Semana de Engenharia 2010
Guimarães, 11 a 15 de Outubro

TAXONOMY FOR THE REPRESENTATION OF SPATIOTEMPORAL DATA

João Peixoto[1] and Adriano Moreira[2]
Mobile and Ubiquitous Systems Group
University of Minho, Guimarães, Portugal
E-mail: [1] peixoto@kanguru.pt, [2]adriano@dsi.uminho.pt

KEYWORDS

Urban Mobility, Spatiotemporal data representation.

ABSTRACT

Several visualization applications of urban mobility have been presented recently, either by the analysis of space occupation, or through the representation of trajectories. However these visualizations are highly dependent on the data type and often only applicable to the selected data type. This paper presents our initial work to create a taxonomy for the representation of spatiotemporal data, sufficiently flexible for various types of mobility data and different forms of visualization of the urban mobility.

INTRODUCTION

The mobility of citizens in an urban area is the source of various problems: traffic congestion, urban planning, among others. For this reason it is important to understand the behaviour of individuals in space and space itself, and that these make use of urban space as a way to reduce and possibly eliminate these difficulties.
However, the dynamics associated with the mobility in an urban area always has two components: Time and Space, rising new questions on how to represent and visualize these dynamics.
As referred by Yu and Shaw (Yu and Shaw 2004) the current Geographic Information Systems (GIS) are structured to represent the spatial component of data but lack the temporal component.
For this reason several studies, presented recently, investigate the spatiotemporal data visualization problem through different forms of visual representation. However, regardless of how the dynamics of an urban space is represented, most of these works focus only one type of mobility data, i.e. a particular situation of information. Although our research work is focused on visualization of the dynamics of urban space, our initial aim is to create a flexible and comprehensive taxonomy for the spatiotemporal representation of movement data that allows the same concepts to be applied to different types of data from different sensors such as GPS, Wi-Fi, GSM, ticketing systems, as well to the different scenarios of urban mobility: through its own car, public transportation, pedestrian. We believe that before explore the different visualization techniques we need a flexible data structure representation of mobility to facilitate the study and application of different visualization techniques.
So this article is organized as follows. The next section describes some of the work developed in the field of visualization of urban mobility. In Section III describes the concepts that are the source of our proposed taxonomy for the representation of spatial-temporal data. In Section IV we instantiate the concepts with two mobility data types in order to validate them. Finally (Section V) we present some conclusions and open questions related to our work.

RELATED WORK

Several works have been presented in the visualization of urban mobility area, using different techniques. One of these techniques is analyze urban mobility using the temporal variation of the occupation that individuals make in the urban space (iSPOTS 2010; Reades et al. 2007; Sevtsuk 2005). This technique is somewhat related to the types of data that can obtain an urban space. It isn't an easy task to get individual data from an urban space, since in some cases the available and collected data are aggregated. This type of representation is based on the creation of temporal snapshots of space occupation. However, due the dynamics of the urban space, this approach may not be the most indicated for the analysis of pattern changes (Hagen-Zanker and Timmermans 2008). Another problem is the definition of mobility in these approaches, because they represent the mobility through the variation of space occupation over the time and not the real movement of individuals. Analyze the mobility based on the use of space does not allow the extraction of more depth conclusions about the urban mobility. We know that there are changes in urban space but we not know the source and destination of movements, the most we can infer that they may have occurred between some areas of space. Thus, the application of this approach may be useful for the planning of urban space based on the detection of concentration areas of individuals, but not adequated for the detection of problems caused by mobility itself, such as traffic congestion.

We believe that deep understanding of the phenomena associated with urban mobility involves the visualization of trajectories, which truly reflect the movements of individuals.

In this area one of the ways to represent trajectories is with vectors, where each vector represents an element in space and time, speed and direction of its movement (Moreira et al. 2010). Through this approach it is possible to have some sense of mobility, since it allows the representation of an individual according to a spatiotemporal reference. However as this representation is based on the observation of the instantaneous movement, the simultaneous perception of the origin and destination of the movement is not easily transmitted. In order to address this issue, some works are using a different technique for representing paths through interconnected source-destination pairs (Brockmann and Theis 2008). Although with this approach it is possible to visualize the trajectories, several questions arise regarding the outcome of the visualization. First, if the interval between samples is large we lose intermediate movements and then our trajectory is twisted, as such there may be behaviours that are not represented. Second, to connect the source to destination we may have to affect the Time component, since the analysis is not done continuously, but by time intervals, consequently losing this representation the notion of space change over the time and creating the same issues associated with representation by snapshots.

Possibly the best approach to the visualization of urban mobility using the representation of trajectories is a compromise between the two techniques described above, take simultaneous the current state of movements, and past and future of trajectories.

In our research work we intend to study these and other issues relating to the visualization of mobility through the representation of trajectories so as to explore new paradigms of representation of mobility. Abandoning the snapshot representation of artefacts (individual or object) and create personal and global maps of mobility. These maps should include the most visited sites, the locations where the user spends more time and the movement axes.

To achieve this goal we believe that it is important to, first, properly structure the information in order to have a taxonomy for the representation of mobility in an urban area and verify what type of data match with this structure, if they already exist or must be acquired.

CONCEPTS OF DATA REPRESENTATION

The work that we have done so far defined four general concepts that characterizes our taxonomy for the representation of data mobility (Figure 1) relating to an artefact. These concepts are designed to fit the data since its acquisition, i.e., data obtained by different sensors without any processing, until we get the same data on the form of trajectories, be they individual or group of individuals.

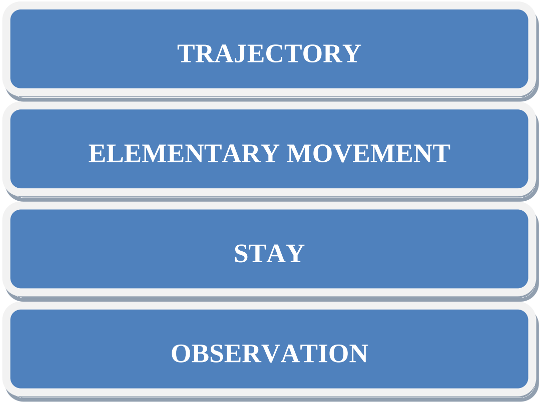

Figure 1: Layered Structure of the Information Concepts

Observation

Regardless the sensor used to acquire the mobility data, that data reflect the observation of a phenomenon and as such the first concept of our representation is the Observation, and can be described abstractly as:

The observation of an artefact in a specific spatiotemporal space

and represented by the following attributes:

<Id_Observation, Artefact, Timestamp, Place>

Stay

As in the work of Reades et al. (Reades et al. 2007) the artefacts observed may reflect the presence of such a space and time. To include this notion, the next layer of our taxonomy defines the Stay as:

Time interval between the first and last observation of an artefact in the same place

and is represented by the following attributes:

<Id_Stay, Artefact, Timestamp_Initial, Timestmap_Final, Place>

Elementary Movement

As the artefacts are characterized by the mobility, it is expected that there are changes in space over the time, so there must be a concept that represents this variation that we called Elementary Movement and described as:

Change of Place to the next one that occurs in time

represented by the following attributes:

<Id_Movement, Artefact, Place_Start, Place_End, Timestap_Initial, Timestap_Final>

Trajectory

Finally, at the top layer of our representation we have the concept of trajectory that represents a set of Elementary Movements ordered in time for the same artefact and is described as:

Time-ordered list of Elementary Movements of an artefact over the space

represented by the following attributes:

<Id_Trajectory, Artefact, Movements>

In turn, the set of existing trajectories at given spatiotemporal moment for a given artefact allows the representation of flows or dynamic movements that exist in the urban space, the final goal that we want achieve with our study.

INSTANTIATION OF CONCEPTS FOR DIFFERENT DATA

Wi-Fi Data

In order to normalize, match and validate the concepts of our taxonomy for the representation of spatiotemporal data that we have been developing, information about the use of the Wi-Fi network at the University of Minho was used for the first instantiation of the concepts, combined with information of the location of several existing Access Points (APs). The information from the logs of multiple APs that define the network was acquired at intervals of five minutes.
From the data collected and to represent them in the form of the concepts previously described in our taxonomy, we created a Relational Database (RDB) composed by a set of tables, where the main tables have a direct correspondence with the four concepts of taxonomy (Figure 2). So to instantiate the concept Observation we have created a table with the same structure of the concept and two auxiliary tables: Artefact and Place. The table of artefacts in this particular case only requires the following attributes:

<Id_Artefact, MacAddress>

being the Mac Address in this case is used to identify the user that logged in to Wi-Fi network. While the Place table contains the following attributes:

<Id_Place, IP, Latitude, Longitude>

where in this particular case the attribute IP is the IP address of the AP, which has an associated pair of coordinates (latitude and longitude).

These two relationship tables are independent of the taxonomy, since its attributes depend of the type of information gained.
When we making the adjustment and loading of data acquired, we have seen immediately a situation that in the future should be studied, the huge amount of information generated.
In this particular case and for a small urban space of just two *campus*, with data collected in only ten days, the table of observations exceeded the one million records and we are just in the first concept of the four existing in our taxonomy, although one that most records will have.

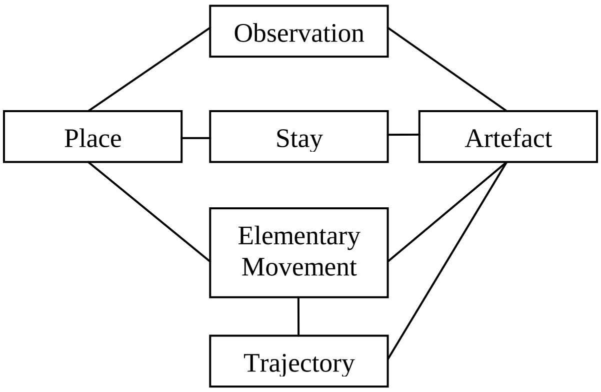

Figure 2: Structure of the Relational Database

After, we made the second instantiation of the concept of taxonomy, the Stay. Since the data have be collected at intervals of five minutes, it was necessary to assume the following commitment, as we do not know the exact moment when the stay expires, we felt that in situations where the individual was not seen in the following five minutes, then stay in this place is over.
According to our definition of Stay and starting with the existing observations, the table that represents the Stays is defined by the following attributes:

<Id_Stay, Id_Artefact, Timestamp_Initial, Timestap_Final, Id_Place>

Thus it is possible to describe the initial moment in time, when a particular individual was in a certain place during a period of time. Just as happened with the table of observations also the table Stay need to relate the Place and Artefact tables.
The following table illustrates a practical example of our concept instantiation of Stay for a particular individual and a reduced period of time:

Table 1: Stay Table

Id_Stay	Id_Artefact	Timestamp _Initial	Timestamp _Final	Id_Place
1	1	16/06/2008 16:00:29	16/06/2008 16:10:43	2
2	1	16/06/2008 17:40:01	16/06/2008 18:58:34	3

In this table it is possible to see a new situation that we faced, the lack of information of the individual status between the two stays presented on the table. This absence may be related to a failure in the mechanism of collecting information or may simply be that the user between the two stays presented has not connected with any AP and as such it is impossible for us to know where it was during this time period.

This difficulty allowed us to start the research on a new concept that at this stage we do not know whether to be part of our current taxonomy or be a parallel concept to it. This concept, which we called *space leap*, will define time periods when "we lost the trail" of the artefact and as such doesn't allow us to determine whether there was a Stay for that time period or an Elementary Movement.

Although conscious of this problem, the instantiation of the concept Elementary Movement for this type of data in particular can be made much simpler in terms of attributes of the table. Since in this particular case, an elementary movement represents a shift between stays for the same artefact, a table for this information can be described only by the following attributes:

<Id_Movement, Artefact, Stay_Start, Stay_End>

Once the tables are related, from the *Stay_Start* and *Stay_End*, we can extract respectively the *Timestamp_Initial Timestamp_Final*, as well *Place_Start* and *Place_End*.

As the concept of trajectory is defined by an ordered list of elementary movements, the instantiation of this concept is described by the following table attributes:

<Id_Trajectory, Artefact, Id_Movement>

GPS Positioning

Many of the works already presented in the visualization of urban mobility area, rely on collections of data sets of readings obtained from GPS receivers. These include data beyond the position, described by a pair of coordinates (latitude and longitude), like the information of the temporal moment in which the position was acquired, corresponding to our definition of *Timestamp*. Since this type of mobility information has frequently used in the study of urban mobility and easy to access, our taxonomy should therefore adapt to this type of mobility data.

Thus, taking into account the normal format of this data acquisition (standard NMEA - National Marine Electronics Association) (NMEA 2010), for example the RMC sentence:

```
$GPRMC,123519,A,4807.038,N,01131.000,E,022
.4,084.4,230394,003.1,W*6A
```

we can easily extract information of the position and the timestamp of artefact associated with the GPS receiver. When instantiating this type of data to our concept of observation, the respective table in the DB can be defined by the following attributes:

<Id_Observation, Artefact, Timestamp, Latitude, Longitude>

In this particular case the attribute *Place* presented in the definition of the concept was replaced by the instantaneous position of the artefact since the GPS positions are typically very dynamic fruit of mobility and as such is simple to immediately include the position via a pair of coordinates to be relate them with a place. However this adaptation of the concept to the data in any way interfere with its own definition.

Because such data reflect an movement in space of the artefact, don't makes sense instantiate the concept Stay since the presence of an artefact in a particular place (considering a place as a single pair of coordinates) often lasts no more that the time intervals between the acquisition of information, which in view of our definition of Stay, would produce null intervals.

Although excluding for this type of data the representation of the concept Stay doesn't imply that didn't make the instantiation of other concepts. The taxonomy is designed so that the absence of a concept influences other concepts to a minimum. Thus, the instantiation of the Elementary Movement will be based not in the Stay concept but in the Observation and at the level of RDB can be represented by a table with the following attributes:

<Id_Movement, Artefact, Lat_Start, Long_Start, Lat_End, Long_End,Timestamp_Initial,Timestamp_Final>

basically in RDB what we are doing is associate two consecutive observations in time for the same artefact and thus describe the elementary movement.

As with the Wi-Fi data, the instantiation of the concept Trajectory is performed similarly.

DISCUSSION AND CONCLUSIONS

The taxonomy proposed in this paper seems to us with huge potential to make it a single representation of mobility data independent of the various sensors used in the acquisition of this information, being flexible enough to adapt to different types of data and thus enabling the creation of visualizations where different types of data complement each other according to a spatiotemporal reference. For example, we thus try to complement the GPS data obtained by individuals who use their car to move, with the data of public transportation and thus better describe the existing behaviours in urban space.

For the data used to make the first instantiation, we found that the collection mechanism used is not the most effective, because every five minutes a new observation is made instead of being made just two (one when the user register at AP and the other when the connection ends), thereby generating a high increase of the number of observations. But how were the data that was available at that time, we could not change the collection mechanism.

Another difficulty that we found in this particular data type and related to the technology itself, was seen when using our visualization prototypes and found that at certain times the user switched between nearby APs which does not mean that there has been in movement, because normally returned to the AP earlier.

Although currently the instantiation of our taxonomy is done only for two types of data, although more common in such studies, we intend to extend our study to other mobility data, such as data from public transportation ticketing, tolls, between others.

One of the next steps is to start the study on the concept *space leap* since its existence will always include errors in the inference of trajectories of individuals because we never know if in this period there exists an Elementary Movement. The definition of this concept is very important because for certain mobility data is often the existence of time intervals where there are no observations. In the particular case of the ticketing systems of public transportation there are situations where we only have, for example, a record of the origin of an elementary movement at the start of the day and a new record at the end of the day. In the time period between these two records we haven't any information about: the place where they finish the elementary movement and what kind of elementary movements existed in that period.

Another of our concerns is related to the space occupied at the level of storage in the DB of the observations data. Because we are working with individual data without any kind of aggregation, leads to a large number of records. So we have to evaluate the results of the visualizations and understand if it makes sense to work with individual data or taking into account these results, the information extracted from visualizations is not compromised when using the data in aggregate form.

REFERENCES

Brockmann, D. and F. Theis. 2008. "Money circulation, trackable items, and the emergence of universal human mobility patterns". *Pervasive Computing*, IEEE, 7(4), 28–35.

Hagen-Zanker, A. and H.J.P. Timmermans. 2008. "A metric of compactness of urban change illustrated to 22 european countries". In Lars Bernard, Anders Friis-Christensen, and Hardy Pundt, editors, AGILE Conf., *Lecture Notes in Geoinformation and Cartography*, 181– 200. Springer.

iSPOTS.Senseable City Lab. http://senseable.mit.edu/ispots/

Moreira, A., M.Y.Santos, M. Wachowicz, and D. Orellana. 2010. "The impact of data quality in the context of pedestrian movement analysis". In Marco Painho, Maribel Yasmina Santos, and Hardy Pundt, editors, *Geospatial Thinking, Lecture Notes in Geoinformation and Cartography*, 61–78. Springer Berlin Heidelberg.

NMEA. National Marine Electronics Association. http://www.nmea.org/

Reades, J., F. Calabrese, A. Sevtsuk, and C. Ratti. 2007. "Cellular census: Explorations in urban data collection." *IEEE Pervasive Computing*, 6(3):30–38, Jul-Sept.

Sevtsuk, A. and C. Ratti. 2005. "iSpots: How wireless technology is changing life on the MIT campus". Technical report, MIT, Cambridge.

Yu, H. and S.L. Shaw. 2004. "Representing and visualizing travel diary data: A spatio-temporal gis approach". In *2004 ESRI International User Conference*.

JOÃO PEIXOTO was born in Gouveia, Portugal and went to the University of Coimbra, where he studied Informatics Engineering and obtained his degree in 2002. He works in Centro de Computação Gráfica since 2001 and started his PhD at the University of Minho in 2008. His main research interests are: Ubiquitous and Urban Computing. His e-mail address is : joao.peixoto@ccg.pt.