Trajectories: analysis

Noam Shoval
The Hebrew University of Jerusalem, Israel

Data about movements of objects are often collected as trajectories in space and time, that is, the movement of each object is recorded as a series of geographic locations with respective time stamps, when the object moves in the basic three-dimensional (3-D) framework of our physical world, defined by geographic space and time.

Trajectories in the geographic context usually refer to the space–time paths of people or cars and are mostly related to the conceptual framework of time geography proposed by Torsten Hägerstrand (1970) that integrates time, as a limited resource, into the thought on spatial behavior.

Qualitative analysis of trajectories

One of Hägerstrand's most profound professional and disciplinary achievements was the ability to represent space and time in a single diagram, unlike an ordinary map, but rather like a snapshot, that reproduces a moment frozen in time. The result was his, now famous, time-geographical diagrams: notational (representational) systems, which formed the basis of much of the subsequent work in the field of time geography, particularly in the realm of analysis and interpretation. The diagrams (Figure 1) consist, as a rule, of two axes: a time-axis and a space-axis, thus making it possible to trace in graphic terms individual time budgets. The effective range of each person is described by a prism, or a series of prisms, whose shape is dependent upon the aforementioned capability constraints. Hence, every pause, regardless of the activity involved, will cause the prism's (or subprism's) range to shrink in direct proportion to the time spent at said stop. But there are also other wider structural features, specific to the social systems within which individuals operate, which, as has long been recognized, help shape people's time budgets and activity patterns (Neutens, Schwanen, and Witlox 2011).

The new computation abilities that have become available within the past two decades have opened new possibilities in the ability to plot space–time data. Kwan (2000) was among the first to produce a "space–time aquarium" within a 3-D geographic information system (GIS) environment using individual activity travel diary data. These images have improved the capability to reveal the characteristics of space–time patterns belonging to different population subgroups, and they have also assisted in improving the ability to identify common patterns.

However, one basic problem in trajectories analysis is the limited ability to aggregate space–time paths in order to create generalized types of trajectories composed of varied activities in order to create patterns fashioned on a quantitative basis while taking into account the sequential element (Andrienko and Andrienko 2010). Previous attempts with quantitative pattern aggregation methods, mainly by transport researchers, did not manage to tackle the issue of the sequential element. Understanding the sequence of activities in space and time allows one to understand an additional integral dimension of activity and to recognize patterns that exist within this dimension.

The International Encyclopedia of Geography.
Edited by Douglas Richardson, Noel Castree, Michael F. Goodchild, Audrey Kobayashi, Weidong Liu, and Richard A. Marston.

DOI: 10.1002/9781118786352.wbieg0611

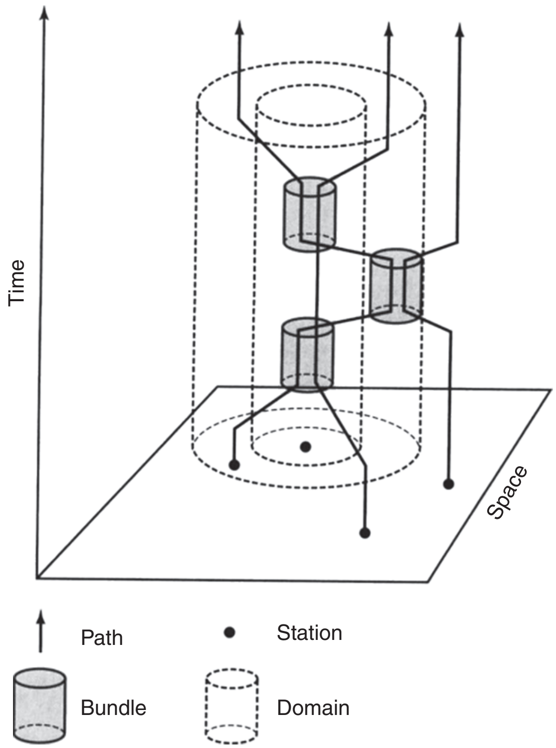

Figure 1 Hägerstrand's time-geographical diagram. Reproduced from Neutens, Schwanen, and Witlox (2001). Reprinted by permission of the publisher (Taylor & Francis Ltd, www.tandfonline.com).

Geography, like many other research and application fields, has moved in recent decades from being a data-poor field to being a data-rich field. This transition happened with the development of fields such GIS and remote sensing, and today public and private sector agencies are creating larger and larger datasets. This change made it imperative to develop new data mining tools to extract and construct knowledge from the huge databases that otherwise will remain undeciphered. The introduction of location devices, such as cellular phones, has created databases of human activities in space–time in great detail, but due to their large size there is need for new techniques in order to be able to extract patterns of behavior from within those large databases.

Quantitative analysis of trajectories

Attempts to aggregate behaviors into patterns and to analyze behaviors have mostly been descriptive; only a handful of studies introduced analytic methods to tackle the new type of high-resolution data available. The realm of trajectory pattern analysis offers many tools that can contribute to the analysis of spatial behavior. Research in this field grows out of the increasing availability of spatiotemporal data and the lack of techniques to analyze it (D'Urso and Massari 2013). The basic entities in this line of research are moving objects' trajectories that have a defined beginning and end time, and are divided into movement segments by stops – pauses in movement, identified in accordance with the scale of analysis. Yet, when considering semantic trajectories, one must consider background geographic information in order to understand and model trajectory patterns.

Trajectory pattern analysis attempts to resolve three major issues: how to measure similarities between trajectories, how to recognize clusters of similar trajectories, and how to treat the sequential element of a trajectory (Grinberger *et al.* 2014). Many measures have been developed to identify the similarity of trajectories – distance-based measures, dynamic time wrapping, and longest common subsequence being some of the most frequently used (Chen, Özsu, and Oria 2005). A variety of clustering methods are also used in pattern analysis, following the methods used for clustering of static data: direct methods that create clusters according to a central value or entity (*k*-means, *k*-medoids, fuzzy *c* means); agglomerative methods, which use a hierarchical tree to cluster trajectories together until a certain threshold is reached (hierarchical clustering); divisive methods, which work in a top-down manner, dividing an entire database into a number of clusters; density-based methods, which extend the cluster's reach until a certain threshold number of objects is crossed; and so on. In order to deal with the elements of time and sequence, three approaches are available (Liao 2005): a raw-data based approach requiring substitution of standard similarity measures with ones that take time into account; the feature-based approach, in which raw data are converted into a feature vector, upon which standard clustering methods can operate; the model-based approach, which likewise converts raw data into prespecified model parameters.

The analysis of trajectories and pattern mining has proven to be useful in many areas aside geography. The fields of travel research and transport planning may be the most relevant fields, with some examples being aggregation of trajectories into *trajectory patterns*, based on regions of interest (predefined or discovered from data), and the sequence and duration of movement between them (Giannotti *et al.* 2007); identification of *mobility profiles* – the routine routes of an individual based on data-defined stops and the sequence between them (Trasarti *et al.* 2011); and the work done by Renso *et al.* (2013), which combines a predefined mobility behavior ontology and movement patterns aggregated from trajectories (measured, synthetic, or semantic) in order to classify patterns into behavior classes. Lately, this form of analysis has proven useful when implemented within the framework of time geography, as is done by the Activity Pattern Analyst, developed by Chen *et al.* (2011).

Sequence alignment methods

Based on the principle of comparing sequences (strings), sequence analysis was developed during the 1980s for use in the natural sciences and was utilized primarily by biochemists to, among other things, analyze DNA sequences. It was adapted for use in the social sciences sometime toward the end of the 1990s (Wilson 1998). Unlike conventional quantitative methods, sequence analysis, as its name suggests, tackles the problem of sequences directly. Sequence analysis could, if applied properly, complement the methods used to collect, present, and analyze temporal and spatial data, particularly the data amassed in the course of empirically based time geography research.

Traditional quantitative methods band similar objects together on the basis of specific shared characteristics, but they cannot expose the hidden patterns buried within sequences (Wilson 1998). Unlike the more generally accepted methods of sequence comparison in which the distance between two sequences of activities is calculated by means of Euclidian-based geometry, such as Euclidean distance, city-block distance, or Hamming distance (the number of positions in which corresponding elements are different), sequence analysis computes the distance between the two on a "biological" basis (Bargeman, Joh, and Timmermans 2002).

The algorithm used in sequence alignment to measure the degree of similarity between two sequences utilizes three elementary operations: insertion, deletion (two operations, which are, on occasion, referred to singly as an "indel"), and substitution (switching the places of two characters). By applying these three processes to one of the sequences, one string is made identical to the other. The more operations needed to make the sequences identical, the longer the distance between the sequences. Hence, the longer the distance between groups, the smaller the similarity. Thus, sequence alignment methods measure the degree of difference between two sequences in terms of their element composition and sequence.

Human spatial behavior is the sum of three parallel dimensions: the "what," the "when," and the "where." The "what" describes the activity taking place, the "when" the temporal dimension of said activity, and the "where" its spatial aspects. Geographical analysis combines these three dimensions to reproduce and understand part of the complex reality of human behavior. Those social scientists who have used sequence analysis to date have tended to focus on the relationship between the "what" (activity) and the "when" (time). In their work, the temporal dimension forms the basis of the sequences analyzed, with each unit (character or word) representing an activity carried out within a specific time frame. Incorporating information such as the frequency, duration, and timing of an activity, the temporal dimension is as a rule analyzed by traditional quantitative methods (Bargeman, Joh, and Timmermans 2002). The location component "where" can be approached in two different ways. The first is "categorized" locations, descriptions of the functionality of the location, such as home, work, and store. Although these words describe locations, they describe the functionality of the location and not the geographical location. These locations are also not unique. Many different homes exist in different locations, and their geographical location cannot be derived from this kind of location description. (For examples of research applying this approach, see Bargeman, Joh, and Timmermans 2002; Stovel and Bolan 2004; Shoval and Isaacson 2007; Shoval *et al.* 2015.) The second way to approach the location component of activity is by using geographic locations. This approach creates locations that are unique in the sense that each set of geographic coordinates symbolizes a specific location in space.

The method offers an effective means to extract sequence patterns from trajectories of human activities and by doing so may also present new ways of analyzing such data. It is equally worth noting that the relatively new and more accurate digital methods of collecting spatial data (i.e., GPS) produce databases that are characterized by extremely high temporal and spatial resolutions. This reinforces the observations to the effect that recent developments in the field of *location aware technologies* (LAT) and *location-based services* (LBS) could trigger an even wider resurgence in time-geographic studies (Miller 2005; Wilson 2008).

SEE ALSO: Representation: trajectories; Time geography and space–time prism

References

Andrienko, G., and N. Andrienko. 2010. "A General Framework for Using Aggregation in Visual Exploration of Movement Data." *Cartographic Journal*, 47(1): 22–40.

Bargeman, B., C.-H. Joh, and H. Timmermans. 2002. "Vacation Behavior Using a Sequence Alignment Method." *Annals of Tourism Research*, 29(2): 320–337.

Chen, L., M.T. Özsu, and V. Oria. 2005. "Robust and Fast Similarity Search for Moving Object Trajectories." In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, Baltimore, Maryland*, 491–502. New York: Association for Computing Machinery.

Chen, J., S.L. Shaw, H. Yu, *et al.* 2011. "Exploratory Data Analysis of Activity Diary Data: A Space–Time GIS Approach." *Journal of Transport Geography*, 19(3): 394–404.

D'Urso, P., and R. Massari. 2013. "Fuzzy Clustering of Human Activity Patterns." *Fuzzy Sets and Systems*, 215: 29–54.

Giannotti, F., M. Nanni, D. Pedreschi, and F. Pinelli. 2007. "Trajectory Pattern Mining." In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by

P. Berkhin, R. Caruano, and X. Wu, 330–339. New York: Association for Computing Machinery.

Grinberger, Y., N. Shoval, B. McKercher, and E. Ng. 2014. “Typologies of Tourists' Time–Space Consumption: A New Approach Using GPS Data and GIS Tools.” *Tourism Geographies*, 16(1): 105–123.

Hägerstrand, T. 1970. “What about People in Regional Science?” *Papers of the Regional Science Association*, 24(1): 7–21.

Kwan, M.P. 2000. “Interactive Geovisualization of Activity–Travel Patterns Using Three-Dimensional Geographical Information Systems: A Methodological Exploration with a Large Data Set.” *Transportation Research C*, 8(1–6): 185–203.

Liao, T.W. 2005. “Clustering of Time Series Data – A Survey.” *Pattern Recognition*, 38(11): 1857–1874.

Miller, H.J. 2005. “A Measurement Theory for Time Geography.” *Geographical Analysis*, 3(1): 17–45.

Neutens, T., T. Schwanen, and F. Witlox. 2011. “The Prism of Everyday Life: Towards a New Research Agenda for Time Geography.” *Transport Reviews*, 31(1): 25–47.

Renso, C., M. Baglioni, J.A. Macedo, *et al*. 2013. “How You Move Reveals Who You Are: Understanding Human Behavior by Analyzing Trajectory Data.” *Knowledge and Information Systems*, 37(2): 331–362.

Shoval, N., and M. Isaacson. 2007. “Sequence Alignment as a Method for Human Activity Analysis in Space and Time.” *Annals of the Association of American Geographers*, 97(2): 282–297.

Shoval, N., B. McKercher, A. Birenboim, and E. Ng. 2015. “The Application of a Sequence Alignment Method to the Creation of Typologies of Tourist Activity in Time and Space.” *Environment and Planning B: Planning and Design*, 42(1): 76–94.

Stovel, K., and M. Bolan. 2004. “Residential Trajectories: Using Optimal Alignment to Reveal the Structure of Residential Mobility.” *Sociological Methods Research*, 32(4): 559–598.

Trasarti, R., F. Pinneli, M. Nanni, and F. Giannotti. 2011. “Mining Mobility User Profiles for Car Pooling.” In *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, edited by C. Aptè, J. Ghosh, and P. Smyth, 1190–1198. New York: Association for Computing Machinery.

Wilson, C. 1998. “Activity Pattern Analysis by Means of Sequence-Alignment Methods.” *Environment and Planning A*, 30(6): 1017–1038.

Wilson, C. 2008. “Activity Patterns in Space and Time: Calculating Representative Hagerstrand Trajectories.” *Transportation*, 35(4): 485–499.