Spatio-Temporal Data Warehouse Design for Human Activity Pattern Analysis

L. Savary, T. Wan, K. Zeitouni
PRISM Laboratory, Versailles University
45 Avenue des Etats-Unis
78035 Versailles - France
{Lionel.Savary, Tao.Wan, Karine.Zeitouni}@prism.uvsq.fr

Abstract

Many applications refer to moving objects or phenomena and require spatio-temporal modelling and specific analysis. Unlike conventional data where attributes are simple values (numerical or textual), object motion defines a continue variation in space and time. This makes it difficult to handle such datasets, for instance, answering the query "how many objects cross a given area during a given time period?" This article addresses spatio-temporal data warehouse schema modelling in order to facilitate those queries and more generally, on-line analytical processing (OLAP) on spatio-temporal databases. It is illustrated on a space-time human activity survey database within the framework of a European IST project[1].

1. Introduction

Given a collection of moving objects, how can the database identify the area and instant of greatest object concentration? For instance, a query like "how many objects cross a given area during a given time period?" These are nowadays major subjects in spatiotemporal databases research areas. However, the spatiotemporal dataset's collection, modeling and organization require important pretreatment before their querying and analysis. This is the case in the European HEARTS (Health Effects and Risks of Transport Systems) project. Indeed, depending on the population activities in time and space, it aims to analyze health risks associated with air pollution and noise and with injuries due to accidents, especially within vulnerable groups such as children and elderly people. It is mainly based on phenomena (pollution, noise, etc.) measures combined with human activity survey (HAS). However, the provided HAS database does not properly represent the trajectory with the people cinematic movements. Instead, only depart and destination zones and times are represented.

Moreover, spatio-temporal modeling is still a challenging field [5] where real dynamic objects in space and time are never implemented in existing GIS.

Thus, we firstly construct an approximation of human motion in space and time to describe population displacements. Due to performance reason and also to facilitate its managements, we make a representation based on discrete space and time.

Secondly, the fundamental characteristic of the data warehouse technology is its multidimensional paradigm. Furthermore, it provides on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data exploring. However, due to spatio-temporal data complexity, they can not be presented within a conventional multidimensional database. Hence, integrating spatio-temporal data in multidimensional paradigm is anther challenge of spatio-temporal data warehouse modeling. In this paper, we give a design of spatio-temporal data warehouse schema which can easily include spatio-temporal information and handle computing operations on it. This data warehouse is developed in the framework of the European HEARTS project and applied to a space-time HAS database.

This paper is organized as follows: the next section presents an overview of related works. Section 3 introduces our problematic and describes our original spatio-temporal data on human mobility. A spatio-temporal multi-dimensional schema is detailed in section 4. Finally the main advantages of our contribution and perspectives are given in conclusion.

2. Related works

[1] This publication was partly funded through HEARTS (Health Effects and Risk of Transport Systems), a research project co-funded by the Energy Environment and Sustainable Development Thematic Programme of the European Commission (contract number: QLK4-CT-2001-00492) and participating institutes (http://www.euro.who.int/hearts).

Many applications refer to moving objects or phenomena and require spatio-temporal modelling and specific analysis, but many commercial data warehouse and OLAP tools for multi-dimensional database analysis are restricted due to the allowable data types for dimensions and measures. Therefore, spatio-temporal databases research has attracted a great deal of attention of many researchers in recent years. We focus here only on works related to spatio-temporal data modelling for human activity, and those dealing with spatial and temporal databases and analysis.

2.1. Spatio-temporal model for human activity

In [8], authors propose a generic spatio-temporal data model to describe activity pattern in a GIS (Geography Information System) environment. These activity patterns are composed of a succession of static and dynamic type of activities carried out at different places and times. In figure 1, an individual (Person) may carry out a list of static and dynamic (Travel Between) activities, where the static activities are georeferenced (Location).

In [7], authors propose a spatio-temporal data model for describing events and episodes along an individual lifeline. They take into account only three lifelines (trajectories): household, residential and career. An individual's history is altered when an event modifies one of his lifelines.

In [3], authors propose a spatio-temporal conceptual model to describe the motion of individuals, in order to analyze their spatio-temporal trajectories and determine spatio-temporal regions where they meet.

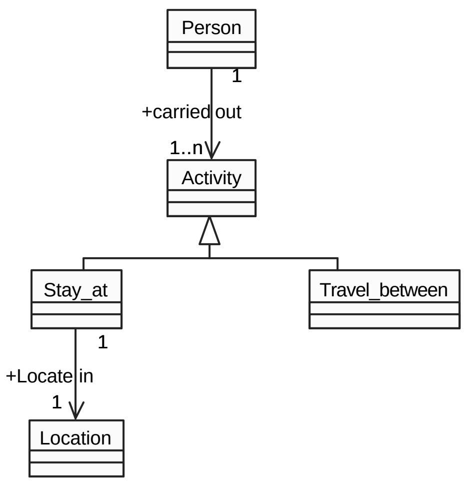

Figure 1. Schema of activity pattern

2.2. Spatial and temporal databases and analysis

Spatio-temporal modeling is still a challenging field [5]. Real dynamic objects in space and time are not yet implemented in commercial GIS. Yeh have proposed a model that maps spatio-temporal features to 3D shape [9]. This reduces spatio-temporal queries to geometrical functions on their 3D representation. Exploring data could be at the first stage based on user queries.

Beyond, exploratory analysis is typically supported by OLAP tools that perform on a multi-dimensional model for most data warehouses. Recent works have considered the extension of those tools and models to deal with spatial data [6]. It proposes to consider dimensions and measures on spatial entities. Further, it proposes different optimization of merging spatial descriptions (geometry) by selectively pre-computing aggregates.

3. Data description and problematic

At the beginning, the original data is a HAS database concerning Lille (a French town) urban mobility. It contained only five principle tables showed in Figure 2: Household, Person, Trip, Journey and Zone. A household contains person residence information and a person may every day do several trips. Each trip may contain several journeys in which more information are provided (e.g. traversed zones, beginning and ending time of each journey, and mode used).

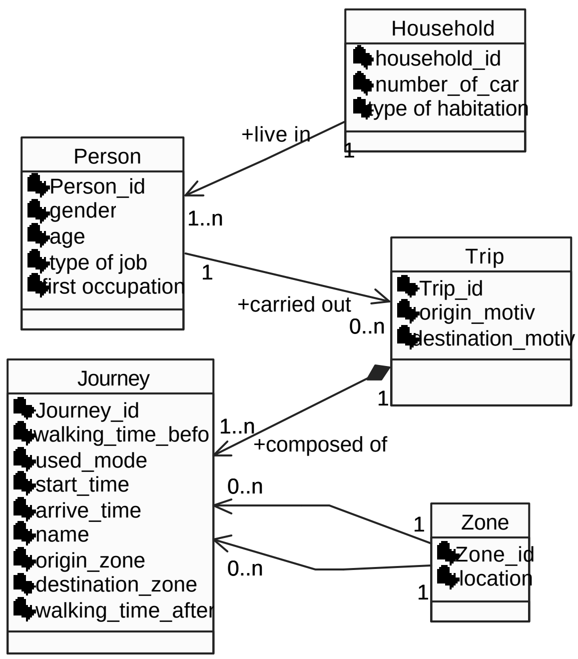

Figure 2. Schema of origin tables

More detailed and vivid for the above figure, figure 3 shows an example of a trip carried out by an individual between 8h00 and 8h45. This trip contains 2 journeys. In the first journey, this person leaves his home in zone1 and walks for 3 minutes to the car parking, then drives his car to tramway station and walks 5 minutes to the platform. For the second journey, he moves from zone2 to zone 3 by the tramway and uses 4 minutes to reach his work office.

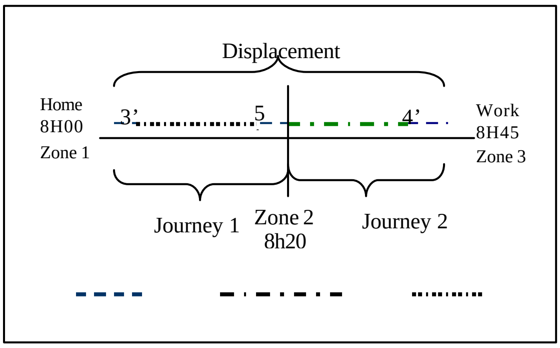

Figure 3. Schema of origin tables

In this HAS database, we have typically spatio-temporal information; however, it does not properly represent the trajectory with the people cinematic movements. Instead, only depart and destination zones and times are represented. Nevertheless, the objective of the project HEARTS is to explore the space-time activity survey and to produce statistics on flexible groups. Hence, the information in this HAS database is too indigent to attain our objective.

Moreover, whether a database is well modelled decides it's following corresponding exploring and analysis. Especially for a HAS database including spatio-temporal information which are difficult presented within a conventional multidimensional database.

Therefore, we propose to estimate an approximation of people cinematic movement trajectory, and then give a data warehouse schema which can facilitate spatio-temporal exploring. They are detailed in the following session.

4. A spatio-temporal representation

Right now, real dynamic objects in space and time have never been implemented in existing GIS. Besides, models such as [9] cannot be implemented with reasonable performances for a large data collection due to the complexity of 3D geometry modeling and their geometrical functions. We propose hereafter another approach.

4.1. Approximation of spatio-temporal data description

The provided MENAGE database does not properly represent the trajectory and the cinematic of people movement. Instead, only depart and destination zones and times are represented.

Firstly, we propose to estimate an approximation of this spatio-temporal information. In theory, this should be trajectory in space and time, i.e. a 3D curve. But for performance reason and also to facilitate its management, we make a representation based on discrete space and time (figure 4).

For space, we use the provided zoning, and for time intervals, we consider the minimal duration of zone crossing. Hence, each movement is defined by a set of pairs (#zone, #timeInterval). Based on this model, we estimate the crossed zones and time intervals within each zone for each journey. Compared with the schema in figure 2, the spatio-temporal entity defined in this schema and the associated discretized units facilitate OLAP operation like drill-down, roll-up process other space and time dimensions (see section 4.2). Note that more sophisticated estimations of trajectory could be implemented, but the target model stays unchanged by associating discrete locations and times. We will show afterwards how this representation is more adequate to Spatio-temporal OLAP (figure 5).

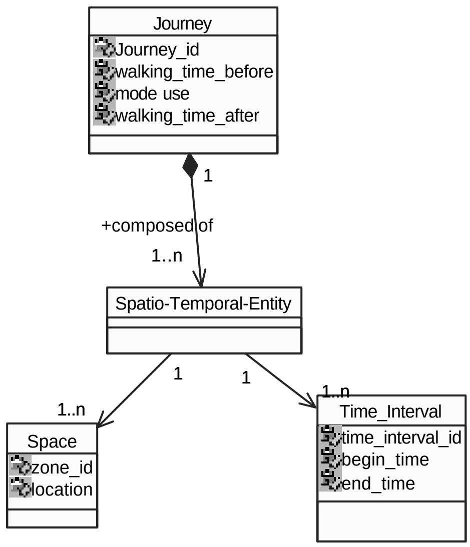

Figure 4. Discretized space and time

The figure 4 describes a part of the original schema (figure 1) where a Journey is composed of the entity *Spatio-Temporal-Entity* which stands for the smallest granularity of a spatio-temporal region crossed by an individual.

The figure 5 shows a spatio-temporal motion of two individuals in a three dimensional space. The three dimensions are composed of two spatial axes (X and Y) representing the map by several zones, and a time axe (T) where the time is discretized into units where the finest granularity is represented by a spatio-temporal region corresponding to one unit of time (the shortest time carried out by an individual to cross a zone) and the smallest zone crossed.

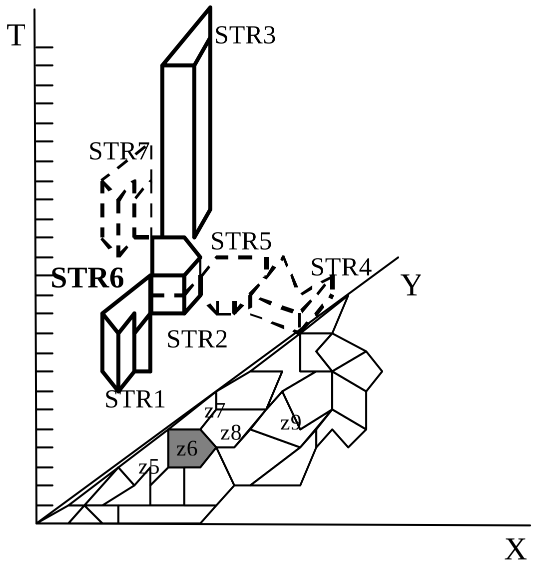

Figure 5. Spatio-temporal region of risks exposure

The first individual (I1) cross three spatio-temporal regions (STR1, STR2, STR3) to go from one place to another. Along his movements, he travels the zone 5 (z5) during five units of time (discretized time), then the zone 6 (z6) during two units of time and finally travels in zone 7 (z7) during ten units of time. The second individual (I2) cross four spatio-temporal regions (STR4, STR5, STR6, STR7). Along his movements, he cross the zones z9, z8, z6, z5 during respectively one unit, one unit, one unit and four units of time. In this figure, we can see that these two individuals are meeting in a spatio-temporal region = STR2 n STR6 = STR6. Thus in zone 6 (z6) during one unit time, I1 and I2 are exposed to the same risks.

4.2. Multi-dimensional schema for spatio-temporal data warehouse

This project aims to explore the space-time activity survey and to produce statistics on flexible groups. This can be carried out by data cubes that generate views with all combinations of dimensions. The specificity of such data is the spatial and temporal dimension. For instance, to estimate the number of persons exposed to a given pollution rate at a given time and location, the analysis will focus on spatial and temporal dimensions counting the number of persons. The difficulty is also involved by the spatial and temporal dimensions, due that the spatio-temporal data can not be presented within a conventional multidimensional database. The [6] is tailored only for spatial databases. However, with time dimension's acceding, objects are mobile not only in space but also in time. Furthermore, our data representation model simplifies the human cinematic movements' description by a set of values, conventional OLAP do not allow multi-valued dimensions. For example, a person trajectory in a journey is defined by a set of pairs (#zone, #timeInterval) while conventional data cube only handles single values. Therefore, we need to design a new database schema in order to adapt our application to data cube capabilities. The proposed solution is to split the journey description in several entities called ST-Journey and to use simple dimensions for zones and time Intervals. Thus, an example of the proposed new spatio-temporal database schema is showed in figure 6.

This figure specifies a snowflake schema of a data warehouse. The fact table is *ST-Journey* and represents human activity and mobility varying in space and time. Dimension tables are from one hand *Space* dimension and *Time* dimension, and from the other hand, the activity of a person. This activity is structured at different levels from the most detailed description (*Journey* dimension) to its generalisation by trip to the most general description (*Person* dimension). Each level contains specific attributes that may have a hierarchical description (such as *Mode*, *Activity* or *Profession*). The *Space* and *Time* dimensions also define a hierarchy with different levels of granularity. The fact table has *trip_duration* as its measure which signifies a duration when a person stays in a space unit at a time interval. Hence, if one wants to calculate the exposure indice in a risky zone, just an addition of all durations in this zone is needed. Moreover, if one wants to calculate the person number crossed in a given time and a given space, just the distinct count of all the records within the time and space interval is needed.

Those query examples show how the multi-dimensional schema proposed here adapts conventional OLAP capabilities to the analysis of the population's spatio-temporal motion as illustrated by the figure 5. The hierarchical views of space and time allow global analysis and mapping of general trend for mobility pattern. Moreover, it puts forward other dimensions than space and time. This brings all the capabilities of OLAP as efficient querying on any

combination of dimensions, drilling down and rolling up granularity levels, slicing and dicing the cubes. For instance, calculating trip duration by grouping each person's journeys, focusing in exposure by mode (such as pedestrian), by age (infants), by profession (students), etc.

This multi-dimensional spatio-temporal schema can be implemented in a data warehouse like Oracle or SQL Sever. In our case, it has been implemented under SQL Server. The estimation of trajectory described in section 41 has been implemented using Oracle that supports spatial data management. It has been tested on the Lille mobility survey database.

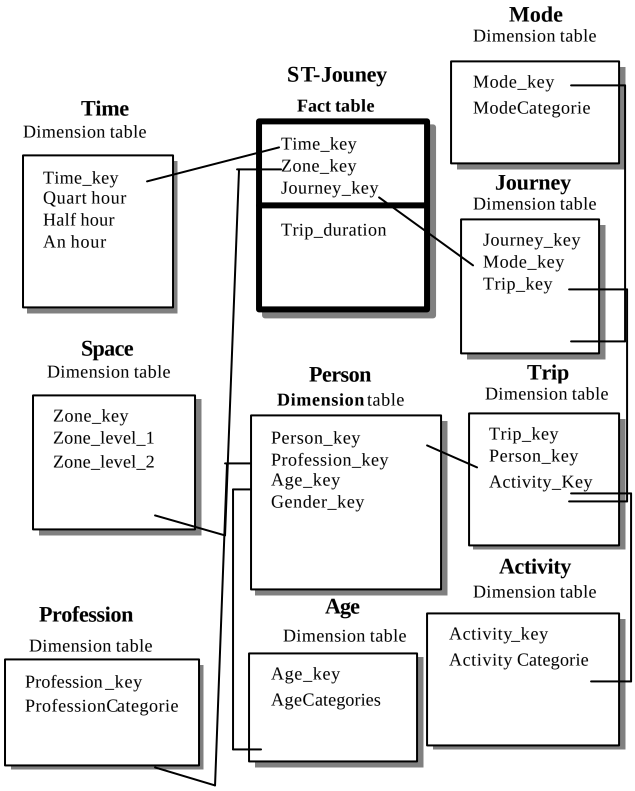

Figure 6. The proposed new spatio-temporal database schema

5. Conclusion

This paper proposed an approach to explore spatio-temporal database. We point out the difficulty to directly use native description of spatio-temporal data. Firstly, we have proposed an approximation of this description by discrete values, i.e. zone subdivision for space and fixed time intervals for time. Then, we have designed a multi-dimensional model including space and time dimensions, and sketched how this schema adapts to spatio-temporal data exploration.

This work has already been prototyped and is at the stage of test for its validation for HEARTS project. Ongoing works deal with its extension to allow computation of spatio-temporal aggregates (e.g. merging adjacent areas) and extend OLAP queries to spatial and temporal queries. Future works will concern the evaluation and the optimization of OLAP queries in this context.

References

[1] T. Carlstein, D. Parkes , N. Thrift, "Afterward: Towards a Time-Space Structured Approach to Society and Environment", Human Activity and Time Geography, London, 1978, Trith (eds.), vol. 2, p. 225-263.

[2] CERTU, *Ministère de l'Equipement, des Transports et du Logement : L'enquête ménages déplacements méthode standard*. Collections du Certu. Octobre 1998. (URL : http://www.certu.fr)

[3] M.C. Fauvet, S.Chardonnel, M.Dumas, P.C. Scholl, P.Dumolard, "Analyse de données géographiques : application des bases de données temporelles", Revue internationale de géomatique. Vol. 8 – n° 1-2/1998, p. 149-165.

[4] J. Gulliver, *Space-time modelling of exposure to air pollution using GIS*. Thesis submitted for the degree of Doctor of Philosophy, University of Leicester, 1995.

[5] D. Peuquet, *Representation of Space and Time*, New York : Guilford Press, 2002..

[6] N. Stefanovic. *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*, Thesis for the degree of Master of science. September 1997.

[7] M. Thériault, C. Claramunt, A. Séguin, P. Villeneuve, "Temporal GIS and Statistical Modelling of Personal Lifelines", 10th International Symposium on Advances in Spatial Data Handling, Ottawa, July 2002, Richardson D. and Van Oosteram P. (eds.), Springer-Verlag.

[8] D. Wang, T. Cheng, "A spatio-temporal data model for activity-based transport demand modelling", International Journal of Geographic Information Science, vol. 15, 2001, p. 561-585.

[9] T.S. Yeh, B. de Cambray, "Time as a geometric dimension for modeling the evolution of entities: a 3D approach", In *GIS and Environmental Modeling : Process and Research Issues*, GIS World Edition Books, Goodchild, Steyart, Parks et al. (eds.), 1996.