelligent Systems and Control (ISCO), 2013 7th International Conference on. IEEE, pp. 421–426.

Kraak, M.J., 2003. The space-time cube revisited from a geovisualization perspective, in: Proc. 21st International Cartographic Conference. pp. 1988–1996.

Krumm, J., Horvitz, E., 2006. Predestination: Inferring destinations from partial trajectories, in: UbiComp 2006: Ubiquitous Computing. Springer, pp. 243–260.

Kuijpers, B., Moelans, B., Othman, W., Vaisman, A., 2009. Analyzing trajectories using uncertainty and background information, in: Advances in Spatial and Temporal Databases. Springer, pp. 135–152.

Laube, P., 2009. Progress in Movement Pattern Analysis., in: BMI Book. pp. 43–71.

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

Laube, P., Dennis, T., Forer, P., Walker, M., 2007. Movement beyond the snapshot-dynamic analysis of geospatial lifelines. Computers, Environment and Urban Systems 31, 481–501.

Laube, P., Imfeld, S., 2002. Analyzing relative motion within groups oftrackable moving point objects, in: Geographic Information Science. Springer, pp. 132–144.

Laube, P., Imfeld, S., Weibel, R., 2005. Discovering relative motion patterns in groups of moving point objects. International Journal of Geographical Information Science 19, 639–668.

Lee, J.G., Han, J., Li, X., Gonzalez, H., 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proceedings of the VLDB Endowment 1, 1081–1094.

Lee, J.-G., Han, J., Whang, K.-Y., 2007. Trajectory clustering: a partition-and-group framework, in: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. ACM, pp. 593–604.

Li, Z., Ding, B., Han, J., Kays, R., Nye, P., 2010. Mining periodic behaviors for moving objects, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1099–1108.

Liao, L., Fox, D., Kautz, H., 2007. Extracting places and activities from gps traces using hierarchical conditional random fields. The International Journal of Robotics Research 26, 119–134.

Liao, L., Fox, D., Kautz, H., 2006a. Location-based activity recognition. Advances in Neural Information Processing Systems 18, 787.

Liao, L., Patterson, D.J., Fox, D., Kautz, H., 2006b. Building personal maps from GPS data. Annals of the New York Academy of Sciences 1093, 249–265.

Lloyd, S.P., 1982. Least squares quantization in PCM. Information Theory, IEEE Transactions on 28, 129–137.

Marketos, G., Frentzos, E., Ntoutsi, I., Pelekis, N., Raffaetà, A., Theodoridis, Y., 2008. Building real-world trajectory warehouses, in: Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access. ACM, pp. 8–15.

Mehra, P., 2012. Context-aware computing: beyond search and location-based services. IEEE Internet Computing 16, 12–16.

Miller, E.J., Vaughan, J., King, D., Austin, M., 2015. Implementation of a “Next Generation” Activity-Based Travel Demand Model: The Toronto Case, in: TAC 2015: Getting You There Safely-2015 Conference and Exhibition of the Transportation Association of Canada//ATC: Destination sÃ\copyright curitÃ\copyright routiÃ" re-2015 CongrÃ" s et Exposition de l’Association Des Transports Du Canada.

Miller, H.J., 2005. A measurement theory for time geography. Geographical analysis 37, 17–45.

Miller, H.J., Han, J., 2009. Geographic data mining and knowledge discovery. CRC.

Monreale, A., Trasarti, R., Pedreschi, D., Renso, C., Bogorny, V., 2011. C-safety: a framework for the anonymization of semantic trajectories. Transactions on Data Privacy 4, 73–101.

Montini, L., Rieser-Schüssler, N., Horni, A., Axhausen, K., 2014. Trip purpose identification from GPS tracks. Transportation Research Record: Journal of the Transportation Research Board 16–23.

Moreno, B., Júnior, A.S., Times, V., Tedesco, P., Matwin, S., 2014. Weka-SAT: A Hierarchical Context-Based Inference Engine to Enrich Trajectories with Semantics, in: Canadian Conference on Artificial Intelligence. Springer, pp. 333–338.

Moreno, B., Times, V.C., Renso, C., Bogorny, V., 2010. Looking Inside the Stops of Trajectories of Moving Objects., in: Geoinfo. pp. 9–20.

Morzy, T., Zakrzewicz, M., 1998. Group bitmap index: a structure for association rules retrieval, in: Proc. ACM SIGKDD. pp. 284–288.

Mountain, D., 2005. Exploring mobile trajectories: An investigation of individual spatial behaviour and geographic filters for information retrieval (PhD thesis). City University, London Department of Information Science School of Informatics.

Mousavi, A., Hunter, A., 2012a. Pointers Extraction of Trajectory Data for Semantic Knowledge Discovery.

Mousavi, A., Hunter, A., 2012b. Using Ontology Based Knowledge Discovery in Location Based Services.

Mouza, C., Rigaux, P., 2005. Mobility patterns. GeoInformatica 9, 297–319.

Nanni, M., Kuijpers, B., Körner, C., May, M., Pedreschi, D., 2008. Spatiotemporal data mining. Mobility, Data Mining and Privacy 267–296.

Nascimento, J.C., Figueiredo, M.A., Marques, J.S., 2010. Trajectory classification using switched dynamical hidden Markov models. Image Processing, IEEE Transactions on 19, 1338–1348.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P.E., 2008. A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences 105, 19052–19059.

Neches, R., Fikes, R.E., Finin, T., Gruber, T., Patil, R., Senator, T., Swartout, W.R., 1991. Enabling technology for knowledge sharing. AI magazine 12, 36.

Nigro, H.O., Císaro, S.E.G., Xodo, D.H., Global, I.G.I., 2008. Data mining with ontologies: Implementations, findings, and frameworks. Information Science Reference.

Ong, R., Wachowicz, M., Nanni, M., Renso, C., 2010. From Pattern Discovery to Pattern Interpretation in Movement Data, in: Proceedings of the 2010 IEEE International Conference on Data Mining Workshops. pp. 527–534.

Palma, A.T., Bogorny, V., Kuijpers, B., Alvares, L.O., 2008. A clustering-based approach for discovering interesting places in trajectories, in: Proceedings of the 2008 ACM Symposium on Applied Computing. pp. 863–868.

Panagiotakis, C., Pelekis, N., Kopanakis, I., Ramasso, E., Theodoridis, Y., 2012. Segmentation and sampling of moving object trajectories based on representativeness. Knowledge and Data Engineering, IEEE Transactions on 24, 1328–1343.

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., undefined, others, 2013. Semantic trajectories modeling and analysis. ACM Computing Surveys (CSUR) 45, 42.

Parent, C., Spaccapietra, S., Zimányi, E., 2006. Conceptual modeling for traditional and spatio-temporal applications: The mads approach. Springer Science & Business Media.

Patterson, D.J., Liao, L., Fox, D., Kautz, H., 2003. Inferring high-level behavior from low-level sensors, in: UbiComp 2003: Ubiquitous Computing. Springer, pp. 73–89.

Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M.-C., 2001. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth, in: Icccn. IEEE, p. 215.

Pelekis, N., Andrienko, G., Andrienko, N., Kopanakis, I., Marketos, G., Theodoridis, Y., 2012. Visually exploring movement data via similarity-based analysis. Journal of Intelligent Information Systems 38, 343–391.

Pelekis, N., Frentzos, E., Giatrakos, N., Theodoridis, Y., 2008. HERMES: aggregative LBS via a trajectory DB engine, in: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. ACM, pp. 1255–1258.

Pelekis, N., Kopanakis, I., Kotsifakos, E.E., Frentzos, E., Theodoridis, Y., 2011. Clustering uncertain trajectories. Knowledge and Information Systems 28, 117–147.

Pelekis, N., Theodoridis, Y., Vosinakis, S., Panayiotopoulos, T., 2006. Hermes–a framework for location-based data management. Advances in Database Technology-EDBT 2006 1130–1134.

Pfoser, D., Theodoridis, Y., 2003. Generating semantics-based trajectories of moving objects. Computers, Environment and Urban Systems 27, 243–263.

Phillips, J., Buchanan, B.G., 2001. Ontology-guided knowledge discovery in databases, in: Proceedings of the 1st International Conference on Knowledge Capture. pp. 123–130.

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

Renso, C., Baglioni, M., de Macedo, J.A.F., Trasarti, R., Wachowicz, M., 2013. How you move reveals who you are: understanding human behavior by analyzing trajectory data. Knowledge and information systems 37, 331–362.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., Andrienko, G., 2008. Visually driven analysis of movement data by progressive clustering. Information Visualization 7, 225–239.

Rocha, J.A.M., Oliveira, G., Alvares, L.O., Bogorny, V., Times, V.C., 2010. DB-SMoT: a direction-based spatio-temporal clustering method, in: Intelligent Systems (IS), 2010 5th IEEE International Conference. IEEE, pp. 114–119.

Sbalzariniy, I., Theriot, J., Koumoutsakos, P., 2002. Machine learning for biological trajectory classification applications. Proc. 2002 Summer Program, Center for Turbulence Research 305–316.

Shen, B., Yao, M., Wu, Z., Zhang, Y., Yi, W., 2006. Ontology-based Association Rules Retrieval using Protege Tools, in: Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on. pp. 765–769.

Shiode, N., Li, C., Batty, M., Longley, P.A., Maguire, D., 2002. The impact and penetration of location-based services.

Sinha, G., Mark, D.M., 2005. Measuring similarity between geospatial lifelines in studies of environmental health. Journal of Geographical Systems 7, 115–136.

Smith, B., Mark, D.M., 1998. Ontology and geographic kinds, in: Poiker, T.K., Chrisman, N. (Eds.), 8th International Symposium on Spatial Data Handling (SDH'98). Presented at the 8th International Symposium on Spatial Data Handling (SDH'98), International Geographic Union, Vancouver, Canada, pp. 308–320.

SOOD, S., 2012. Behavior Computing: Modeling, Analysis, Mining and Decision.

Spaccapietra, S., Cullot, N., Parent, C., Vangenot, C., 2004. On spatial ontologies.

Spaccapietra, S., Parent, C., 2011. Adding meaning to your steps (keynote paper), in: Conceptual Modeling–ER 2011. Springer, pp. 13–31.

Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F., Vangenot, C., 2008. A conceptual view on trajectories. Data & Knowledge Engineering 65, 126–146. doi:10.1016/j.datak.2007.10.008

Spinsanti, L., Celli, F., Renso, C., 2010. Where you stop is who you are: understanding peoples' activities, in: Proceedings of the 5th Workshop on Behaviour Monitoring and Interpretation—user Modelling. Karlsruhe, Germany.

Srikant, R., Agrawal, R., 1996. Mining sequential patterns: Generalizations and performance improvements. Springer.

Steiniger, S., Neun, M., Edwardes, A., 2006. Foundations of location based services. CartouCHe1-Lecture Notes on LBS 1.

Studer, R., Benjamins, V.R., Fensel, D., 1998. Knowledge engineering: principles and methods. Data & knowledge engineering 25, 161–197.

Thériault, M., Claramunt, C., Villeneuve, P.Y., 1999. A spatio-temporal taxonomy for the representation of spatial set behaviours, in: Spatio-Temporal Database Management. Springer, pp. 1–18.

Toivonen, H., Klemettinen, M., Ronkainen, P., Hätönen, K., Mannila, H., 1995. Pruning and grouping discovered association rules.

Trasarti, R., Pinelli, F., Nanni, M., Giannotti, F., 2011. Mining mobility user profiles for car pooling, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1190–1198.

Trasarti, R., Rinzivillo, S., Pinelli, F., Nanni, M., Monreale, A., Renso, C., Pedreschi, D., Giannotti, F., 2010. Exploring real mobility data with M-atlas. Machine Learning and Knowledge Discovery in Databases 624–627.

Tsoukatos, I.I., Gunopulos, D., 2001. Efficient mining of spatiotemporal patterns. Advances in Spatial and Temporal Databases 425–442.

Tuzhilin, A., Adomavicius, G., 2002. Handling very large numbers of association rules in the analysis of microarray data, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 396–404.

Tuzhilin, A., Liu, B., 2002. Querying multiple sets of discovered rules, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 52–60.

Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F., 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. Web Semantics: science, services and agents on the World Wide Web 4, 14–28.

Valente, A., Breuker, J., 1996. Towards principled core ontologies, in: Proceedings of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems. Citeseer.

Viktoratos, I., Tsadiras, A., Bassiliades, N., 2014. Providing a context-aware location based web service through semantics and user-defined rules, in: Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14). ACM, p. 9.

Wachowicz, M., Ong, R., Renso, C., Nanni, M., 2011. Finding moving flock patterns among pedestrians through collective coherence. International Journal of Geographical Information Science 25, 1849–1864.

Wang, W., Yang, J., Muntz, R., others, 1997. STING: A statistical information grid approach to spatial data mining, in: VLDB. pp. 186–195.

Wang, X., Gu, W., Ziebelin, D., Hamilton, H., 2010. An ontology-based framework for geospatial clustering. International Journal of Geographical Information Science 24, 1601–1630.

Wannous, R., Malki, J., Bouju, A., Vincent, C., 2013a. Time integration in semantic trajectories using an ontological modelling approach, in: New Trends in Databases and Information Systems. Springer, pp. 187–198.

Wannous, R., Malki, J., Bouju, A., Vincent, C., 2013b. Modelling mobile object activities based on trajectory ontology rules considering spatial relationship rules, in: Modeling Approaches and Algorithms for Advanced Computer Applications. Springer, pp. 249–258.

Xie, K., Deng, K., Zhou, X., 2009. From trajectories to activities: a spatio-temporal join approach, in: Proceedings of the 2009 International Workshop on Location Based Social Networks. ACM, pp. 25–32.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K., 2013. Semantic trajectories: Mobility data computation and annotation. ACM Transactions on Intelligent Systems and Technology (TIST) 4, 49.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., Aberer, K., 2011. SeMiTri: a framework for semantic annotation of heterogeneous trajectories, in: Proceedings of the 14th International Conference on Extending Database Technology. ACM, pp. 259–270.

Yan, Z., Macedo, J., Parent, C., Spaccapietra, S., 2008. Trajectory ontologies and queries. Transactions in GIS 12, 75–91.

Yan, Z., Parent, C., Spaccapietra, S., Chakraborty, D., 2010. A hybrid model and computing platform for spatio-semantic trajectories, in: The Semantic Web: Research and Applications. Springer, pp. 60–75.

Yan, Z., Spaccapietra, S., 2009. Towards Semantic Trajectory Data Analysis: A Conceptual and Computational Approach., in: VLDB PhD Workshop. Citeseer.

Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 186–194.

Zaki, M.J., 2001. SPADE: An efficient algorithm for mining frequent sequences. Machine learning 42, 31–60.

Zhang, J., Goodchild, M.F., 2002. Uncertainty in geographical information. CRC press.

Zhang, T., Ramakrishnan, R., Livny, M., 1996. BIRCH: AnEcient DataClusteringMethod forVery Large Databases, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, Montreal, Canada.

Zheng, Y., Chen, Y., Li, Q., Xie, X., Ma, W.-Y., 2010. Understanding transportation modes based on GPS data for web applications. ACM Transactions on the Web (TWEB) 4, 1.

Zheng, Y., Zhang, L., Ma, Z., Xie, X., Ma, W.-Y., 2011. Recommending friends and locations based on individual location history. ACM Transactions on the Web (TWEB) 5, 5.

Zheni, D., Frihida, A., Ghezala, H.B., Claramunt, C., 2009. A semantic approach for the modeling of trajectories in space and time, in: Advances in Conceptual Modeling-Challenging Perspectives. Springer, pp. 347–356.

Zhixian, Y.A.N., 2011. Semantic Trajectories: Computing and Understanding Mobility Data. ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE.

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L., 2004. Discovering personal gazetteers: an interactive clustering approach, in: Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems. ACM, pp. 266–273.

n.d. . Time Ontology. URL http://www.w3.org/2006/time (accessed 10.20.14).