EuroVis Workshop on Visual Analytics (2013)
M. Pohl and H. Schumann (Editors)

Towards Privacy-Preserving Semantic Mobility Analysis

N. Andrienko[1], G. Andrienko[1] and G. Fuchs[1]

[1]Fraunhofer IAIS, Sankt Augustin, Germany

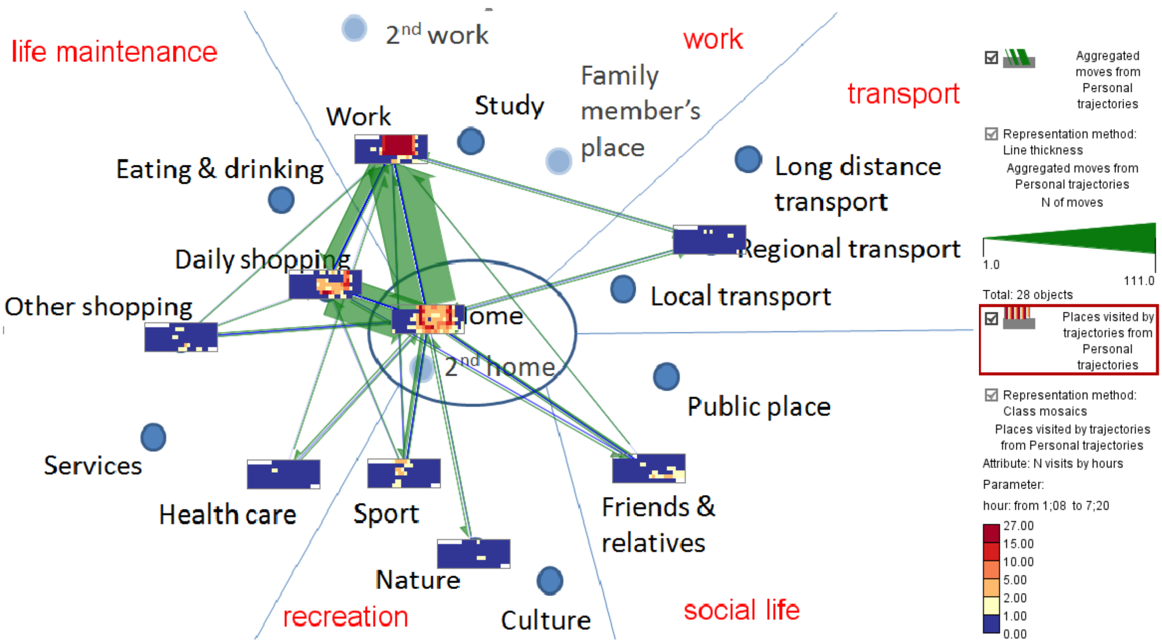

Figure 1: *A exemplary chorematic representation of visit dynamics to semantic places and magnitudes of flows between them.*

Abstract

By analyzing data reflecting human mobility, one can derive patterns and knowledge that are tightly linked to the underlying geography and therefore cannot be applied to another territory or even compared with patterns obtained for another territory. Another problem of mobility analysis is compromising personal privacy, since person identities can be determined based on the regularly visited geographical locations. We here propose an idea for novel approach based on transformation of the spatial component of movement data from the geographic space to an abstract semantic space, inspired by the concept of cartographic chorems. We demonstrate that many visual analytics procedures developed for geographic movement data can be adapted for privacy-preserving mobility analysis based on semantic spaces.

1. Introduction

Mobility is an essential component of human life and economic activity. Yet, scientists did not come close to understanding the phenomenon of human mobility in its full variety of interplaying aspects: spatial, temporal, social, psychological, economical, and environmental. While lots of mobility-related data are available nowadays, there is a principal obstacle to gaining deep understanding of the mobility phenomenon from these data. The problem lies in the

© The Eurographics Association 2013.

currently existing methodology for mobility analysis, which can only enable discovering patterns and gaining knowledge that are specific to the territory covered by the available data. Analysis results obtained for different territories are incomparable because they refer to specific geographic locations with their specific properties and interrelations. Even highly general statistics, such as statistical distribution of traveled distances, are dependent on the extent and specific spatial properties of a territory under analysis. This geographic specificity precludes gaining more general knowledge about human mobility that could apply to multiple territories.

To solve this problem, it is necessary to represent and analyze mobility-related data and information on a higher level of abstraction than is possible on the basis of geographic space. The spatial component of mobility data and information needs to be transformed from geographic space to an abstract semantic space. The components and structure of the semantic space depend on the level and goals of the intended analysis and not on the specific geographic properties of the territory the available data refer to. References to specific geographic locations in mobility data are converted into references to abstract semantic locations in the semantic space. These abstracted data and information representations are then analyzed for gaining knowledge that allows comparison between territories, generalization and application to other territories, and essentially contributes to the overall understanding of the phenomenon of human mobility.

We are here presenting the core idea of a novel approach we are developing with the goal to attain a sophisticated, scalable, privacy-preserving semantic mobility analysis methodology. This approach is based on semantic abstractions of geographic space inspired by the concept of chorems [Bru86]. Chorems are schematic visual representations of task-relevant spatial or spatio-temporal information at a high abstraction level. We extend the meaning of the term to include not only visual but also formal, computer-processable representations of abstract semantic spaces and semantically abstracted data and information populating these spaces. An important property of chorematic abstractions is their capability to conceal sensitive individual information related to individual's positions in geographical space. The abstraction from geographic space to a universal semantic space common for all individuals facilitates privacy-preserving analysis of individuals' mobility data. In this paper we are discussing the key concepts behind this chorematic movement analysis along with illustrative examples to show the principal feasibility and utility of our proposed approach.

2. Related work

Extensive research on trajectory analysis has been conducted in information visualization and Visual Analytics [AA13], knowledge discovery in databases [GP08], spatial computing [Lau09], and moving object databases [GS05]. According to the recent detailed overview of the state-of-the-art in semantic trajectory analysis [PSR*13], the existing research belongs to three major categories: annotation of trajectory segments with references to visited places of interest (POI), determining the most probable transportation modes, and discovery of predefined movement patterns, such as 'meet', 'flock', 'leadership', etc. In the first category, the annotation is done using a database of public POI. This approach cannot find personal POI, such as home, work place, friend's home, etc. Ahas et al. [ASJ*10] extract frequently visited locations from mobile phone use records and classify some of them as home or work places based on the call frequency and the time of the day when the calls are made. Meanings of personal places can also be interpreted based on the daily and weekly patterns of the place visit times [AAB*13]. Papers [AA12] and [AAK*11] discuss privacy issues in applying visual analytics methods to movement data.

3. From maps of trajectories to chorems of semantic trajectories

Based on existing typologies of human activities (e.g., [JFG12]), we create a schematic hierarchical map of personal places and activities that can serve as a reference frame (semantic space) for representing movement data and analytical artifacts, see Figure 2.

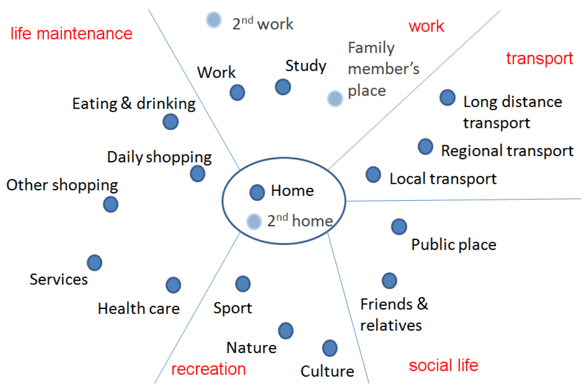

Figure 2: *A chorematic representation of a semantic space of human activities. Note this figure represents only an illustrative, manually laid out example.*

To convert trajectories from geographic to semantic space, the geographic locations occurring in the trajectories need to be semantically interpreted. Sometimes trajectory data include annotations describing the visited places and/or activities of the individuals. If this is not the case, the following procedure can be applied. Personal POI are extracted from trajectories as described in [AAH*11b]. First, each person's stop events with a chosen minimum duration are extracted. Next, spatial clusters of each person's stop events are discovered by means of density-based clustering. By outlining the clusters, personal POI are defined. The POI are classified into semantic categories based on the profiles of the hourly counts of the visits during the work days and weekends. The procedure is described in [AAB*13], section 7.2.6.

© The Eurographics Association 2013.

Geographic trajectories enriched with place annotations are transformed into abstracted semantic trajectories by replacing the geographic positions of the stops that occurred in the annotated places by the positions of the types of these places in the semantic space. The trajectory segments between the stops are omitted.

4. Mobility analysis in semantic spaces

Semantically abstracted trajectories can be analyzed using various visual analytics methods ([AA13], [AAB*13]) focusing on four aspects of mobility: movers (moving objects) with their trajectories, events positioned in space and time, locations visited by movers, and time units with respective spatial situations (Figure 3). We shall present examples of addressing these aspects using daily trajectories of two persons, one recorded during a year in a small town in Germany and the other observed over a period of two month in a large city in the USA. Note that in particular the concrete geographic regions and the time periods under observation as well as the number of observations (count of semantic trajectories) differ significantly between the data sets of both persons, which would present a formidable obstacle to the majority of existing analysis methods.

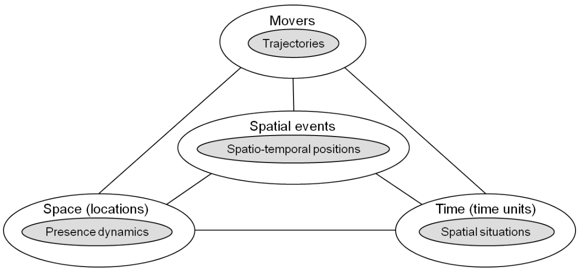

Figure 3: *A multi-perspective view of movement.*

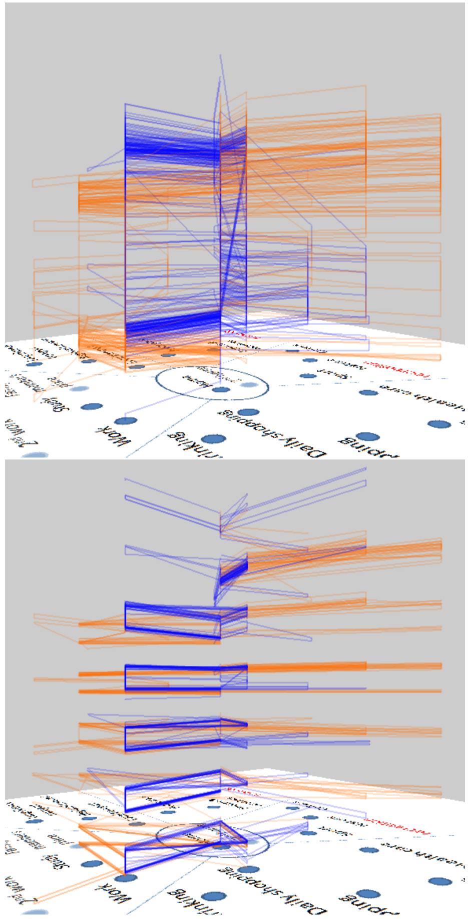

Figure 4: *STC views of two persons' (blue – Germany, orange – USA) semantic trajectories aligned to show daily (top) and weekly (bottom) temporal cycles.*

4.1. Analysis focusing on movers

To find typical daily sequences of activities, semantic trajectories of one or several persons can be transformed with respect to their temporal references. Figure 4 shows space-time cube (STC) representations [Häg70] of two persons' trajectories aligned to different temporal cycles. In the upper STC, trajectories of working days have been aligned to show overall patterns and principal differences between the two persons' daily activity patterns. In the bottom STC, alignment by weekly cycles was selected, thus also revealing the expected differences between working days and weekends for both persons; but also differences in semantic places visited primarily on weekends by either person thus indicating differences in lifestyles and recreational activities.

4.2. Analysis focusing on events

Semantic trajectories can also be viewed as events of visiting locations in the semantic space. It is possible to analyze the temporal distribution of these events, in particular, with regard to the temporal cycles [AAH11a]. We omit obvious examples due to the space limits.

4.3. Analysis focusing on space

To analyze the temporal patterns of person's activities, which are reflected in person's presence in different semantic places, we compute the presence counts in each semantic place for the 7x24 hourly intervals of the weekly cycle. The results for the person from Germany are shown by mosaic diagrams in Figure 1. Blue pixels represent zero counts and the colors from yellow to red represent positive values. Rows of the diagrams correspond to the days of the week and columns to the hours of the day. We can see that the *work* activity occurs mostly at regular times of the working days and occasionally on Saturday and Sunday. *Sport* activities occur mostly in the morning. *Friends and relatives* are visited from Friday to Sunday. The transitions between the activities are represented by the flows between the semantic places.

© The Eurographics Association 2013.

Figure 5 shows the differences between the activity patterns of the two persons along two different dimensions: the flows (aggregated movements) between the semantic locations and the visits of the locations. In Fig. 5 top, the flows of two persons are overlaid on the same map for visual comparison. The widths of the arrows represent the flow magnitudes and color identifies the respective person (blue – Germany, orange – USA). Fig. 5 bottom demonstrates the possibility to compare by arithmetically subtracting data of one person from data of the other. The differences between the relative frequencies of the hourly visits are encoded by a bi-variant color scale (legend on bottom right). The shades of purple mean higher presence frequencies of the German person and shades of orange mean higher values of the other person.

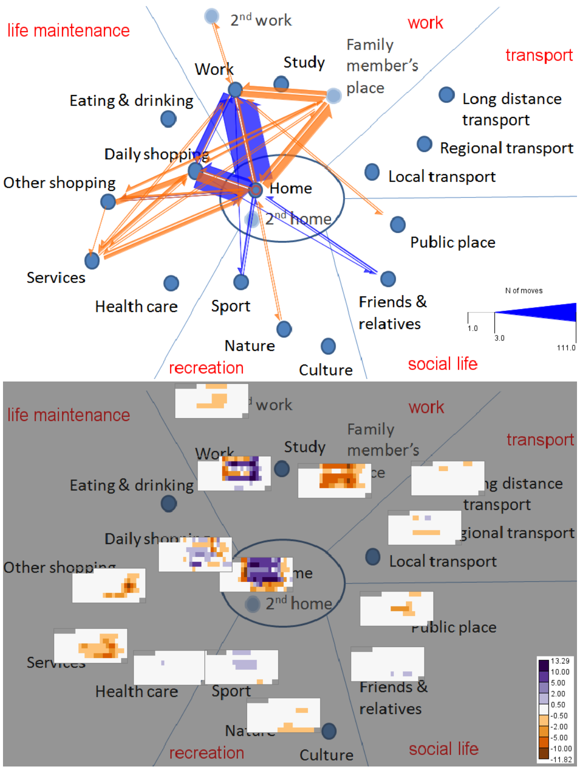

Figure 5: *Comparison between semantic trajectories of two different persons: magnitudes of flows between (top) and relative frequencies of visits to (bottom) semantic places.*

4.4. Analysis focusing on time units

This analysis is done for the person from Germany. The other person's data can be analyzed analogously. We cluster the hourly time intervals by similarity of the spatial situations in terms of the presence in the semantic places and the flows between the places. Colors are assigned to the clusters so as to reflect the similarities between their average feature profiles ([AAB*10], [AAS*12]). Due to the space limits, we cannot include the images of the spatial situations corresponding to the clusters. The time mosaics in Figure 6 show the positions of the time clusters within the daily and weekly time cycles. Each colored square represents one hourly interval; the color represents the cluster membership.

Clustering according to the presence (Figure 6 left) shows that during the working days there are two one-hour periods starting from 10:00 and 19:00 that are dissimilar to the others. In these times, the presence is shared between *home*, *work*, and *daily shopping*. All daytime hours of the working days are similar (the person mostly stays in the work place), with increasing variation towards the end of the week. On Saturday, there is a distinct time period dedicated to shopping. Clustering by flows (Figure 6 right) reveals the usual times of the morning and evening moves during the working days and shows that they are very different. The situations during the work hours are similar to those in the night; hence, the person rarely moved during the work hours.

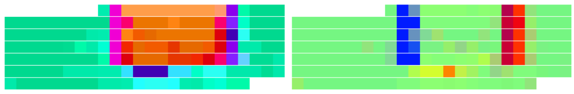

Figure 6: *A calendar view of the clusters of hourly intervals according to the distributions of the presence (left) and flows (right). The colors represent the cluster membership of the time intervals; similar colors correspond to similar clusters.*

5. Discussion and conclusions

A good feature of the proposed methodology is that the analysis results are comparable across different individuals, regions, and spatial scales. Thus, using the same abstract semantic space as shown in Figure 1, Figure 5 compares the mobility of two different persons from a small town in Germany as well as a large city in the USA. We see some similarities between their life styles: both work from Monday to Friday at regular times and sometimes spend evenings of the working days and Saturdays with their friends or relatives. We also see differences: the second person rarely goes directly from *home* to *work* or from *work* to *home* but usually stops at a *Family member's place* twice a day, probably, bringing and picking up a family member.

We have demonstrated that a number of visual analytics procedures developed for traditional movement data can be successfully applied or adapted to semantically abstracted movement data. Since the transformation conceals specific geographic locations visited by individuals, they cannot be identified from the transformed data and detected patterns; hence, the analysis is done in a privacy-preserving way.

The work presented here are the only first steps towards a comprehensive application of the chorematic approach. In the future, we are going to extend our study to creating automated visual analytics procedures supporting the whole process of the semantic analysis of mobility, including identification of individual POI, creation of chorematic representations of semantic spaces, and analysis of movement data in the semantic space.

© The Eurographics Association 2013.

References

[AA12] ANDRIENKO G., ANDRIENKO N.: Privacy issues in geospatial visual analytics. *Advances in Location-Based Services* (2012), 239–246. 2

[AA13] ANDRIENKO N., ANDRIENKO G.: Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization 12*, 1 (2013), 3–24. 2, 3

[AAB*10] ANDRIENKO G., ANDRIENKO N., BREMM S., SCHRECK T., VON LANDESBERGER T., BAK P., KEIM D. A.: Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Comput. Graph. Forum 29*, 3 (2010), 913–922. 4

[AAB*13] ANDRIENKO G., ANDRIENKO N., BAK P., KEIM D., WROBEL S.: *Visual Analytics of Movement*. Springer, 2013. 2, 3

[AAH11a] ANDRIENKO G., ANDRIENKO N., HEURICH M.: An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science 25*, 9 (2011), 1347–1370. 3

[AAH*11b] ANDRIENKO G. L., ANDRIENKO N. V., HURTER C., RINZIVILLO S., WROBEL S.: From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *IEEE VAST* (2011), IEEE, pp. 161–170. 2

[AAK*11] ANDRIENKO G. L., ANDRIENKO N. V., KEIM D. A., MACEACHREN A. M., WROBEL S.: Challenging problems of geospatial visual analytics. *J. Vis. Lang. Comput. 22*, 4 (2011), 251–256. 2

[AAS*12] ANDRIENKO N., ANDRIENKO G., STANGE H., LIEBIG T., HECKER D.: Visual analytics for understanding spatial situations from episodic movement data. *KI 26*, 3 (2012), 241–251. 4

[ASJ*10] AHAS R., SILM S., JÄRV O., SALUVEER E., TIRU M.: Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology 17*, 1 (2010), 3–27. 2

[Bru86] BRUNET R.: La carte-modèle et les chorèmes. *Mappemonde 86*, 4 (1986), 2–6. 2

[GP08] GIANNOTTI F., PEDRESCHI D. (Eds.): *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer, 2008. 2

[GS05] GÜTING R. H., SCHNEIDER M.: *Moving Objects Databases*. Morgan Kaufmann, 2005. 2

[Häg70] HÄGERSTRAND T.: What about people in regional science? *Papers in Regional Science 24* (1970), 6–21. URL: http://dx.doi.org/10.1007/BF01936872, doi:10.1007/BF01936872. 3

[JFG12] JIANG S., FERREIRA J., GONZÁLEZ M. C.: Clustering daily patterns of human activities in the city. *Data Min. Knowl. Discov. 25*, 3 (2012), 478–510. 2

[Lau09] LAUBE P.: Progress in movement pattern analysis. In *BMI Book* (2009), Gottfried B., Aghajan H. K., (Eds.), vol. 3 of *Ambient Intelligence and Smart Environments*, IOS Press, pp. 43–71. 2

[PSR*13] PARENT C., SPACCAPIETRA S., RENSO C., ANDRIENKO G., ANDRIENKO N., BOGORNY V., DAMIANI M. L., GKOULALAS-DIVANIS A., MACEDO J., PELEKIS N., THEODORIDIS Y., YAN Z.: Semantic trajectories modeling and analysis. *ACM Computing Surveys 45*, 4 (2013). 2

© The Eurographics Association 2013.