Sixth Framework Programme

Information Society Technologies

Future Emerging Technologies

Information Society Technologies

Project Acronym:

GeoPKDD

Project full title:

Geographic Privacy-aware Knowledge Discovery and Delivery

Project Number:FP6-014915
Instrument: Specific Target Research Project

Project Deliverable D4.1

Technical Report on the Characterization of the Class of Applications of the GEOPKDD Paradigm

Date of preparation: 10/12/2006
Revision: final

Operative commencement date of Contract: 01/12/2005
Project Coordinator: Fosca Giannotti
Project institute coordinator: Knowledge Discovery and Delivery-LAB / ISTI-CNR

Programme Name: ....IST
Project Number:............................014915
Project Title: ....GEOPKDD

Document Number: D4.1
Work-Package:..............................WP4
Document Type: ....Deliverable
Contractual Date of Delivery: 01/12/2006
Actual Date of Delivery: 10/12/2006
Title of Document: ....Technical Report on the Characterization of the Class of Applications of the GEOPKDD Paradigm

Author(s): ....Monica Wachowicz, Arend Ligtenberg (WUR - Alterra), Salvo Rinzivillo (KDDLAB)

Dissemination Level.....................Public

GeoPKDD consortium partecipants

Partic. Role	Partic. N.	Participant name	Short name	Country
CO	**1**	**KDD Lab.** joint research group of **ISTI-CNR**, Istituto di Scienza e Tecnologie dell'Informazione, Pisa. http://www-kdd.isti.cnr.it/ and **Univ. Pisa**, Dept. of Computer Science http://www.di.unipi.it	**KDDLAB**	I
CR	**2**	**Univ. of Hasselt**, Theoretical Computer Science Group. http://alpha.uhasselt.be/research/groups/theocomp/	**HASSELT**	B
CR	**3**	**EPFL - Ecole Polytechnique Fédérale de Lausanne**, Lab. DB, Lausanne. http://lbdwww.epfl.ch/e/ **and** University of Milan - Computer Science and Communication Department http://www.dico.unimi.it/	**EPFL**	CH
CR	**4**	**Fraunhofer Institute for Autonomous Intelligent Systems**, Sankt Augustin. http://www.ais.fraunhofer.de/	**FAIS**	D
CR	**5**	**Wageningen UR,** Centre for GeoInformation. http://cgi.girs.wageningen-ur.nl/	**WUR**	NL
CR	**6**	**Research Academic Computer Technology Institute,** Research and Development Division. http://www.cti.gr/ and Univ. Piraeus, Dept. of Informatics http://www.unipi.gr	**CTI**	GR
CR	**7**	**Sabanci University**, Faculty of Engineering and Natural Sciences. http://www.sabanciuniv.edu/	**UNISAB**	TK
CR	**8**	**WIND Telecomunicazioni SpA,** Direzione Reti Wind Progetti Finanziati & Technology Scouting. http://www.wind.it	**WIND**	I

CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

1.1. This report

This report addresses the deliverable D4.1 Characterization of the class of applications of the GeoPKDD paradigm of the Work Package 4 "Application Demonstrators". Chapter 2 outlines the role of the GeoPKDD process in supporting sustainable mobility. The main outcomes from the Workshop on Sustainable Mobility are also presented. Chapter 3 describes the three promising applications domains for the demonstration of the GeoPKDD process. Chapter 4 shows five potential application demonstrators, from which two have been selected as potential candidates to be used during the Phase 2 – Application Demonstrations of the GeoPKDD project.

1.2. Rational of the WP 4

This WP aims to develop a preliminary design of the GeoPKDD platform, comprising the collection and management of a representative sample of the database (e.g. aggregated spatio-temporal locations of mobiles), finalized to the realization of the selected application demonstrators. Towards this goal, the task ***T4.1 – Characterization of class of applications of the GeoPKDD paradigm*** has been realized by the responsible contractor: WUR-Alterra; and the participants: KDDLAB and WIND, with involvement of EPFL, Hasselt University and FAIS. This task aimed to provide a set of requirements and the recognition of the potential applications supported by aggregative location-based services, societal and economical applications enabled by aggregative geographic knowledge extractable from the GeoPKDD process.

2. THE ROLE OF THE GEOPKDD PARADIGM IN SUPPORTING SUSTAINABLE MOBILITY

The proliferation of mobile technologies for "always-on" at "any-time" and "any-place" has facilitated the generation of huge volume of positioning data sets containing information about the location and the movement of entities through the geographic environment. In principle, every time an entity moves through space, it creates a trajectory (i.e. track or path) representing the history of its past and current locations. Examples of interesting trajectories of moving entities may range from hurricane and tornado tracks (Georges *et al.* 1993) up to individual trajectories of animals (Hunter *et al.* 2005) and planes (Brunk and Davis 2002). Specially designed sensors can provide the location of a mobile entity as well as information about the geographic environment where this entity is moving. Current research on mobile technologies such as sensor web, wireless communication, and portable computers has been crucial for the development of multi-sensor systems. Their use to sense a geographic environment and mobile entities can include photodiodes to detect light level, accelerometers to provide tilt and vibration measurements, passive infrared sensors to detect the proximity of humans, omni-directional microphones to detect sound, and other built-in sensors for temperature, pressure, and CO gas (Chen and Kotz 2000).

Moreover, there are many types of mobile applications which have been developed in meeting many of society's needs for economic development, experience and culture. Some examples include the applications based on Location Based Services (LBS) such as tourism, marketing, and transportation management. Existing LBS can already provide tourists the information about their current location in a way that they can find directions, retrieve geographic information, and leave comments on an interactive map (Abowd *et al.* 1997, Oppermann and Specht 2000). A travel diary can be automatically compiled using the history where a tourist has traveled over time. Some systems are also capable of making suggestions on places of interest to visit by visualizing the required information within an augmented reality environment. The positioning data is usually collected by the Global Positioning System (GPS) for outdoor tracking and infrared (IR) positioning for indoor tracking.

Although positioning data sets containing information about the location of mobile entities may be available to develop mobile applications, how to effectively use that information is still a challenging problem. New methods and tools are needed in the fields of databases, statistics, geography, remote sensing, and artificial intelligence that can automatically transform these very large positioning data sets into information about the movement of entities, and furthermore, be the source of geographic knowledge. One of the main challenges still remains on to extracting new, insightful information embedded within the large heterogeneous databases that contain private information about the location of the mobile entities and their surrounding geographic environment.

We need to go beyond the collection of positioning data sets to the delivery of information and knowledge derived from these data. A knowledge discovery process empowers the experts of an application domain to extract relevant and useful geographic knowledge from very large positioning data sets. It also supports the development of the next generation of mobile applications through its ability to cope with data warehousing, target data selection, cleaning, and pre-processing, as well as data mining, model selection, evaluation and interpretation of the hidden patterns embedded within very large heterogeneous databases. However, one of the main research issues is privacy, which is concerned with the safeguard of processing positioning data sets. It is important to realize that privacy concerns are very important with respect to the social acceptance of the use of a geographic knowledge discovery process for developing mobile applications (Smyth 2001). The latest developments in privacy-preserving techniques in databases are the primordial importance if the aim is to reconstruct a vast number of individual trajectories which will allow an efficient and effective storage of these trajectories as well as suitable access methods to support analysis and data mining tasks. Such privacy-preserving techniques should support mechanisms that prevent the disclosure of sensitive data, both explicitly (e.g., providing individual's identity) and implicitly (providing non-sensitive data from which sensitive information can be inferred).

2.1. What is needed to achieve sustainable mobility?

Transportation is at the heart of modern civilization: it drives progress and brings people together. Throughout the world, the average person spends about one hour a day traveling, a constant figure over the past 50 years. According to the World Business Council for Sustainable Development, ***Sustainable Mobility*** is the ability to meet the needs of society to move freely, gain access, communicate, trade, and establish relationships without sacrificing other essential human or ecological values today or in the future (Banister *et al.* 1997). A holistic study carried out by research and stakeholders during the Mobility 2001 meeting, has pointed out the interdependencies among transport modes, technologies, systems, and institutions crucial. The report highlights eight grand challenges:

- ensuring that transport systems serve essential human needs;
- adapting vehicles to the evolving requirements on emissions, fuel use, capacity, and ownership structure;
- reinventing the relationship between public transport and the private car;
- reinventing the process of planning, developing and managing mobility infrastructure;
- reducing carbon emissions;
- resolving the competition between personal and freight transport for the use of infrastructure;
- tackling congestion;
- building institutional capacity.

One of the basic problems to deal with these challenges is related to the lack of knowledge about movement. We know very little about how human behavior can affect sustainability, the environment, and the scenic quality, as we see in processes such as urban sprawl, intensive outdoor recreation, city expansion, and infrastructure developments. The main reason is the lack of information about the behavioral patterns of people's movements on the landscape. In principle, the Information and Communication Technology (ICT) made possible to gather very large data sets containing location information from mobile devices over time.

The GeoPKDD project is developing an innovative knowledge discovery process for finding patterns of trajectories of mobile devices. This project is a pioneer attempt to provide information about how often, how long, and where people have been. Privacy-aware methods will also be developed to protect the individual. These new kind of information will be useful to several applications.

The workshop on Sustainable Mobility was organized for getting a community input and feedback on the issues and repercussions of discovering behavioral patterns from people in motion using privacy preserving data mining, spatio-temporal ontologies, and visualization techniques that can enable innovative movement-aware applications for sustainable mobility (See Appendix 1 for the CFP). Specifically, the workshop provided a forum for discussion based on different perspectives from data providers, scientists, application developers, and users. The two-day workshop was host by the Waag Society in Amsterdam, and 33 participants from Europe, Canada and USA attended the event (See Appendix 2 for a detailed list of participants).

Several topics were considered, but were not limited to (See Appendix 3 for the Workshop Programme):

Requirements

- Experiences from mobile telecommunication companies in providing the location of mobile customers
- Experiments on tracking people using mobile devices
- How is the mobile user experience different from the fixed user experience?

Strategic Issues

- How can we enable tracking people to be made as seamless, uncomplicated and reliable experience on mobile devices?
- How urban form characteristics (e.g. urban shapes and transportation network types at different scales) may be expected to influence people's movements on the landscape?
- How trajectory patterns can be used to test implicit and explicit claims underlying mobility and policies?
- How to unravel trust and confidence problems, as well as security and privacy rights?

Technological Issues

- How can we make better use of existing mobile technologies? Are there gaps? Is there a need for new technologies?
- What technologies are needed to enable uniquely mobile capabilities that are aware of people's behavior on the landscape?

Two main outcomes from the workshop have been identified by the participants. They were:

1. We need to realize that making new concepts matter, if we are going to deal with the main challenges in sustainable mobility. The strategies for sustainable mobility and transport have, in recent years, been launched in many countries and also on the international level. But so far, limited success has been recorded. However, the questions arise how the sustainability of transport systems and policies can in fact be measured, and how the information about movement of people can be used for this purpose.

2. There are three major strategic/technological issues which need considering if the aims of the GeoPKDD project are to be incorporated more adequately:

    - Will privacy prevent the collection of data in the first place? The regulation must be in place to ensure that privacy issues of the citizens are guaranteed.

    - How much data is needed for knowledge discovery? It is not yet clear how much data will need to be collected in order to obtain the patterns (new knowledge) that will help on defining the strategies for sustainability of transport systems and policies.

    - What kinds of patterns can be expected from the GeoPKDD process? Several applications have been described in the workshop, but most of them manipulated small positioning data sets which have been collected for a limited number of people during a short period of time.

2.2. The process of knowledge discovery in databases

The term "knowledge discovery in databases" was coined in 1989 in an effort to describe the overall process within which data mining is a step in extracting patterns from data. In general, it has been defined as “the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad *et al.* 1996, p.6). The proliferation of such a process coincides with an exponential increase in disparate data sets being linked together across place, scale, time, theme and discipline and available to science, government, and industry. In particular, very large databases that are rich in terms of attribute depth and large in the sense of having many records or objects represented.

The process usually involves experimentation, iteration, user interaction, and many design decisions and customizations. Different delineations have been proposed for a knowledge discovery process, including the nine-step process described as following (Fayyad *et al*. 1996):

1. Developing an understanding of the knowledge domain, the relevant prior knowledge, and the goals of the user.

2. Creating a target data set, selecting a data set, or focusing on a sub-set of variables or data samples, on which discovery is to be performed.

3. Data cleaning and pre-processing.

4. Data reduction and transformation.

5. Choosing the data mining task.

6. Choosing the data mining algorithm(s).

7. Data Mining for a particular form of representation such as classification rules or trees, regression, clustering, etc.

8. Evaluating and validating the results.

9. Consolidating discovered knowledge: incorporating this knowledge into the performance of the system, or simply documenting it and reporting it to users.

Although this list might suggest a sequence of steps, a knowledge discovery process is in fact a *random process* in which the steps are often carried out by an unsystematic approach and do not follow a straightforward analysis. Moreover, many people do not realize that these steps have been treated as separated activities, which with their own principles, procedures and limitations. One of the main reasons relies on the fact that a knowledge discovery process has been a combination of individual techniques that are built from a base in the fields of databases, pattern recognition, artificial intelligence, machine learning, and has strong ties to related efforts in information visualization and to exploratory data analysis in statistics (Cook *et al*. 1997, Card *et al*. 1998). Table 1 summarizes some of the approaches developed by different scientific communities to carry out the steps of mining (step 7), validating (step 8) and reporting the findings (step 9) for classification purposes. The aim of this table is not to describe all the techniques already developed for classification, but rather to illustrate how disparate the examples are, and how unfortunately, there is little research integrating and comparing these techniques. The most representative example is given by the database community who has taken different conceptual and implementation techniques for data mining and reporting that are not necessarily compatible with each other. These approaches are joined together at a merely system level that does not guarantee a better understanding of what a knowledge discovery process is for, and as result, hampering a useful exploration of large databases.

Table 1: Some KDD steps and the respective techniques developed by different scientific communities for classification

Scientific Community / KDD Step	Databases	Statistics	Artificial Intelligence	Information Visualization
Mining	Classification Rules	Local pattern analysis and global inferential tests	Neural Networks	Visual data mining
Validating	Computational models for interestingnes s, confidence and support measures	Significance tests	Learning followed by verification using a test data set	User usability tests
Reporting	Rule lists	Significance power	Likelihood estimation and information gain	Visual communication

A different delineation has been proposed by Ramakrishman and Grama (1999) based on a taxonomy in which a knowledge discovery process is described according to four perspectives on how knowledge is acquired in the process. The first perspective, and actually the most common, is *induction* with its origin in Artificial Intelligence and Machine Learning, in which the process is based on the "learning-from-examples" concept. This is reflected by the extensive number of existing data mining algorithms that can extract generalized rules from a target data set and summarize the relationships between attributes at higher concept levels. Some examples include the attribute-oriented induction method (Cai *et al.* 1991, Han and Fu 1995) which has integrated learning-from-examples algorithms with database operations (e.g. group by). Some authors have investigated attribute-oriented induction methods for extracting generalisation hierarchies for spatial data (Wang *et al.* 1997, Han *et al.* 1997).

The *compression perspective* emerges from the work of the 14th century philosopher William of Occam, in which the Occam's razor concept is stated as "entities are not to be multiplied beyond necessity". The developments in computational learning theory and the feasibility of models based on minimum encoding inference, such as MML-Minimum Message Length (Wallace 1990), have played an important role in establishing a solid theoretical foundation to this perspective. The Occam's razor is often used as a guiding principle in model selection in data mining, which suggests a "good" model should use any relevant variable, relationship, or behavior but ignore all irrelevant ones. Models should capture the essence of an application domain under study by searching for simplicity. Some examples of modeling algorithms are projection pursuit, neural networks, decision trees, and adaptive splines (Fayyad *et al.* 1996). All these models assume the availability of a training data set, and the goal is to find a model to predict *y* from *x* that will perform well on a new data set.

In contrast to the previously described perspectives, the *querying perspective* is based on discovering knowledge through database query languages. In general, database models have been developed for storing and querying data, and they still need to be proven to be "good" models for data mining. Most database management systems do not allow the type of data interaction that a knowledge discovery process requires. Nevertheless, several research efforts have been focused on enhancing query languages such as SQL (Structured Query Language), mainly because most of the data is available from commercial databases and warehouses. Some examples are the semantic query optimization approach by using semantic rules to reformulate a query (Hsu and Knoblock 1996, Siegel 1988 and Shekhar *et al.* 1993) and the FOIL (Quinlan 1990) approach using Horn-clause definitions in a query.

Finally, the query perspective is closely related to the *approximation perspective*, which relies on the previous knowledge of a model (e.g. a database schema) in order to find some hidden structure in the data. For example, linear algebraic matrix approximations have been developed to identify hidden structures in text data without using a simple keyword matching (e.g. Latent Semantic Indexing, patented by Bellcore).

2.3. The geographic knowledge discovery process

The term "geographic knowledge discovery" was coined later, having the acronym of GKDD, and representing a special case of knowledge discovery in databases since it required specialized tools and provided unique research challenges to deal with space and time. The process has been defined using the previous mentioned steps, but critical research areas have been identified as developing and supporting geographic data warehouses, richer geographic data types, better spatio-temporal representations, and user interfaces (Harvey and Han 2001, Gahegan *et al.* 2001, Wachowicz 2001). In an effort to frame a geographic knowledge process in the context of spatio-temporal environmental data, MacEachren and Wachowicz (MacEachren *et al.* 1998) have proposed a conceptual framework for the integration database and visualization techniques, emphasizing a merger of meta-operations for the GKDD steps. The knowledge discovery system specifically considered location, time, and attribute aspects of each data entity during all steps of analysis (from pre-processing, through application of data mining tools, to interpretation).

A different approach was proposed by Aldrige (2000) on advocating the concepts of extensional knowledge (i.e. facts) and intensional knowledge (i.e. rules) based on Pawlak's theory on notions of equivalence relations, generalization, induction, deduction, and supervised and unsupervised learning. Empirical knowledge of real-world phenomena was applied to represent extensional knowledge on choropleth maps, and the results shown GKDD as a process of inducing non-trivial, potentially useful intensional geographic knowledge from databases.

However, successful applications of GKDD are not common, despite the vast literature on knowledge discovery in databases. The reason is that, although it is relatively straightforward to find patterns in very large spatio-temporal databases, establishing their relevance and explaining their causes are both very complex problems. In practice, most of the patterns found in a GKDD process may well already be background knowledge of an application domain. Large databases may contain a vast number of hidden patterns which are not necessarily novel or useful. At the moment, a geographic knowledge discovery process has no concept of what is known by experts in a way that the patterns make sense within the context of the current application domain. Addressing these issues requires to consider a knowledge discovery process as a human-centred process, not only in the sense that users need to dynamically interact with the system, but also that knowledge can only be inferred from very large and possibly poorly-understood databases if the effective form of *metaphor* is used.

This is even more relevant for geographic knowledge discovery process where there is a scarce geographic knowledge on the forms of metaphors on inferring knowledge from spatio-temporal databases. Therefore, the need here is for more complex reasoning modes, which could provide the mapping, required for a systematic set of correspondences between metaphors that we try to understand and the patterns found in a geographic knowledge discovery process. This is discussed in more detail in Chapter 3.

2.4. Privacy issues: Involving the stakeholders in a GeoPKDD process

Privacy needs to be addressed from the beginning of the geographic knowledge discovery and therefore needs to be integrated into the complex relationship between patterns and information metaphors. In legal frameworks as well as in the database community, the concept of privacy is generally translated into data protection, or more specifically, the protection of personal data. Personal data is "any information relating to an identified or identifiable natural person [...]; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (EU Directive 95/46/EC, Art. 2 (a)). This definition has two effects: first, it focuses the attention on data (as opposed to people), and second, it focuses the attention on the identification of a natural person as the problem. Thus, it implicitly declares data-processing activities to be "privacy-preserving" if and only if they do not (better, if they provably cannot) involve the identification of the natural person who is the "carrier" of a record of attributes. This notion is reflected in the knowledge discovery literature by regarding goals such as k-anonymity as the defining properties of a privacy-preserving data mining algorithms.

However, the privacy literature suggests that a people view of privacy involves not one but many identities –which can be supported by a number of identity management schemes--, and that there are concerns over profiles independently of identification, and that context is all-important. A true geographic knowledge discovery process should transcend the algorithm-centric and data-centric views. As a result, it should involve a better understanding of an application domain, and the evaluation of the lessons learned from previous cycles of a geographic knowledge discovery process. This means that the complexities of software development and software use must be considered. Different mobile entities release trajectory data in different contexts as a result of complex sets of interactions. Moreover, different mobile entities have partially conflicting interests in the information processed and available through such systems, and each particular system exists in an environment populated by many systems.

Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, the Fair Information Practices (FIP) notice, choice, access, and security that were set down as a recommendation in the US and updated by the Federal Trade Commission, or the principles of the EU Privacy Directives, define privacy not only as a matter of concealment of personal information, but also as the ability to control what happens with it (Gürses *et al.* 2006). These are sometimes also referred to as the Eight Principles of "Fair Information Practices" which are one of the following:

- **Collection limitation**: Data collectors should only collect information that is necessary, and should do so by lawful and fair means, i.e., with the knowledge or consent of the data subject.

- **Data quality**: The collected data should be kept up-to-date and stored only as long as it is relevant.

- **Purpose specification**: The purpose for which data is collected should be specified (and announced) ahead of the data collection.

- **Use limitation**: Personal data should only be used for the stated purpose, except with the data subject's consent or as required by law.

- **Security safeguards**: Reasonable security safeguards should protect collected data from unauthorized access, use, modification, or disclosure.

- **Openness**: It should be possible for data subjects to learn about the data controller's identity, and how to get in touch with him.

- **Individual participation**: Data subjects should be able to query data controllers whether or not their personal information has been stored, and, if possible, challenge (i.e., erase, rectify, or amend) this data.

- **Accountability:** Data controllers should be accountable for complying with these principles.

Beyond regulations and legal frameworks, privacy brings a new component to a GeoPKDD process; that is: the stakeholders. There may be different degrees of privacy interests from different stakeholders within a GeoPKDD process. Basically, three main groups of stakeholders might play an important role in dealing with the privacy issues in a GeoPKDD process. They are:

- **sensor carriers**: the ones who produce the positioning data sets. They should authorize the level of privacy expected for collection, use, openness, and individual participation in a GeoPKDD process.

- **data collectors and miners**: the ones interested in collecting the positioning data and developing the data mining algorithms. They must ensure the level of privacy required for data collection, data quality, and security safeguards.

- and **the experts of an application domain**: the ones interested in applying the outcomes of a GeoPKDD process. They should also define the level of privacy required for the collection, purpose, and use of the results of a GoPKDD process.

Figure 1 illustrates the relationship among the three groups of stakeholders who are involved in a GeoPKDD process. For example, the collection of data about AIDS patients and their movement may be seen differently from a public health or epidemiology perspective –with the objective of contact tracing-- versus from an individual perspective –with the objective of protecting against social or workplace discrimination. Such cases point out to the necessity of multilateral privacy and security requirements analysis to accompany a GeoPKDD process in which the privacy and security interests of all the stakeholder need to be documented, conflicts worked out, and constraints elicited for the privacy-aware GeoPKDD process.

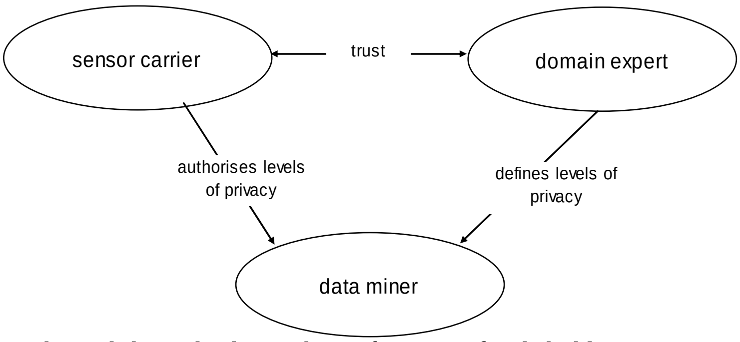


Figure 1 – The multilateral relationships of groups of stakeholders in a GeoPKDD process

Further, positioning data, and hence trajectory models, are particularly sensitive information to sensor carriers because of the specific characteristics of a geographic environment. The sensor carriers cannot avoid being at a location at any point of time (one cannot “opt-out of being somewhere”). Therefore, their impression of lacking self-control and comprehensiveness of surveillance is particularly pronounced in comparison to data miners and experts concerns about privacy. In addition, positioning data sets in combination with geographic information allow many inferences because of rich social background knowledge. One reason is that because of physical constraints, the number of interpretations of being at one location is often bounded (e.g., visits to a doctor specialist in the treatment of AIDS victims). Another reason is that there are typical spatio-temporal behavior patterns (a location where a person habitually spends the night without moving is most likely that person’s home). This may bring further constraints on the positioning data to be collected, processed and released by data miners as well as experts of an application domain.

The preservation of privacy in ubiquitous environments during data collection is difficult, if not impossible. Multiple channels collect different kinds of physical information (e.g. radio frequency fingerprints) that can be used to precisely identify the trajectories of a given sensor carrier. Therefore, opting out may not be a possibility during data collection, and may bring about legal or social conflicts. Especially since trajectories coupled with a-priori knowledge on places and social contexts, can be analyzed by data miners to infer individual identities used to link profiles of individuals, or classify persons into previously defined groups.

Although groups are made up of aggregate patterns, which are not intended to identify single individuals or entities, they may still provoke the problem of group discrimination as a result of the processed data. All members of a group in a specific area or with certain trajectory patterns can be categorized into one group and the discovered category may be attributed to all members when later the discovered patterns are applied. For example, if a specific region is identified as having a high-rate of cancer, then all persons linked to this region may be offered insurance at higher-rates than in other areas, regardless of how heterogeneous the group members in reality are. Therefore, for an expert of an application domain, the concealment of exact rules, which cover the complete population, or even weaker rules used to classify groups into sensitive categories may also breach privacy of individuals, although they do not identify individuals.

3. FUTURE APPLICATION DOMAINS FOR A GEOPKDD PROCESS

This chapter relates the GeoPKDD process with the prospect of potential application domains. Three application domains have been identified as the most promising on implementing a GeoPKDD process. They are: **transport management**, **spatial planning** and **marketing**.

3.1. Transport Management: The integration of multimodal choices

Transport or transportation refers to the movement of people and goods from one place to another. The term is derived from the Latin trans (across) and portare (carry). Transport management is aimed at solving the problems between infrastructure (e.g. transport networks) and operations (e.g. road traffic control). Modalities are a combination of infrastructure and operations. In this scheme, both private and public transport modalities are managed by a planning authority having control of some decision variables: road pricing, transit ticket prices, and the service characteristics of transit. The multimodal transport system is subject to some constraints: physical and environmental capacity constraints, and budget constraints. For example, in some cases an upper bound is imposed on the ticket price, in order to help people who are captive of transit.

The overall goal of the integration of multimodal transportation is to develop innovative solutions to fundamental problems, working in multidisciplinary teams that explore a range of expertise including statistical analysis, operational research, psychology, engineering, marketing, visual culture and aesthetics, and IT (Richman 1998). One of the problematic issues of multimodal transportation is the routing of people towards and through inner cities. Cities tend to expand at their boundaries. As a consequence the city-centre needs to handle increasing traffic flows of different modalities coming into the centre. As result of the growth of many cities the intensity of the traffic in many centers exceed by far the capacity of the centers infrastructure as it was designed originally. Congestion, conflicts between various modes of transportation, parking problems and the exploitation of alternative routes through residence areas causing nuisance to the residence are the result. Traditionally the problem of congestion of inner cities is dealt with by imposing all kind of parking restrictions, one way-street policies, and road toll.

The prospects of providing travelers and planners with multimodal information about their trajectory behavior will for the first time, integrate people's behavior information with transportation information (infrastructure and operations). If such integration can be achieved based on the patterns observed from the trajectories of people, and their relations between space, time, and activities; we will have the possibility of presenting to travelers, and in particular drivers, with comparable information on travel options across modalities. An integrated multimodal information service will have a great potential to inform and influence travel choices, as well as identify the requirements from, and potential benefits of knowing about the patterns of people's trajectories. Geographic knowledge about the dynamics of movement, the distribution and composition of modalities might improve the knowledge about accessibility of inner cities as well.

The travel data is usually described as in terms of individual modalities (e.g. car or public transport), commuting time or distance, spatial distribution of jobs and housing locations, total vehicle miles traveled, mode of average trip lengths, and congestion on links and intersections. However, evidence from the National Travel Survey data for Great Britain suggests that measures such as driving speed at different times of the day do not show large variations (Stead and Marshall 2001). Moreover, Crane (2000) firmly states that there is evidence to suggest that the travel diaries systematically overstate household travel, and as a result, short journeys may be under-recorded.

Most common metaphors used in this application domain are related to the movement-as-accessibility metaphor which defines how people move from one location to another as pedestrians, or taking cars, bikes, or public transportation. For a geographic knowledge discovery process this might imply the deductive search as one of the following examples:

- Discover how a set of point patterns evolves from time t1 to time t2, in terms of a specific mode of transportation;

- Discover an observation window (spatial and temporal extends) where point patterns reveal a change of mode of transportation;

- Discover the rules which explain a spatial distribution of a set of point patterns at a given time in terms of a specific mode of transportation.

The geographic knowledge discovery process of patterns of moving people and their respective trajectories will have an impact on understanding the relationship between modality choices and trajectory patterns. For example, an increase of trajectory patterns at a local scale may result in a reduction on accessibility, which is defined as the ease of reaching a particular destination. Moreover, the knowledge of such trajectory patterns will play an important role in studying the route conditions from effects such as hazards, noise, traffic jams, and visual pollution.

Metaphors are needed to allow the choice of the scale on what the type of trajectory patterns can be found. One example is given by information about accessibility which is usually known by gathering data from models and empirical surveys about how various individuals organize their daily activities in time and space, as well as the travel involved (Axhausen and Gärling 1992).. Simulations, questionnaires as well as statistical data (when available) are some examples of sources used to infer knowledge about how people commute during the day, month, or year. Most common socio-economical statistics used are population growth (e.g. annual growth rate), social statistics (e.g. levels of education, sex, age), economic growth (e.g. GNP, household incomes), employment structure (e.g. commercial, financial and service sectors), land use policy and regulations, land use patterns and built up areas (e.g. urban growth), transport statistics, commuter's demographic and household characteristics.

This kind of information would allow the experts to check if information can be inferred from the trajectory representation that may breach the privacy requirements of any of the sensor carriers. If so, the experts may consider using one or a number of privacy preserving methods to protect these actors. Trajectory patterns are the outcome of a highly complex interplay between personal and household characteristics and features of the urban/rural environment. The development of a geographic knowledge discovery process will be essential to be able to gain knowledge on how some assumed spatio-temporal relationships between urban/rural forms and trajectories behavior will result in activity patterns such as direction of commuting (e.g. within inner, within outer or cross commuting in a metropolitan area), modality (distance, type of transport – private, public, non-motorized), and time (e.g. time spend by commuters who traveling to work). Some examples are given below:

- Discover accessibility patterns that explain the occurrence of shopping activity with its corresponding transportation modality.

- Discover the dependencies between working and leisure activities according to a specific transportation modality.

- Detect the occurrence of an unexpected activity.

Here, the data miners and domain experts need to observe the classifications that they infer and check to see whether the "category anonymization" requirements of the different sensor carriers are breached.

3.2. Spatial Planning: The adaptation of space to human behavior

Spatial planning is aimed to change the organization of a geographic environment to meet the demands of society. Demands of society continuously change as the result of change in the society but also due to change in the geographic environment itself. Demands result into claims upon existing spatial functions. As space becomes a limited resource the geographic environment is expected to fulfill multiple functions (Valk 2002). People compete for the same resources. Especially rural areas are under increasing pressure and need to fulfill multiple functions (Cammen and Lange 1998). They have to be attractive for recreation but also productive in terms of agriculture while at the same time they provide the space necessary to meet the claims of expanding urban areas. There is a clear shift of rural areas having primarily a production function towards rural areas that are regarded as differentiated residence area. At the same time planning shifts from planning bases upon primarily hierarchical principles towards more actor oriented and participatory types of planning (Geertman 1996, Woerkum 2000).

In spatial planning, location-allocation representations and methods have been developed when positioning data sets were scarce, difficult to obtain, and models were deterministic, or entirely predictable (Hägerstrand 1970). In principle, mobile technologies made possible to gather very large data sets containing movement information from mobile devices over time. This has opened the opportunity to deal with the location-allocation problems from a people's perspective in spatial planning.

Currently many of the decisions taking in spatial planning take into account the land uses of a spatial environment. Currently, land use describes the activity the landscape is used for. At the national level, current types of land use are important to decide about new development scenarios for a region in terms of defining recreational areas, nature and urban areas. At local levels, land use is an important parameter for deciding about locations for living, industrialization, and leisure.

The geographic knowledge about the patterns of the movement of people on a spatial environment might refine the concept of land use in spatial planning (Ratti 2004, Pulselli *et al.* 2005). Currently land use is mostly based on the knowledge obtained from a spatio-temporal classification of features on a landscape. These classifications are static as they do not include human behavior. Knowledge about movement of people and masses might add additional insight beyond that of the traditional land use concept, mainly in terms of understanding the effects of the landscape on the movement of people and vice-versa. As a result, land use could become the activity metaphor based on the movement of people, rather than the location of its features.

A second metaphor is related to the general concept of functionality of space. Functional spaces are spaces which are designed to fulfill a specific task. Typical examples focused on in this research are shopping areas, airports, areas for large scale events, and parks. Knowledge of the movement of people in these types of areas is required for the situating of shops, shop-types, and checkpoints. More over the dimensioning of pathways, gateways and emergency evacuation routes might benefit from additional knowledge about the spatial temporal dynamics of moving crowds.

The above mentioned examples of applications in spatial planning would benefit from the gathering of positioning data of moving people on a landscape into a trajectory data warehouse. This type of data is not commonly used in the process of designing spaces nor is it commonly available. Many of the design decision are traditionally based on estimations, extrapolation from known cases and simulation models.

Very little is known about the growth, shift, or even decrease of movement patterns of people in urban and rural areas. There is no universally accepted standard classification of human activities, and the association among activities that generates patterns of trajectories at individual, organizational, and urban/rural form levels is not well understood. The basic assumption is that individuals and households try to meet their basic needs and preferences by participating in activities, while the environment (urban/rural forms) they live in offers them the opportunities and constraints to do so. The geographic discovery process needs to be essentially targeted to finding linear patterns of trajectories which can be understood be a planner and designer. This implies that attributes of trajectories depend on the type; the scale and the goal of the planning and the function the space need to fulfill. Little or no knowledge exist about these issues yet.

The main metaphor is movement-as-urban form. Mobility is the trajectory of individuals that is dependent on urban forms such as transportation networks and land use (Spence and Frost 1995. Land uses support human activities. Those activities are spatially separated. People need transport to go from one place to another (from home to work to shop back to home for instance). Transport is a "derived demand," in that transport is unnecessary but for the activities pursued at the ends of trips. Therefore, a certain land use type might enable common activities to occur close a specific place (e.g. housing and food shopping), as well as places with higher-density development closer to transportation lines and hubs. Poor land use concentrates activities (such as jobs) far from other destinations (such as housing and shopping).

Multifunctional land use can be defined as combining various socio-economic activities in the same area. The basic idea is to save scarce space (Vreeker et. al. 2004). An important aspect in the planning of multifunctional land use is to try to integrate activities as harmonized as possible and when possible strive to a synergy between two (or more) activities. In many European countries there is a constant friction between the aim of keeping particular environments as undisturbed as possible to facilitate the habitat functions for flora and fauna or allowing these environments to functions leisure areas for citizens to divert from their demanding daily lives. The concept of multi-functional land use tries to integrate both activities of a geographical environment. The challenging factor is to provide ample opportunities for recreational activities while preserving and developing nature. Traditional instruments which are part of the “toolbox” of the planner are zoning, routing, and temporary closing. Mountain bikes are for example bared from highly sensitive area or only allowed on special tracks. Some breeding areas are closed for traffic during certain hours a day or during the mating and breeding seasons.

Most of these decisions are made in absence of knowledge about the effects of the dynamics recreational activities on the quality of the nature in that area. No knowledge about the current behavior of the various leisure seekers is present. Information about intensity and followed routes are mostly estimated or based on incidental counting. Knowledge about the patterns of movement of visitors of nature areas and the type of activities might improve the harmonization of multifunctional use of nature areas. Decisions can be made based on measured patterns of movement and related to observed effects on the environment. This might lead to more precise or flexible zoning, routing, or access policies.

Additionally geographic knowledge discovery might be aimed at finding patterns that represent conflicting behavior. In relative small areas (like the ones found in the Netherlands) often there are irritations between for example walkers and mountain bikers or horseman. Insight in the periods and locations of potential conflicting trajectories might allow managers to improve the way they handle the various types of visitors leading to fewer conflicts. The research challenge for these applications is to discriminate amongst the different type of movement patterns or trajectories and assign the properties which are relevant to these trajectories. The question of how to analyze, the differences between, for example, the patterns of trajectories of a hiker and a mountain-biker is not a trivial one. No methods are known yet that can deal with recognition, and classification of these subtleties.

3.3. Marketing: The shift towards movement aware marketing

Currently marketing mostly is done based on customers' profile which, normally, is statically defined. Such profiles are based on characteristics like gender, age, income, family situation, and purchase history. Based on the customers characteristics strategies can be developed to determine a need for purchasing certain products amongst potential or existing customers. Traditionally the success of marketing depends on what is called the marketing mix of four P's: product, price, promotion, and placement. This rather traditional view on marketing has been criticized since it has a main focus on a company or marketer rather than being a consumer centered strategy. Furthermore, the traditional marketing mix model does not serve the marketing of services very well. Last decades various elements of the marketing mix have been suggested to be changed. For example, the placement needs to be converted to convenience and promotion to communication. More elaborate knowledge about potential customers is also needed to be improved in marketing mix models.

Current computer and database technology enable the storage, processing and analyzing of much more variables of people behaviors that are relevant to marketing. An important development is that of geo-marketing. Geo-marketing implies the use of Geographic Information Systems (GIS) to include location information into the marketing mix. Based on spatial analysis, additional knowledge and insight might be gained about for example the spatial distribution of income and demographic composition of districts.

The main metaphor is that of movement-as-personalization. Using movement data, marketeers and service organizations can better target their information and services to specific users depending on their activities, relations, and locations. The "scaring" outlook of many people to be spammed by location based advertisements, generated by relative dumb LBS can be alleviated by providing more intelligent information based of movement behavior.

Recently, Location Based Services (LBS) have been added to the geo-marketing sector as a new marketing tool. Using LBS, marketers can pinpoint their marketing mix and enhance their communication with potential customers based on their exact location and time. Most obvious examples are push marketing base on SMS messages when a customer passed a shop he or she might be interested in (mobile advertising).

As a next step, LBS might develop further towards Movement Based Services (MBS). One of the differences with "traditional" LBS is that it will take into account the history, behavior and relation with other movements. LBS only provides the context from the users, and the environment (who is where a time t). MBS has he potential to add to this, knowledge about what he/she did, how he/she did it and with whom.

The data used for LBS is usually only describes the location in space, a caller id and a time stamp. For MBS, information about the followed tracks, the movement characteristics (speed, acceleration, periodicity in movements etc) need to be added and stored. This is a substantial shift in how to deal with the data. Currently LBS do not need the analysis and storage of locations per se. The majority of LBS are user or event based. They discrete event approach of LBS only require the data for the moment the data is requested. MBS typically requires the maintenance and storage of data to be able to infer patterns out of it. One of the concerns at this stage is privacy. As data is required and requesting of the movements of individual people preserving privacy is a first requirement. The ubiquitous nature of the collection of movement data makes privacy even a more pressing concern. The main metaphors in this application domain are related to construction of the public/private divide and freedom from intrusion. The control about what and when data about movement activities should be private or public need to be clear to and perhaps in control of the person carrying a sensor. The right to be let alone is a basic right of humans. Especially in marketing based application the control of the right should be part of the decision making about what part of the reality space should be sample and registered.

Little is known about the use of movement patterns for marketing purposes. There is barely research after the marketing related behavior and movement behavior. In principle there are two models that make use of movement data: the first, the consent model is based on informing, or assisting users with information or services based on authorization given by them. This means that people decide what type of services or information they are interested in. Based on their movement behavior these services/information are provided to them when required or needed in an intelligent fashion. The second model, the informed model, users receive targeted information based on their movement behavior, location, time and the behavior of others, i.e. the behavioral pattern they are part of. The challenge of the informed model is to couple information about movement behavior with other sources of (behavior) information like shopping history, and non-behavior information. For both models the information should, in principle, be able to infer the following basic knowledge on:

- what someone is doing;
- how someone is moving and with whom;
- who else are moving (coinciding patterns).

Privacy-aware services can be developed which are currently hard to realize such as presenting services of information based on the type of movement. If you are, for example, driving a car on a crowded highway you probably like to have information presented differently than when walking around in a city. So movement based behavioral information might facilitate also the means and methods by which information is presented to customers.

Therefore, the inferential space tries to discover the causes and consequences of movement behavior. The discovery of geographic knowledge related to privacy aware marketing is mainly targeted to finding groups that show similar behavior and to determine if this behavior is interesting given a certain marketing goal. Examples of typical queries are:

- Discover the general patterns that explain the behavior of certain groups of people given a marketing perspective. Using the characteristics of these groups marketing can be targeted and personalized.

- Discover the dependencies between movement behavior and the effects of personalized movement aware marketing. Can movements of people be influenced by certain marketing actions or are the effects of marketing dependent of movement behavior?

- Discover the type of information appreciated by people when they are moving at a certain time, modality and location.

Currently the above types of knowledge discovery cannot be carried out or only limited based on marketing research.

4. POTENTIAL APPLICATIONS FOR THE GEOPKDD DEMONSTRATORS

This Chapter describes briefly the main characteristics and requirements of five potential applications to be developed for the GeoPKDD application demonstrators of Phase 2 – Application Demonstration of the GeoPKDD project.

4.1. Demonstration Case 1: Netherlands

- Data provider: LNV (Ministry of Agriculture and Fisheries, the Netherlands)
- Type of data: large volume of GPS positioning data sets containing information on the routes of manure cargos in the Netherlands. Source and destination are known.
- User needs: to discover anomalies among the routes (trajectories) that can be used to identify irregular dumps of manure in non-authorized areas.
- Data availability: The data is currently used within the Manure Transportation System, and it could be available through an agreement contract.
- Application Domain: Transportation Management
- Application requirements:
  1. Discover anomalies such as stopovers at unexpected locations or deviation from a normal route; the source of origin could be different from what was originally planned.
  2. Discover associations of routes (trajectories) with particular locations that are not obvious. For example, there is a possibility of a hidden association apart from the pre-defined points such as the origin and destination of the route. In this case, although the route shown is from A to B, it is possible to find if the route has any association with other locations, such as C, which can then help in identifying potential illegal dumps.

4.2. Demonstration Case 2: Noord Brabant

- Data providers: Vodaphone and Logica CMG
- Type of data: large volume of phone usage (GSM data) which was filtered to obtain the data when the caller is in a car.
- User needs: to discover patterns of change in traffic density and velocity, including causes and effects, in order to support sustainable mobility.
- Data availability: the data is currently used within the Monitoring Transportation System (MTS) to generate traffic views in the province of Noord Brabant, the Netherlands. It could be available through a server agreement with LogicaCMG and the province of Brabant.
- Application Domain: Transportation Management

- Application requirement:

1. Discover knowledge rules of traffic behavior that can identify changes in traffic density and velocity.

4.3. Demonstration Case 3: Pisa

- Data provider: WIND

- Type of data: WIND data (Figure 2) and the data available from the survey of individual daily behavior in the municipality of Pisa (3500 trips, 806 individuals, 38 administrative geographical areas).

- User needs: discover similarities among trajectories of human activities (commuting, family visits, shopping) at different levels of spatial and temporal hierarchies.

- Data availability: Contact with N. Salvati, University of Pisa needs to be made.

- Application Domain: Spatial Planning

- Application requirements:

1. Discover clusters of similar co-locations of trajectories at different levels of spatial hierarchy (e.g. land use and municipality region), as well as temporal (daily, month, etc.)

2. Discover association rules that explain human activities. For example, at neighborhood level, people living in similar neighborhood in terms of income, have similar movement patterns. Or people with similar commuting patterns in the same neighbor have different shopping patterns.

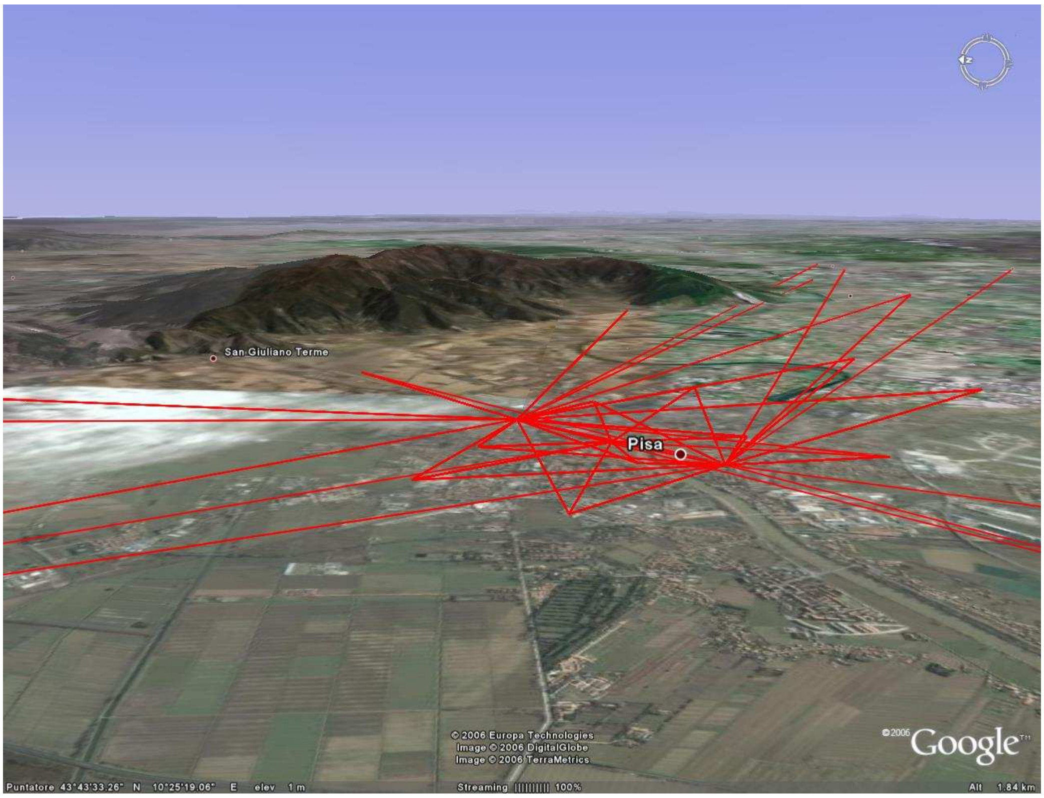


Figure 2 – Visualization of the GSM data sets provided by WIND

4.4. Demonstration Case 4: Wageningen

- Data provider: WUR - ALTERRA
- Type of data: GPS positioning data and socio-economic statistics (CBS). The use of the data generator might be needed (Figure 3).
- User needs: discover similarities among trajectories that identify human activities (commuting, family visits, shopping) at different levels of spatial and temporal hierarchies.
- Data availability: Data is already available.
- Application Domain: Spatial Planning
- Application requirements:

1. Discover clusters of similar co-locations of trajectories at different levels of spatial hierarchy (e.g. land use and municipality region), as well as temporal (daily, month, etc.)
2. Discover association rules that explain human activities. For example, at neighborhood level, people living in similar neighborhood in terms of income, have similar movement patterns. Or people with similar commuting patterns in the same neighbor have different shopping patterns.

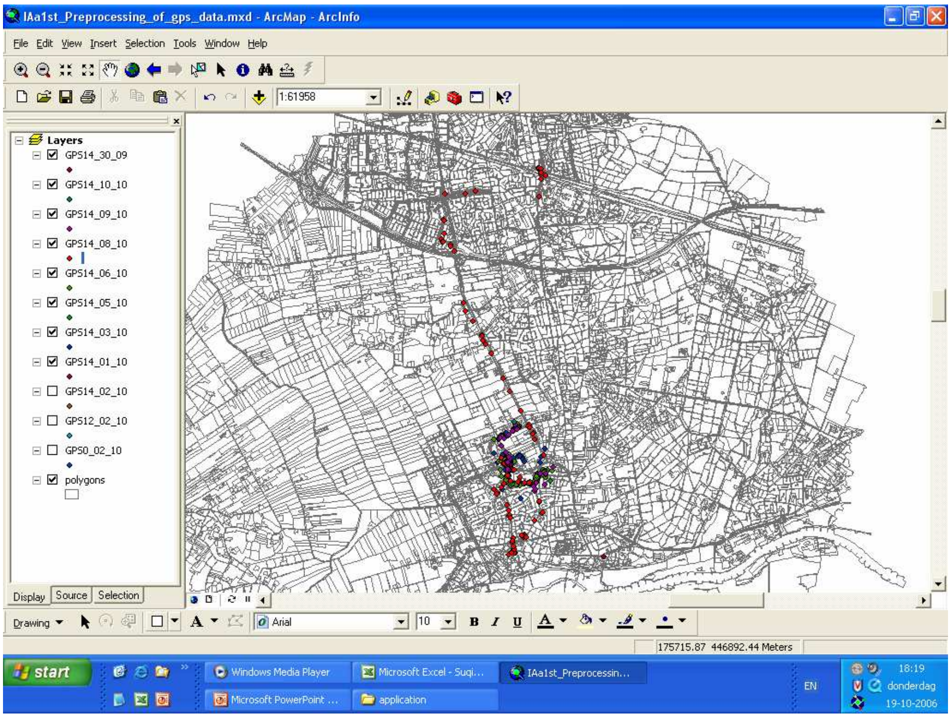


Figure 3: Example of trajectories obtained from the Wageningen experiment

4.5. Demonstration Case 5: Milano

- Data provider: Commune di Milano
- Type of data: large data sets containing GPS positioning data of cars
- User needs: discover similarities among trajectories that identify human activities (commuting, family visits, shopping) at different levels of spatial and temporal hierarchies.
- Data availability: It will be available through a contract agreement with the Commune di Milano.
- Application Domain: Transport Management
- Application requirement:

1. Discover behavioral patterns that can help a better management of the local transport network.

4. CONCLUSIONS

This report describes three application domains, focusing on their main characteristics and requirements. It also illustrates the potential applications supported by aggregative location-based services, societal and economical applications enabled by aggregative geographic knowledge extractable from the GeoPKDD process. Three feasible and promising application domains have been highlighted as being transportation management, spatial planning, and marketing. The selection of these application domains were based on the following criteria:

- availability of movement data such as GPS and GSM data sets;
- need for innovative solutions for the current problems and issues within a specific application domain;
- focus on the challenges on supporting sustainable mobility.

Based on these criteria, five applications have been selected as potential demonstrators to illustrate the GeoPKDD process. Two of them (Demonstration cases 2 and 5) have been selected as the potentials candidates to be used for the implementation of the GeoPKDD demonstrators.

One of the main conclusions from this task T4.1, is that the availability of large positioning data sets is not a reality yet. The potential data providers such as operators, municipalities, and companies, lack the operational and organizational infrastructure to gather and store large volumes of positioning data sets.

Another important conclusion is related to the user needs and requirements. Most of spatial planners, transport managers, and marketing managers had difficulties in understanding the potential use of the outcomes of the GeoPKDD process. As a result, the identification of user requirements has been a difficult task.

REFERENCES

Abowd, G.D. Atkeson, C.G., Hong, J., Long, S., Kooper, R. and Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, 3(5):421-433, October 1997.

Aldridge, C.H. (2000). A theoretical foundation for geographic knowledge discovery in databases. *Proceedings First International Conference on Geographic Information Science*, Georgia, USA.

Axhausen K.W. and Gärling, T. (1992). Activity-Based Approaches to Travel Analysis: Conceptual Frameworks, Models and Research Problems. *Transport Reviews*, **12**, pp.324-341.

Banister, D.; Watson, S. and Wood C. (1997) Sustainable cities, transport, energy, and urban form. *Environment and Planning B: Planning and Design*, **24**(1), pp. 125-143.

Brunk, B., and Davis, B. (2002). SDAT enterprise: Application of geospatial network services for collaborative airspace analysis. *Proceedings of the CADD/GIS Symposium*, San Antonio.

Cai, Y., Cercone, N. and Han, J. (1991). Attribute-Oriented Induction in relational databases. In *Knowledge Discovery in Databases*, G. Piatesky-Shapiro and W.J. Frawley, AAAI Press, pp. 213-218.

Cammen, H. v. d. and Lange, M. A. d. (1998). *Ontwikkelingen in wetenschap en technologie : sturingstheorieen en landelijke gebieden*, Nationale Raad voor Landbouwkundig Onderzoek, Den Haag.

Card, S., Mackinlay, J. and Shneiderman, B. (1998). Information visualization. *Readings in Information Visualization* (Eds. Card, S., Mackinlay, J. and Shneiderman, B.), San Francisco: Morgan-Kaufmann, pp. 1-34.

Cook, D., Buja, A., Cabrera, J. and Hurley, C. (1995). Grand tour and projection pursuit. *Computational and Graphical Statistics*, **4**(3), pp. 155-172.

Chen, G and Kotz, D. (2000). *A survey of context-aware mobile computing research.* Dartmouth Computer Science Technical Report TR2000-381. 16p.

Crane, R. (2000) The Influence of Urban Form on Travel: An Interpretive Review. *Journal of Planning Literature*, **15**(1), pp. 3-23.

Deadman, P. (1999). Modelling individual behaviour and group performance in an intelligent agent-based simulation of the tragedy of the commons. *Journal of Environment Management*, **56**,159–172.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery: An overview. In: *Advances in Knowledge Discovery and Data Mining* (Eds. Fayyad, U., Piatetsky-Shapiro, G, Smyth, P. and Uthurusamy, R.), Cambridge, MA: AAAI/MIT Press, pp. 1-34.

Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T.M. (2001). The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Science, Special issue on research challenges in geovisualization*, **28**(1), pp.29-44.

Geertman, S. and Stillwell, J. (2003). *Planning support systems in practice*, Springer, Berlin [etc.]. Isn1685707 en.

Georges, T. M., Harlan, J. A., Meyer L. R., and Peer, R. G. (1993). Tracking hurricane Claudette with US Air Force One over the horizon radar. *Journal of Atmospheric and Oceanic Technology*, **10**, pp.441–451.

Gürses, S.F., Berendt, B., Santen, T. (2006). Multilateral security requirements analysis for preserving privacy in ubiquitous environments. *ECML PKDD 2006 Workshop on Ubiquitous Knowledge Discovery for Users (UKDU'06).* Submitted.

Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, **24**, 7–21.

Han, J. and Fu, Y. (1995). Discovery of multiple-level associations rules from large databases. *Proceedings of the International Conference on Very Large Databases*, pp.430-431.

Han, J., Koperski, K. and Stefanovic, N. (1997). GeoMiner: A System Prototype for Spatial Mining. *Proceedings ACM-SIGMOD 1997*, Arizona

Harvey, M.J. and Han, J. (2001). Geographic data mining and knowledge discovery: an overview. In: *Geographic Data Mining and Knowledge Discovery*. (Eds. Harvey J. Miller and Jiawei Han), London: Taylor and Francis, pp.3-32.

Hsu, C.N. and Knoblock, C.A. (1996). Using Inductive Learning to generate rules for semantic query optimisation. In: *Advances in Knowledge Discovery and Data Mining*. U.M Fayyad *et al.* (eds.) AAAI Press/The MIT Press, pp.425-445.

Hunter, A., El-Sheimy, N., and Stengouse, G. (2005). "Close and Grizzly GPS/Camera Collar Captures Bear Doings", *GPS World*, February Issue, 2005, pp 24 – 31.

MacEachren, A.M., Wachowicz, M., Edsall, R., Haug, D., and Masters, R. (1998). Constructing knowledge from multivariate spatio-temporal data: Integrating GVis and KDD methods. *International Journal of Geographic Information Science, Special issue on geographic visualization*, **13**(4), pp.311-334.

Oppermann, R. and Specht, M. (2000). A context-sensitive nomadic exhibition guide. In: *Proceedings of Second International Symposium on Handheld and Ubiquitous Computing, HUC 2000*, Bristol, UK, Springer Verlag, pp.127-142.

Pulselli, R.M., Pulselli, F.M., Ratti, C. and Tizzi, E. (2005). Dissipative structures for understanding cities: Resource flows and mobility patterns. *Proceedings of the First International Conference on built Environment Complexity, BECON 2005*, Liverpool, England, pp. 271- 279.

Quinlan, J.R. (1990). Learning logical definitions from relations. *Machine Learning*, **5**, pp.239-266.

Ramakrishman, N. and Grama, A.Y. (1999). Data Mining: from serendipity to science. *IEEE Computer*, **32**(8), pp.34-37.

Ratti, C. (2004). Space syntax: some inconsistencies. *Environmental and Planning B: Planning and Design*, **31**, pp.487-499.

Richmond, J.E.D. (1998). Simplicity and complexity in design for transportation systems and urban forms. *Journal of Planning Education and Research*, **17**, pp.220-230

Shekhar, S., Hamidzadeh, B., Kohli, A. and Coyle, M. (1993). Learning transformation rules for semantic query optimisation: a data driven approach. *IEEE Transactions on Knowledge and Data Engineering*, **5**(6), pp.950-964.

Siegel, M.D. (1988). Automatic rule derivation for semantic query optimiser. *Proceedings of the International Conference on Expert Systems*, pp.371-385.

Smyth, C.S. (2001). Mining mobile trajectories. In: *Geographic data mining and knowledge discovery*, H. J. Miller and J Han (eds.), London: Taylor and Francis, pp. 337-361.

Spence, N. and Frost, M. (1995) Work travel responses to changing workplaces and changing residences. In: *Cities in Competition. Productive and sustainable cities for the 21st century*, Brotchie, J.; Batty, M.; Blakely, E.; Hall, P. and Newton, P. (eds.) Longman Australia Pty Ltd., Melbourne. pp. 359-381.

Stead, D. and Marshall, S. (2001). The relationships between urban form and travel patterns: An international review and evaluation. *European Journal of Transport and Infrastructure Research*, **1**(2), pp.113-141.

Valk, van der, A. (2002). The Dutch planning experience, *Landscape and Urban Planning* 58, pp.201–210.

Vreeker, De Groot, H.L.F., Verhoef, E.T. (2004). Urban Multifunctional Land Use: Theoretical and Empirical Insights on Economies of Scale, Scope and Diversity, Built Environment, **20**(4), pp.289-307.

Wachowicz, M. (2001). GeoInsight: an approach for developing a knowledge construction process based on the integration of GVis and KDD methods. In: *Geographic data mining and knowledge discovery*, H. J. Miller and J Han (eds.), London: Taylor and Francis, pp. 239-259.

Wallace, C.S. (1990). Classification by minimum-message-length encoding. *Proceedings of the International Conference on Advances in Computing and Information* (ICCI’90), Springer Verlag, **468**, pp.72-81.

Wang, W., Yang, J. and Muntz, R. (1997). STINGA: Statistical information grid approach to spatial data mining. *Proceedings of VLDB*, pp.186-196.

Woerkum, C. v. (2000). *Communicatie en interactieve beleidsvorming*, Sam-som, Alphen aan den Rijn.

APPENDIX 1 - THE CALL FOR PARTICIPATION, WORKSHOP ON SUSTAINABLE MOBILITY

CALL FOR PARTICIPATION

WORKSHOP ON KNOWLEDGE DISCOVERY FOR SUSTAINABLE MOBILITY

The challenges and repercussions of discovering behavioral patterns from people in motion

Waag Society, Amsterdam, The Netherlands

September 11th and 12th , 2006

Introduction

We know very little about how human behaviour can affect sustainability, the environment, and the scenic quality, as we see in processes such as urban sprawl, intensive outdoor recreation, city expansion, and infrastructure developments. The main reason is the lack of monitoring information about the behavioural patterns of people's movements on the landscape. In principle, the Information and Communication Technology (ICT) made possible to gather very large data sets containing location information from mobile devices over time.

This workshop aims at getting a community input and feedback on the issues and repercussions of discovering behavioural patterns from people in motion using privacy preserving data mining, spatio-temporal ontologies, and visualisation techniques that will enable innovative movement-aware applications. Specifically, the workshop will provide a forum for discussion based on different perspectives from data providers, scientists, application developers, and users.

Scope of the Workshop

Topics considered in scope for this workshop include, but are not limited to:

Requirements

- Experiences from mobile telecommunication companies in providing the location of mobile customers
- Experiments on tracking people using mobile devices
- How is the mobile user experience different from the fixed user experience?

Strategic Issues

- How can we enable tracking people to be made as seamless, uncomplicated and reliable experience on mobile devices?
- How urban form characteristics (e.g. urban shapes and transportation network types at different scales) may be expected to influence people's movements on the landscape?
- How trajectory patterns can be used to test implicit and explicit claims underlying mobility and policies?
- How to unravel trust and confidence problems, as well as security and privacy rights?

Technology Issues

- How can we make better use of existing mobile technologies? Are there gaps? Is there a need for new technologies?
- What technologies are needed to enable uniquely mobile capabilities that are aware of people's behaviour on the landscape?

How to participate

To ensure productive discussions, the workshop is limited to 30 attendees.

Position papers are required in order to participate in this workshop. Each organisation or individual wishing to participate must submit a position paper explaining their interest in the workshop no later than August 27th . The intent is to make sure that participants have an active interest in the area, and that the workshop will benefit from their presence.

Send papers PDF format (1 to 2 pages) to Monica.Wachowicz@wur.nl

All position papers will be available from the workshop Web site. Speaker slides will also be available at the Web site after the workshop. There will not be printed proceedings.

To attend, you must register by filling out the registration form. The URL for the registration form will be sent to you after your position paper is accepted.

There will be no participation fee and participants are responsible for organizing their own travel and accommodation, however the availability of lower-cost accommodation will be available.

Important Dates

- **August 27th** : Deadline for position papers
- **September 1st** : Notification of acceptance of position papers
- **Sept 5th** : Registration deadline
- **11-12 September**: WORKSHOP ON KNOWLEDGE DISCOVERY FOR SUSTAINABLE MOBILITY

Workshop supported by

Project Name: Geographic Privacy-aware Knowledge Discovery and Delivery
Project Acronym: GeoPKDD
URL: http://www.geopkdd.eu
Project Number:IST-6FP-014915

Project Name: Mensen in beweging: Het plannen van mobiliteit in het landschap
Project Acronym: People in Motion
Project Number: RGI -160

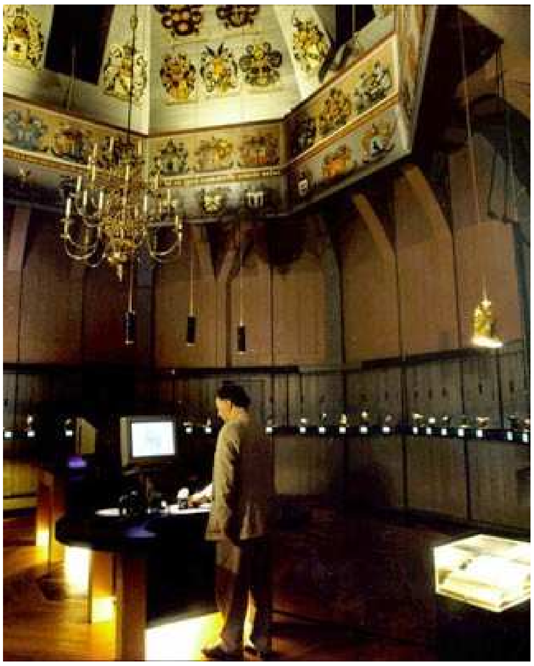
BSIK - Ruimte voor Geo-Informatie

Networks of Centers of Excellence

Location

The workshop will be held at the 17th century Theatrum Anatomicum located in the Waag Society building (http://www.waag.org/) in the heart of Amsterdam, the Netherlands.

De Waag, one of Amsterdams oldest monuments



The main conference room, the Theatrum Anatomicum, here Rembrandt painted his 'Anatomical Lesson'.

Please, do not hesitate to contact me if you have any questions.

Monica Wachowicz

APPENDIX 2 – LIST OF PARTICIPANTS

Name	Organization	Country
Arend Ligtenberg	Wageningen UR, Centre for Geo-Information	Netherlands
Aske Hopman	Waag Society, Locative Media	Netherlands
Ayse Goker	Ambiesense	United Kingdom
Bart Kuijpers	Hasselt University	Belgium
Carlo Ratti	MIT, SENSEable City Laboratory	USA
Chiara Renso	KDDLAB	Italy
Daniela Faur	Politechnica University of Bucharest	Romania
Dino Pedreschi	University of Pisa	Italy
Eduardo Dias	GEODAN, Mobile Solutions	Netherlands
Fosca Giannotti	KDDLAB	Italy
Gennady Andrienko	FAIS	Germany
Gideon Bazen	Logica CMG	Netherlands
Hans Myrhaug	Ambiesense	United Kingdom
Henk Eertink	Telematica Instituut, INCA Group	Netherlands
Jandirk Bulens	Wageningen UR, Centre for Geo-Information	Netherlands
Joanne Heyink Leestemaker	CityWorks	Netherlands
Jon Reades	University College London	United Kingdom
Marius Thériault	University of Laval	Canada
Monica Wachowicz	Wageningen UR, Centre for Geo-Information	Netherlands
Naser El-Sheimy	University of Calgary	Canada
Natalia Andrienko	FAIS	Germany
Ouri Wolfson	University of Illinois at Chicago	USA
Riccardo Mazza	WIND Telecomunicazioni SpA	Italy
Rob Postema	Logica CMG	Netherlands
Ronald Lenz	Waag Society, Locative Media	Netherlands
Salvo Rinzivillo	University of Pisa	Italy
Selim Volkan Kaya	Sabancı University	Turkey
Simone Puntoni	KDDLAB	Italy
Vania Bogorny	Hasselt University	Belgium
Wendy Bohte	OTB - Tu Delft	Netherlands
Wilko Quak	OTB - Tu Delft	Netherlands
Yücel Saygin	Sabancı University	Turkey