Mobility, Data Mining and Privacy: The GeoPKDD Paradigm

Fosca Giannotti (1, 2) and Roberto Trasarti (1)
(1) KDD Lab - ISTI - CNR Pisa, Italy
(2) Center for Complex Network Research, Northeastern University, Boston, MA

Abstract

The technologies of mobile communications and ubiquitous computing pervade our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. Miniaturization, wearability, pervasiveness of mobile devices are producing traces of our mobile activity, with increasing positioning accuracy and semantic richness: location data from mobile phones (Global System for Mobile Communications: GSM cell positions), Geographic Positioning System (GPS) tracks from mobile devices receiving geo-positions from satellites, etc. The objective of the GeoPKDD (Geographic Privacy-aware Knowledge Discovery and Delivery), a project funded by European Commission under the Future and emerging technologies (FET) program of the 6th Framework(FP6), has been to discover useful knowledge about human movement behavior from mobility data, while preserving the privacy of the people under observation. Pursuing this ambitious objective, the GeoPKDD project has started a new exciting multidisciplinary research area, at the crossroads of mobility, data mining, and privacy. This paper gives a short overview of the envisaged research challenges and the project achievements.

1 Introduction

Research on moving-object data analysis has been recently fostered by the widespread diffusion of new techniques and systems for monitoring, collecting and storing location aware data, generated by a wealth of technological infrastructures, such as GPS positioning and mobile phone networks. These have made available massive repositories of spatio-temporal data recording human mobile activities, that call for suitable analytical methods, capable of enabling the development of innovative, location-aware applications. This is a scenario of great opportunities and risks: on one side, mining this data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems; on the other side, individual privacy is at risk, as the mobility data may reveal, if disclosed, sensitive personal information. The GeoPKDD project [1], since 2005, investigates how to discover useful knowledge about human movement behavior from mobility data, while preserving the privacy of the people under observation. GeoPKDD aims at improving decision-making in many mobility-related tasks, especially in metropolitan areas:

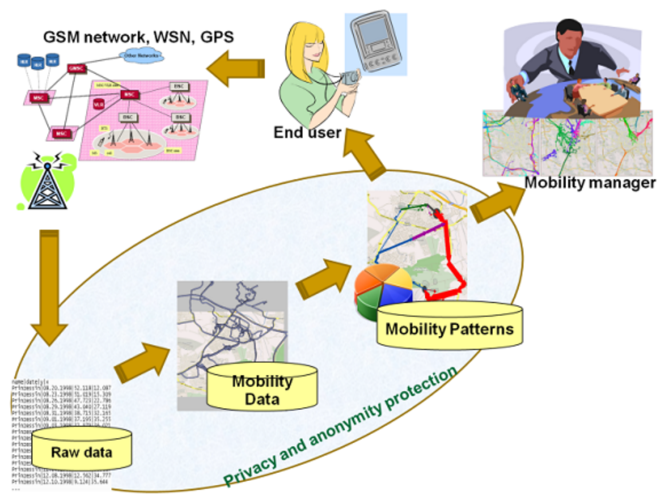

Figure 1: The GeoPKDD process

- Monitoring and planning traffic and public transportation systems
- Localizing new facilities and public services
- Forecasting/simulating traffic-related phenomena
- Geo-marketing and location-based advertising
- Innovative info-mobility services
- Detecting deviations in collective movement behavior.

Many challenging questions emerge from the above scenario. How to reconstruct a trajectory from raw logs, how to store and query trajectory data? Which spatio-temporal pattern and models are useful analytical abstractions of mobility data? How to compute such patterns and models efficiently? How to classify trajectories according to means of transportation (pedestrian, private vehicle, public transportation vehicle)? Privacy protection and anonymity - how to

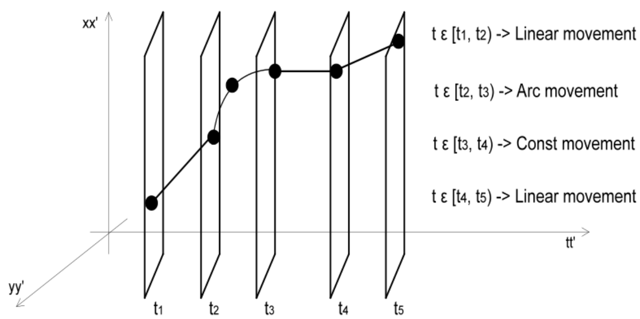

Figure 2: Representing trajectories in Hermes

make such concepts formally precise and measurable? How to find an optimal trade-off between privacy protection and quality of the analysis?

To answer these questions, the basic assumption of GeoPKDD is that *the movement data are at the base of an integrated knowledge discovery process, capable of supporting management, querying, analysis and interpretation of this form of data and patterns [13].* Unfortunately, discovering useful knowledge from mobility data cannot be achieved by simply invoking some off-the-shelf automated tool: as data miners know, successful analytics is the fruit of an overall knowledge discovery process, from raw data to knowledge. Figure 1 depicts the steps of the knowledge discovery process on movement data. Here, raw positioning data are collected from mobile devices and stored in the data repository. Trajectory data are then built, stored and analyzed by data mining algorithms to discover patterns hidden in the data. This process is typically iterative, since the composition of subsequent data mining methods is needed, both on data and pattern themselves, to obtain useful results. Finally, the extracted patterns have to be interpreted in order to be deployed by the final users. During its four years of activity, the researchers within the GeoPKDD project provided a wealth of methods and technologies aimed at support the various steps of such process. Some of these methods are embedded into analytical platforms, which provide different interaction metaphors. The next sections of this paper describe some of the GeoPKDD methods and platforms. Section 2 shortly describes the methods for storing and warehousing trajectories. Section 3 introduces the key trajectory mining algorithms developed in the project. Section 4 presents two platforms which integrate several analysis methods and provide two different interaction metaphors.

2 Mobility Data Management and Warehousing

A trajectory, the basic form of mobility data, is a sequence of time-stamped locations, sampled from the itinerary of a moving object. A database management system and a data warehouse have been designed around this specific form of movement data.

The design of the trajectory database has been influenced by the research on Moving Object Databases (MOD), which extends the traditional database technology for modeling, indexing and querying trajectory data. In MODs, the spatial and temporal dimensions are first-class citizens and both past and current (as well as anticipated future) positions of moving objects are of interest [4, 5, 6]. Among the available possibilities, GeoPKDD adopted the MOD Hermes [7], which, beyond storing and querying mechanisms for massive trajectory data, also provides efficient means for reconstructing trajectories from raw location data. Trajectory reconstruction transforms sequences of raw sample points into meaningful trajectories in accordance with different filters: temporal gaps, spatial gaps, maximum speed, tolerance distance, among others.

Hermes models moving object assigning to each time interval an arc, i.e., a function returning the position of the object for each instant within the interval; arcs are either lines or curves (See Figure 2.). Hermes provides dedicated indexing mechanisms that support efficient queries on trajectory data, such as:

- Spatial (range or nearest-neighbour) search : *Find all trajectories that were inside area A at time instant t (or time interval I)* or *Find the trajectory that was closest to point B at time instant t (or time interval I).*
- Topological / directional search: *Find all trajectories that entered (crossed, left, bypassed, etc.) or were located west (south, etc.) of an area* or *Find all trajectories that crossed (met, etc.) or were located left of (right of, in front of, etc.) a given trajectory T*
- Most-similar-trajectories: *Find the k most similar trajectories to a given trajectory T*.

Hermes is implemented on the top of a relational object oriented database management system, the ORACLE Spatial Cartridge.

Within GeoPKDD, the trajectory data warehouse, T-Warehouse in short, has been created on top of Hermes as a first analytical tool. The T-Warehouse is a spatio-temporal data cube representing various aggregated measures of the moving objects, such as spatial density and speed. The T-Warehouse supports a variety of dimensions (temporal, spatial, thematic) and measures (about space, time end their derivatives), which enable exploratory analysis, drilling up and down over space and time dimensions. The most challenging investigations addressed the definition of adequate aggregate functions for trajectories: a new method to compute holistic functions has been developed, such as the aggregate presence measure (number of distinct trajectories in a spatial unit) [8, 9, 10].

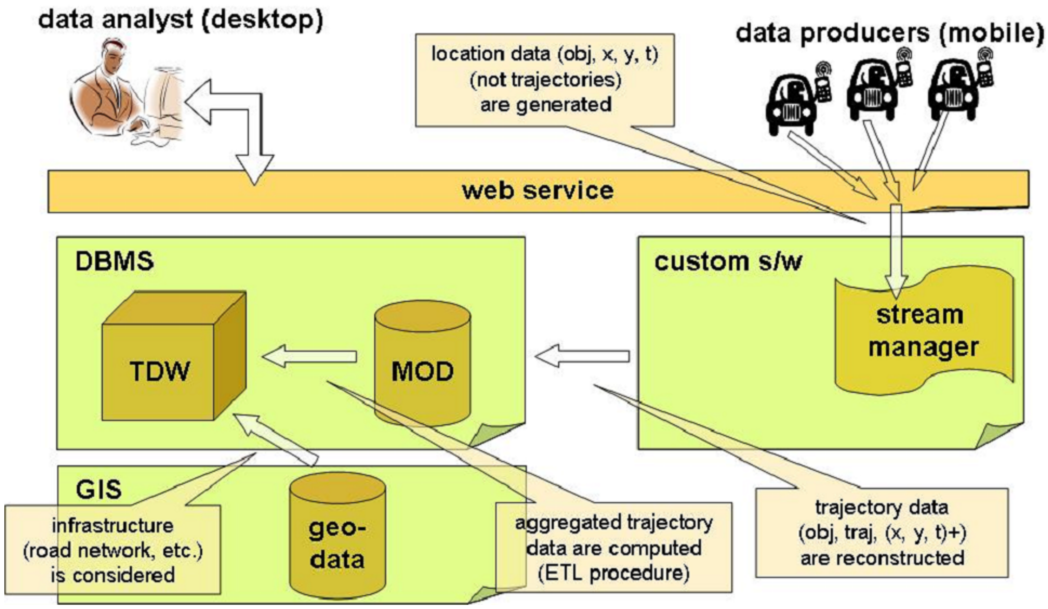

Figure 3: The trajectory database management system and warehouse

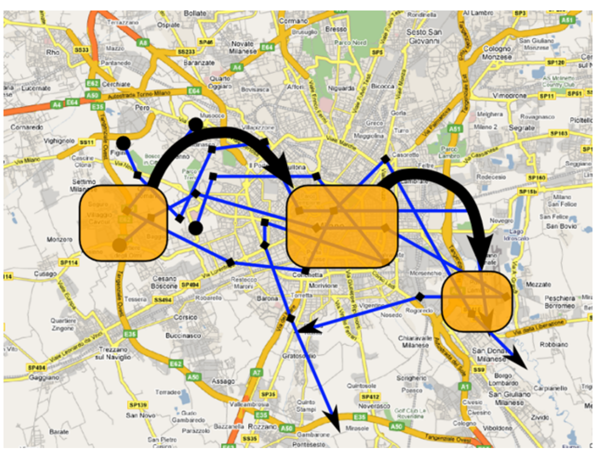

Figure 4: An Example of Trajectory Pattern

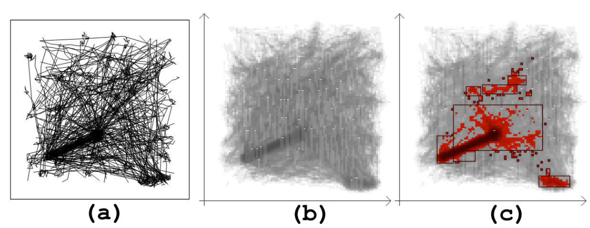

Figure 5: Roi construction: (a) original dataset, (b) dense cells and (c) regions of interest.

3 Mobility Data Mining

While the T-Warehouse analytical tools concentrate on *presence* of moving objects, mobility mining is aimed at analyzing *movement*. A method for mobility data mining tackles two different tasks: first, to define the format of spatio-temporal patterns and models to be extracted from trajectory data, and second, to design and implement efficient algorithms for extracting such patterns and models. This section provides a brief account of the different mining tasks developed within GeoPKDD focusing on trajectory patterns, trajectory clustering, and trajectory classification.

3.1 Trajectory Pattern Mining.

We introduce a novel notion of spatio-temporal pattern, which formalizes the idea of aggregated movement behaviors. A trajectory pattern, as defined in [12], represents a set of individual trajectories that share the property of visiting the same sequence of places with similar travel times. Therefore, two notions are central: (i) the regions of interest in the given space, and (ii) the typical travel time of moving objects from region to region. In this approach a trajectory pattern is a sequence of spatial regions that, on the basis of the source trajectory data, emerge as frequently visited in the order specified by the sequence; in addition, the transition between two consecutive regions in such a sequence is annotated with a typical travel time that, again, emerges from the input trajectories. For instance, consider the following two trajectory patterns over regions of interest in the center of a town:

$$\text{Railway Station} \overset{(15min)}{\rightarrow} \text{Castle Square} \overset{(2h15min)}{\rightarrow} \text{Museum (a)}$$
$$\text{Railway Station} \overset{(10min)}{\rightarrow} \text{Middle Bridge} \overset{(10min)}{\rightarrow} \text{Campus (b)}$$

Here, pattern (a) may be interpreted as a typical behavior of tourists that rapidly reach a major attraction from the railway station and spend there about two hours before getting to the adjacent museum. Pattern (b), may highlight the pedestrian flow of students that reach the university campus from the station: for them, the central bridge over the river is a compulsory passage. It should be observed that a trajectory pattern does not specify any particular route among two consecutive regions, while a typical travel time is specified, which approximates the (similar) travel time of each individual trajectory represented by the pattern. More formally:

DEFINITION 3.1. *A T-pattern is a pair (S,A), where*
$$S =< R_0, ..., R_n >$$
is a sequence of locations, and
$$A = \alpha_1, ...\alpha_n \in {R_k}^+$$
are the transition times (annotations). A T-pattern is also represented as:
$$R_0 \overset{\alpha_1}{\rightarrow} R_1 \overset{\alpha_2}{\rightarrow} \ldots \overset{\alpha_n}{\rightarrow} R_n.$$

The basic trajectory patterns discovery process consists of four steps:

Popular regions detection: The input set of trajectories $T = \{t_1 \ldots t_n\}$ is intersected with a $N \times N$ spatial grid and the number of trajectories traversing each cell is counted. If the count of a cell is greater than a frequency threshold τ the cell is considered *dense* (Fig.5(b)).

Region of interest construction: To obtain a larger region, neighboring dense cells are merged together to obtain larger regions, called Region of interest (Roi) (Fig.5(c)).

Trajectories translation: Each trajectory in the input dataset is transformed into a corresponding sequence of Roi with appropriate time-stamps, according to an approximated matching based on spatio-temporal interpolation.

Pattern discovery: Applying the algorithm introduced in [2] to the set of time-stamped sequences of Roi, the frequent sequences are extracted together with their typical transition times, with reference to a specified temporal tolerance threshold.

The complexity of the algorithm resides with the dynamic discretization of space and time needed to match fragments of different trajectories. When the Roi's are known

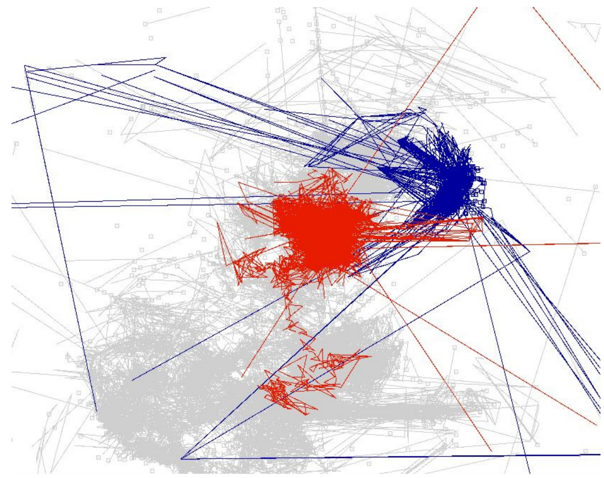

Figure 6: Trajectory clustering using the common-destination distance

a-priori as part of the background knowledge of the analysis, the algorithm simplifies considerably, as the first two phases above are not needed; this is the case of mobile phone data when the area of influence of towers (antennas) are considered as Roi's.

3.2 Trajectory Clustering.

Clustering is one of the general approaches to explore and analyze large amounts of data, since it allows the analyst to consider groups of objects rather than individual objects, which are too many. Clustering associates objects in groups (clusters) such that the objects in each group share some properties that do not hold (or hold much less) for the other objects. Spatial clustering builds clusters from objects being spatially close and/or having similar spatial properties (shapes, spatial relationships among components, etc.). Clustering of trajectories implies considering space, time and movement characteristics within a similarity notion: simple distance-based clustering methods are not effective in separating trajectory clusters that exhibit a non convex (non globular) shape, as it often occurs in practice. Therefore, we decided to design trajectory clustering on the basis of density-based clustering [23], extended according to the following principles:

Parametric w.r.t. to distance function : There are several interesting notions of distance useful to characterize diverse movement behaviors. [14] introduced the idea of a generic density-based clustering algorithm which is **parametric** with respect to a distance function. The distance functions may take into account the spatial locations visited, their order, and, in some cases, the time.

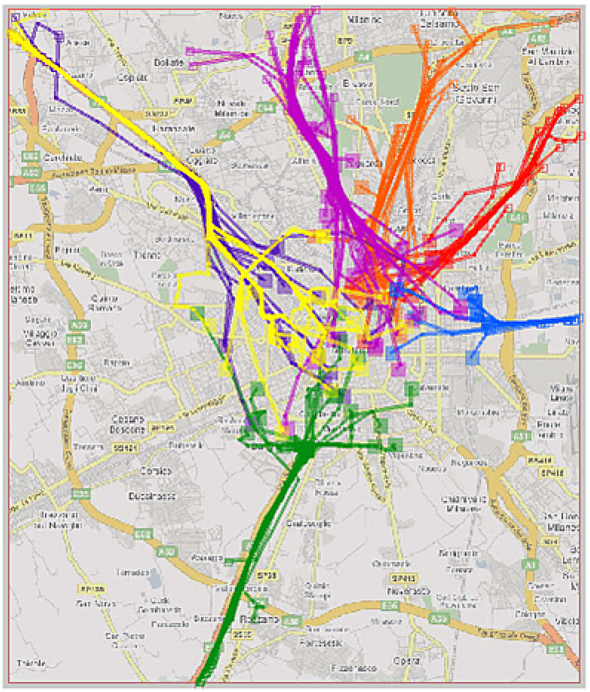

Figure 7: Trajectory clustering using the route-similarity distance

Progressive clustering : Trajectories are complex objects, so the clustering computation hardly scales up and may become prohibitively expensive. In [15] we introduced the idea of **progressive** clustering, using simpler distances at the beginning in order to separate the trajectories into broad clusters, and then focus on interesting sub-clusters using more complex and discriminative distance functions.

A variety of distance functions have been proposed for trajectories, including the basic Euclidean distance (assuming that trajectories are represented by vectors of fixed length), spatial Euclidean distance average along the time, time series-inspired functions such as (dynamic) time warping distance and Least Common Sub-Sequence (LCSS) measure, and direction-oriented distances. Some simpler distances that are often useful in practice are:

Common destination: Compare the destination of two trajectories and compute the distance between the two points. This simple distance allow the clustering algorithm to discover groups of trajectories ending in the same area (See Figure 6).

Common origin: Similarly to the previous one, this function considers the starting point of the two trajectories allowing the clustering to discover groups with a common start area.

Route similarity: The function finds the best alignment in space and time of the two trajectories for comparison (See 7).

In density-based trajectory clustering, a cluster of moving objects contains all elements that are density-reachable with respect to a density threshold. The algorithm presented in [14] computes an augmented cluster-ordering of the database objects. This algorithm is the basic brick of the interactive tool defined in [15], which allows the user to progressively refine the search. This method uses a sample of the original set of trajectories to compute the clustering and to reduce the time and space complexity needed. Then a subset of representative trajectories are extracted from each clusters. At this point the algorithm computes only the distances between the original set of trajectories and the representative ones to create a complete clustering. The result approximates the output obtained applying the classical algorithm on the entire set, but the experiments show that the error is very small and can be estimated with precision. The tool supports a step-wise analytical procedure called *progressive clustering*: a simple distance function with a clear meaning is applied at each step, while successive application of different functions yields sophisticated interpretation of clusters. Visualization and interaction techniques play a crucial role.

In [17] the notion of progressive clustering is further extended by combining clustering and classification, which are driven by a human analyst through an interactive visual interface. First, the analyst takes a manageable subset of the objects and applies a density based clustering to it. The analyst experiments with the clustering parameters for gaining meaningful results with respect to the analysis goals. Then, the analyst builds a classifier, which can be used for attaching new objects to the existing clusters. The analyst may also modify the clusters for their better understandability and/or conformance to the goals. The produced classifier is applied to the whole dataset. Each object is either attached to one of the clusters or remains unclassified, if it does not fit in any cluster. When necessary, the analyst may repeat the procedure (take a subset - cluster - build a classifier - classify) to the unclassified objects.

3.3 Trajectory Classification and Location Prediction.

Predictive models for trajectory data include a classification method for inferring the category of a trajectory, (e.g., the transportation means associated to a trajectory: private car, public transportation, pedestrian, etc.), and a predictor of the next location of a moving object given its past trajectory. There is strong current interest in next location prediction, in that it enables several intelligent location-based services. In the literature, this task is achieved by applying various learning methods to the history of each moving object for the purpose of creating an individual location predictor. Our

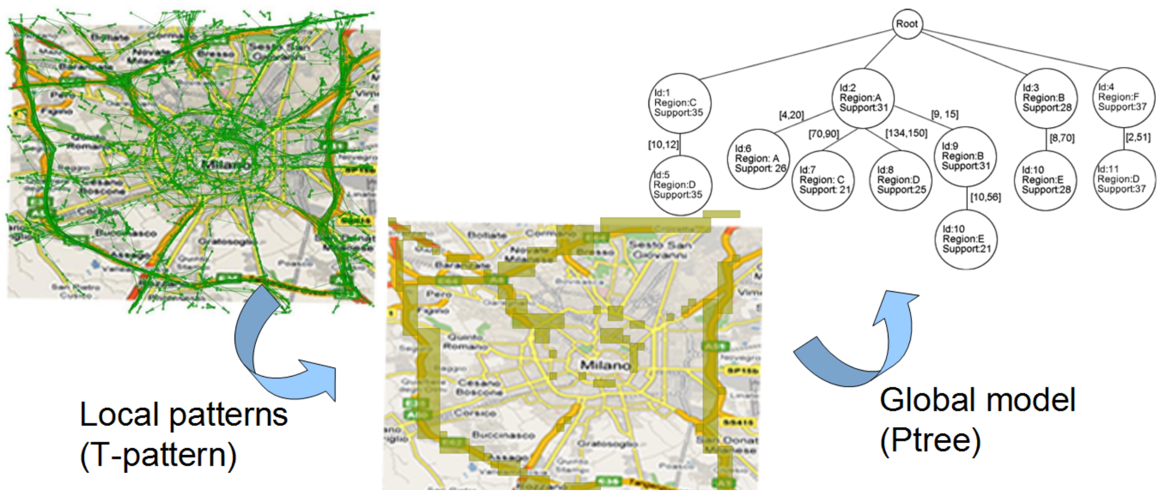

Figure 8: From local to global models: the prediction tree

proposed method, called WhereNext [19] predicts the future location of a moving object on the basis of the collective behavior, synthesized by the previously extracted T-patterns: this is coherent with the idea that global models can be built out of a collection of (many) local patterns. Using trajectory patterns as predictive rules has the following implications:

- the learning depends on the movement of all available objects in a certain area, instead of on the individual history of an object;
- the collection of trajectory patterns intrinsically contains the spatio-temporal properties emerging from the data.

The data structure used to build the predictor is a *prediction tree* constructed by merging the trajectory patterns. In the prediction tree each node contains entries of the form $\langle id, region, support, children \rangle$, where:

- id is the identifier of the node.
- $region$ represents a region of a T-pattern
- $support$ is the maximum support of the T-patterns of the node is part.
- $children$ is the list of child nodes.

The prediction tree uses the notion of T-Pattern prefix, defined as follows.

DEFINITION 3.2. *Let (S, A) and (S', A') be two T-patterns such that $(S, A) = R_0 \xrightarrow{\alpha_1} R_1 \xrightarrow{\alpha_2} \ldots \xrightarrow{\alpha_n} R_n$ and $(S', A') = R_0 \xrightarrow{\beta_1} R_1 \xrightarrow{\beta_2} \ldots \xrightarrow{\beta_k} R_k$. (S', A') is a prefix of (S, A) if and only if $k \leq n$ and $\forall i = 1 \ldots k$: α_i is included in β_i.*

To insert a T-pattern T_p into a prediction tree, we look for the path in the tree matching the longest prefix of T_p. Then, we append to the identified path a branch corresponding to the rest of the elements of T_p. The overall prediction tree is obtained by inserting all input trajectories progressively. The prediction tree is then used to assign the most likely next locations for a given moving object. The main idea behind the prediction is to find the best path on the tree, namely the best T-pattern, that matches the given trajectory. Hence, for a given trajectory we compute the best matching score among all admissible paths for the trajectory into the prediction tree. The children of the last node in the best matching path are selected as next possible locations.

3.4 Trajectory Anonymity.

In the context of personal mobility data, privacy is a big concern: location data allow inferences which may help an attacker to discovery private information, such as individual habits and preferences. Hiding explicit identifiers and replacing them with pseudonyms is insufficient to guarantee anonymity, since location represents a property that may allow re-identification: for instance, characteristic locations such as home and work place can be easily uncovered with the use of visual analytics methods, given detailed personal trajectories. Therefore, in all cases when privacy concerns are relevant, the trajectory data cannot be disclosed without appropriate safeguards. Anonymization techniques are data transformations that aim at a double goal: decrease the probability of re-identification below an acceptable threshold, while at the same time maintaining the analytical utility of the data. One of the basic objectives of GeoPKDD was to create analytical methods that natively took into account the privacy requirements. Therefore, the researchers involved in the project studied many different methods for protecting individual privacy, which applied to different steps of the knowledge discovery process, and aimed at preserving data utility with reference to various mining tasks. In the following we give a brief account of two methods for trajectory anonymity. The very basic notion is $k-anonymity$ for trajectories: a $k-anonymous$ trajectory dataset is one where the itinerary of each person is indistinguishable from that of other $k-1$ persons – anonymity viewed as *hiding in the crowd*. A k-anonymity method transforms a dataset of trajectories into a new one where all trajectories are k-anonymous.

The Never Walk Alone method [11] proposes a novel concept of k-anonymity based on co-localization that exploits the inherent uncertainty of the moving objects whereabouts. The notion of $(k, \delta) - anonymity$ is proposed for moving objects databases, where δ represents location imprecision. This approach is based on trajectory clustering and spatial translation: first, groups of k similar trajectories are formed by clustering, and then random spatial perturbation is applied to each group 9. The resulting trajectory dataset is k-anonymous, and some basic analytical properties are preserved, such as spatial density.

A more recent proposal is in [18], where the anonymization of movement data is obtained by combining the notions of spatial generalization and k-anonymity. The main idea is to hide locations by means of generalization, specifically, replacing exact positions in the trajectories by approximate positions, i.e. points by centroids of areas. The main steps of the proposed methods are:

- constructing a suitable tessellation of the territory into areas depending of the input trajectory dataset;
- applying a spatial generalization of the original trajectories;
- transforming the dataset of generalized trajectories to ensure that it satisfies the notion of k-anonymity.

We conducted a thorough study on a real-life GPS trajectory dataset, and provided strong empirical evidence that the proposed anonymity techniques achieve the convicting goals of data utility and data privacy: in practice, the achieved anonymity protection is way stronger than the theoretical worst-case, while the quality of cluster analysis on the trajectory data is accurately preserved.

4 Mastering the GeoPKDD Process

In order to support the interactive, iterative, combined usage of the various tools for the purpose of discovering mobility knowledge, GeoPKDD developed two prototype platforms: a semantic-based query & reasoning systems, and a visual analytic environment.

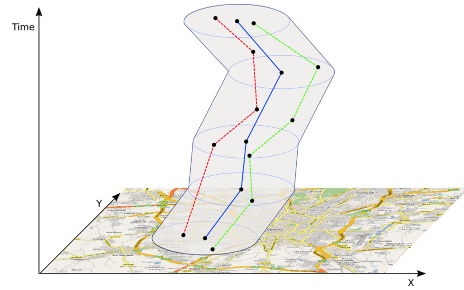

Figure 9: Trajectory anonymity: Never Walk Alone

This system allows the user to describe the entire knowledge discovery process using a set of primitives [20], based onto a **Data Mining Query Language** (DMQL) (See Figure 10). The spatio-temporal query primitives support selection and pre-processing of trajectory data w.r.t. geographic background knowledge, as well as anonymization. The trajectory mining primitives allow extracting and validating mobility patterns and models. The Data Mining Query Language managing the whole knowledge discovery process exhibits several advantages:

progressive querying and mining : Supporting the progressive combination of data selection, execution of data mining algorithms on the selected data, storing of the discovered patterns or models, querying the patterns or the models, selection of of data supporting such models and analyze it in more detail, e.g., by applying further mining tasks. An example is shown in Figure 11, where first a clustering task is applied, secondly the trajectories of a selected cluster are considered, and third the T-patterns over these trajectories are extracted. This iterative process allows to use the models not only as static knowledge to be presented as a result, but also as active elements of the process.

Repeatability of the process : The entire process is coded by a script that can be re-applied on different data.

The DMQL incorporates a reasoning component, which allows one to specify domain-driven ontologies using the Web Ontology Language (OWL); the typical use of this feature is to specify different types of trajectories and patterns [22]. The main objective of the semantic component is to enrich both trajectories and mined patterns with domain knowledge, thus making an explicit representation of semantic concepts. Examples of semantic enrichments for trajectories are the concepts of *stops* and *moves* [3], defined respectively as properties of movement and absence of movement in the segments of a trajectory. The semantic component of the DMQL allows one to specify such concepts and to perform the semantic tagging of trajectories, which can be used as input for mining tasks. Another example of semantic tagging in the transportation domain is the concept of systematic vs non systematic movement, which can be defined on the basis of the routine behavior of commuters in their home-work-home trips.

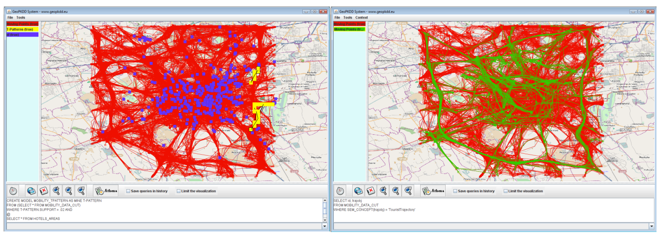

Figure 10: The Semantic-based query & reasoning system

4.1 Visual Analytics.

The aim of this system is to help the analyst to navigate through mobility data and patterns and to visually drive the analytical process [21]. The key features include: the visualization of T-patterns to support the navigation of the extracted patterns in the spatial and temporal dimensions; the progressive refinement of T-clusters, through user-driven exploration and evaluation of the discovered T-clusters [16] and the visual exploration of various measures provided by the T-Warehouse [9].

5 Conclusion

The analytical power of the tools developed within the GeoPKDD project has been to assess by experiments over massive collections of trajectory data, in particular data sensed by vehicular GPS devices at a fine spatio-temporal resolution. We ran a large scale experiment of urban mobility analysis, based on a real life GPS dataset, obtained from 17,000 private cars with on-board GPS receivers, tracked during one week of ordinary mobile activity in the metropolitan area of the city of Milan, Italy. The observed population consists of anonymous and heterogeneous car drivers participating in a specific car insurance program. On the basis of this experiment, we showed how a comprehensive atlas of urban mobility can be created, which reveals the relevant mobility behaviors: commuting trips, frequently followed itineraries, convergent patterns, slow-down patterns, etc. This concept goes beyond the O-D matrix, the typical tool of transportation engineering: not only the flux among locations is analyzed, but also the movement patterns obtained by learning from trajectory micro-data. The mobility patterns and clusters can be browsed by a mobility analyst (by the hours of the day, the days of the week, the geographic area, etc.), in order to explore the typical mobility of a city in

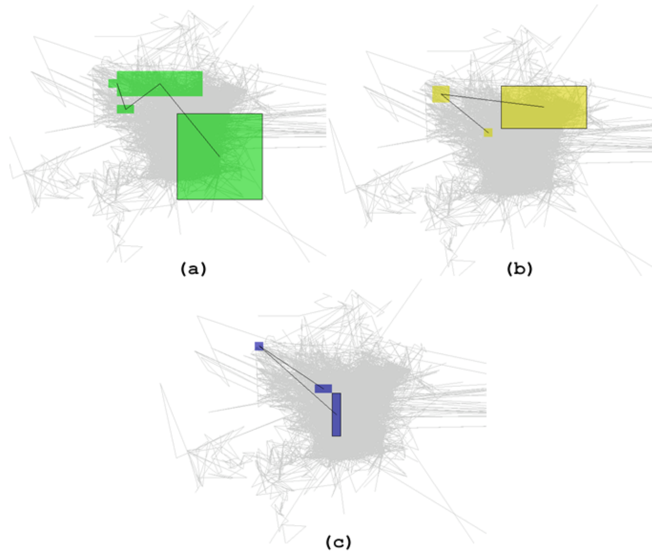

Figure 11: The T-Patterns discovered on the trajectories in a cluster

varying circumstances, also to observe emerging deviations from normal.

This complex experiment confirmed the original vision of the GeoPKDD project: in order to turn raw GPS tracks into useful forms of mobility knowledge and accomplish complex analytical tasks such as the creation of an urban mobility atlas, a thorough infrastructure for supporting the knowledge discovery process is needed, designed around a core of models and algorithms for trajectory data mining and analysis.

Acknowledgments: The authors wish to acknowledge Octotelematics S.p.A. for providing the anonymized dataset of private cars with on-board GPS receivers, tracked during one week of ordinary mobile activity in the metropolitan area of the city of Milan, Italy.

References

[1] Geographic Privacy-aware Knowledge Discovery and Delivery Project. GeoPKDD, http://www.GeoPKDD.eu/

[2] F. Giannotti, M. Nanni, and D. Pedreschi. Efficient mining of sequences with temporal annotations. In Proc. SIAM Conference on Data Mining, .SIAM, 2006.

[3] S. Spaccapietra et al. A conceptual view on trajectories. Data Knowl. Eng., 65(1):126-146, 2008.

[4] R.H. Guting, , et. al. SECONDO: An Extensible DBMS Platform for Research Prototyping and Teaching. In Proceeding of the International Conference on Data Engineering, ICDE, pages 1115-1116, Tokyo, Japan, April 2005.

[5] M.F. Mokbel, et al. PLACE: A Query Processor for Handling Real-time Spatio-temporal Data Streams (Demo). In Proceeding of the International Conference on Very Large Data Bases, VLDB, pages 1377-1380, Toronto, Canada, August 2004.

[6] O. Wolfson, et al. (2002) Management of Dynamic Location Information in DOMINO (Demo). In Proceeding of the International Conference on Extending Database Technology, EDBT, pages: 769-771

[7] N. Pelekis et al. (2006) Hermes - A Framework for Location-Based Data Management. Proceedings of EDBT. Pelekis, N. et al. (2008) HERMES: aggregative LBS via a trajectory DB engine. Proceedings of ACM SIGMOD.

[8] G. Marketos, et al. (2008) Building Real World Trajectory Warehouses. Proceedings of MobiDE.

[9] S. Orlando et al. (2007) Spatio-Temporal Aggregations in Trajectory Data Warehouses. Proceedings of DaWaK.

[10] N. Pelekis et al. (2008) Towards Trajectory Data Warehouses. Chapter in Mobility, Data Mining and Privacy: Geographic Knowledge Discovery. Springer-Verlag. 2008.

[11] O. Abul, F. Bonchi, and M. Nanni. Never Walk Alone: Uncertainty for anonymity in moving objects databases. In Proc. of the 24nd IEEE Int. Conf. on Data Engineering (ICDE'08)

[12] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory pattern mining. In KDD, pages 330-339, 2007.

[13] F. Giannotti and D. Pedreschi, editors. Mobility, Data Mining and Privacy - Geographic Knowledge Discovery. Springer, 2008.

[14] M. Nanni and D. Pedreschi. Time-focused clustering of trajectories of moving objects. J. Intell. Inf. Syst., 27(3):267-289, 2006.

[15] G.Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni and Dino Pedreschi. A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data. SSTD, pag. 432-435(2009)

[16] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Visually-driven analysis of movement data by progressive clustering. Information Visualization, 7((3/4)):225-239, 2008.

[17] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Interactive Visual Clustering of Large Collections of Trajectories. In proceedings IEEE VAST 2009.

[18] A. Monreale, G. Andrienko, N. Andrienko, F. Giannotti, D. Pedreschi , S. Rinzivillo, S. Wrobel. Movement Data Anonymity through Generalization. In PILBA 2009.

[19] A. Monreale, F. Pinelli, R. Trasarti and F. Giannotti: WhereNext: a Location Predictor on Trajectory Pattern Mining.KDD 2009. 15th ACM SIGKDD Conference on Knoledge Discovery and Data Mining

[20] R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, Y. Theodoridis: The DAEDALUS Framework: Progressive Querying and Mining of Movement Data. ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS 2008).

[21] G.L. Andrienko, N.V. Andrienko: A Visual Analytics Approach to Exploration of Large Amounts of Movement Data. VISUAL 2008: 1-4

[22] M. Baglioni, J.A. Fernandes de Macdo, C. Renso, R. Trasarti, M. Wachowicz: Towards Semantic Interpretation of Movement Behavior. AGILE Conf. 2009: 271-288

[23] M. Ankerst, M. Breunig, H.P. Kriegel and J. Sander Optics: Ordering points to identify the clustering structure. In Proc. ACM SIGMOD 1999, 4960.