Journal of Transport Geography 19 (2011) 394–404

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Exploratory data analysis of activity diary data: a space–time GIS approach

Jie Chen [a,*], Shih-Lung Shaw [b], Hongbo Yu [c], Feng Lu [a], Yanwei Chai [d], Qinglei Jia [a]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 2309, 11A Datun Road, Chaoyang District, Beijing 100101, PR China
[b] Department of Geography, University of Tennessee, 304 Burchfiel Geography Building, Knoxville, TN 37996-0925, USA
[c] Department of Geography, Oklahoma State University, 337 Murray Hall, Stillwater, OK 74078-4073, USA
[d] Department of Urban and Economic Geography, Peking University, 5 Yiheyuan Road, Haidian District, Beijing 100871, PR China

ARTICLE INFO

Keywords:
Time geography
Space–time GIS
Activity diary data

ABSTRACT

Study of human activities in space and time has been an important research topic in transportation research. Limitations of conventional statistical methods for analysis of individual-level human activities have encouraged spatiotemporal analysis of human activity patterns in a space–time context. Based on Hägerstrand's time geography, this study presents a space–time GIS approach that is capable of representing and analyzing spatiotemporal activity data at the individual level. Specifically, we have developed an ArcGIS extension, named Activity Pattern Analyst (APA), to facilitate exploratory analysis of activity diary data. This extension covers a set of functions such as space–time path generation, space–time path segmentation, space–time path filter, and activity distribution/density pattern exploration. It also provides a space–time path based multi-level clustering method to investigate individual-level spatiotemporal patterns. Using an activity diary dataset collected in Beijing, China, this paper presents how this Activity Pattern Analyst extension can facilitate exploratory analysis of individual activity diary data to uncover spatiotemporal patterns of individual activities.

1. Introduction

Studying human activities in space and time has been an important topic in transportation research (e.g., Szalai, 1972; Kitamura, 1988; Golob and McNally, 1997; Robinson and Godbey, 1997; Bhat and Koppelman, 1999; Joyce and Stewart, 1999; Lu and Pas, 1999; Stinson, 1999; Goulias, 2002; Pendyala and Goulias, 2002). Activity diary data have served as a major data source in many of these studies. An activity diary dataset records activities conducted by sample individuals within a particular time period (e.g., one day or multiple days), including information such as when an activity began and ended, where an activity took place, characteristics of each activity, and individuals who participated in those activities (Michelson, 1973). As a result, an activity diary dataset provides a means to record individual activities in a spatial and temporal context. With a detailed activity diary dataset, it is possible to gain insight into human activity patterns that can help researchers explore and better understand how individuals interact with other people and the environment.

Activity patterns often are analyzed with statistical methods (e.g., Cullen et al., 1972; Chapin, 1974; Shapcott and Steadman, 1978; Klepeis et al., 2001; Chai et al., 2002; Vrotsou et al., 2007; Jim and Chen, 2009). Although these statistical methods are very useful in studying aggregate characteristics of individual activities, they are less helpful in analyzing individual activity patterns and interactions in a space–time context. As activities occur in both space and time (Anderson, 1971), it is essential to treat space and time jointly in activity studies (Pred, 1977). An integrated space–time analytical environment also allows researchers to investigate individual activities as processes rather than separate events. Thus, exploratory data analysis functions supported by an integrated space–time framework can make important contributions to activity-based transportation studies.

Hägerstrand's time geography offers a useful conceptual framework to study individual activity patterns under various constraints in a space–time context (Hägerstrand, 1970, 1978, 1989). Its space–time path concept, which represents the spatial movements of an individual over time, presents an effective form to model spatiotemporal characteristics of individual activities. There have been a number of efforts incorporating time-geographic concepts into a geographic information system (GIS) environment to represent and analyze individual activities in both spatial and temporal dimensions (e.g., Miller, 1991, 1999; Kwan and Hong, 1998; Kwan, 2000; Kwan and Lee, 2003; Buliung and Kanaroglou, 2006; Yu, 2006; Neutens et al., 2007, 2008a,b; Kang and Scott, 2008; Yu and Shaw, 2008; Shaw et al., 2008; Shaw and Yu, 2009; Kraak and Huisman, 2009). These efforts demonstrate

* Corresponding author. Tel.: +86 10 6488 9015x3; fax: +86 10 6488 9630.
E-mail addresses: chenj@lreis.ac.cn (J. Chen), sshaw@utk.edu (S.-L. Shaw), hongbo.yu@okstate.edu (H. Yu), chyw@pku.edu.cn (Y. Chai), jiaql@lreis.ac.cn (Q. Jia).

0966-6923/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2010.11.002

that time geography and GIS together can provide a useful analytical environment to visualize and explore individual-level activity data in a space–time context. As individual-level activity diary datasets are becoming easier and more affordable to collect due to the advancement of location-aware technologies, there is a growing need for a spatiotemporal analytical environment that can help researchers investigate complex human activity and interaction patterns hidden in activity dairy datasets. Built upon previous studies, we develop a space–time GIS implemented as an ArcGIS extension to facilitate spatiotemporal exploratory analysis of activity diary datasets. Using a large activity diary dataset collected in Beijing, China in 2007, this study demonstrates that this extension is capable of managing, querying, analyzing, and visualizing complex activity data at the individual level in a space–time GIS environment.

The rest of this paper is organized as follows. The next section provides a brief review of research related to the development of a space–time GIS for exploring individual activity diary datasets. Section 3 discusses the functions available in the Activity Pattern Analyst extension developed in this study for examining spatiotemporal patterns of human activities hidden in activity diary datasets. Section 4 uses an activity diary dataset collected in Beijing to present a case study illustrating the benefits of employing this space–time GIS approach to gain insight of human activity patterns. Concluding remarks are presented in the final section.

2. Related research

An individual-level activity diary dataset stores both spatial and temporal information of individuals' activities and offers a valuable source to gain understanding of disaggregate and aggregate activity patterns in space and time. When the sample size of an activity diary dataset is large, it presents a challenge of extracting useful information from the dataset to reveal complex human activity and interaction patterns. Exploratory data analysis can serve as a useful first step for researchers to gain insight of the data and formulate hypotheses for further studies (Gahegan, 2000; Guo et al., 2005).

Hägerstrand (1970) proposed an approach of studying human activities conducted under various constraints in a space–time context, which becomes known as time geography. This time geography approach represents individual activities in an integrated space and time environment (Gren, 2001). This space–time system is a three-dimensional orthogonal system that consists of two spatial dimensions and a temporal dimension (Fig. 1) (Shaw et al., 2008). The 2D spatial dimensions track locational changes of individuals, while the temporal dimension arranges human activities according to their chronological order. The space–time path concept, which provides an efficient form to represent an individual's movement history in a space–time system (Miller, 2004), denotes a series of characteristics of an individual's daily activities including location, time, duration, sequence, frequency and type of activities (Kwan, 2002; Scott, 2006; Ren and Kwan, 2009). Time geography considers that everything being done, including "do nothing", is an activity (Ellegåd, 1999). Therefore, every point on a space–time path is associated with at least one activity. To be consistent with this definition, an activity in this paper can be either a mobile activity that moves from one location to another location or a stationary activity that is conducted at a fixed location. A stationary activity appears as a vertical line segment and a mobile activity shows as a tilted line segment in a space–time path (Fig. 1).

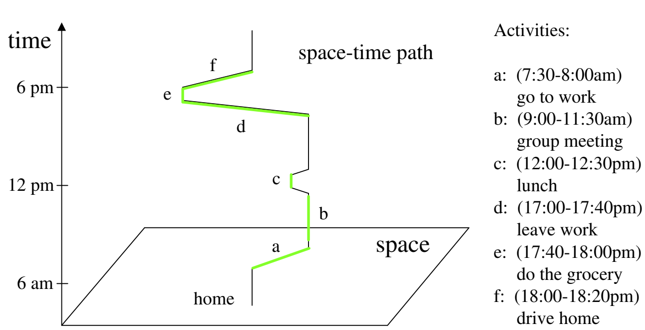


Fig. 1. Space–time system and space–time path.

With the modern geocomputational and visualization capabilities, GIS have been recognized as a potential approach to implementing the time geography framework and supporting assessment of spatiotemporal characteristics of human activities (Pipkin, 1995). Attempts have been made to store and manage individual activities in GIS to support basic queries of spatial and temporal characteristics (e.g., Shaw and Wang, 2000; Wang and Cheng, 2001; Frihida et al., 2002; Buliung and Kanaroglou, 2004). In recent years, a number of studies have incorporated time-geographic concepts into GIS to study individual accessibility (e.g., Kwan and Hong, 1998; Kwan, 1998; Miller, 1999; Weber and Kwan, 2002; Weber, 2003; Neutens et al., 2007, 2008a,b; Forer, 1998; Huisman, 2006; Kraak and Huisman, 2009; Miller and Bridwell, 2009), to implement visualization of space–time paths (e.g., Kwan 1999a,b, 2000; Kwan and Lee 2003; Yu, 2007) and to explore human activities and interactions (e.g., Yu, 2005, 2006; Buliung and Kanaroglou, 2006; Yu and Shaw, 2007, 2008; Kang and Scott, 2008; Kang et al., 2009). In particular, several GIS toolkits have been implemented accordingly (e.g., Buliung and Kanaroglou, 2006; Shaw et al., 2008; Yu and Shaw, 2008; Kang and Scott, 2008). These studies indicate that a space–time GIS (i.e. 2D space + 1D time) offers a powerful environment to represent the key concepts of time geography. Moreover, analysis functions developed in a space–time GIS framework present new methods of studying complex spatiotemporal human activity patterns and interactions (Yu and Shaw, 2008; Shaw and Yu, 2009).

There is a growing interest in exploring spatiotemporal characteristics and patterns of activities at the individual level as individual activity datasets have become increasingly available to the research community. Who share similar spatiotemporal activity patterns? How does a group of people organize their daily travels and activities differently in space and time from another group of people (e.g., office workers vs. household wives)? Methods are needed to help researchers identify and investigate these patterns in an individual-level activity dataset. A number of measures have been suggested to tackle these research needs. Examples of such measures include Hausdorff distance (e.g., Huttenlocher et al., 1993; Brakatsoulas et al., 2005), Fréchet distance (e.g., Alt and Godau, 1995; Alt et al., 2003), the longest common subsequence (LCS) algorithm (e.g., Agrawal et al., 1995; Chen et al., 2005) and the dynamic time warping (DTW) algorithm (e.g., Sankoff and Kruskal, 1983; Chen et al., 2005; Sakurai et al., 2005). Each of these measures basically examines the similarity between two curves using different algorithms. However, when it comes to measuring similarity of space–time paths, each of them has certain advantages and disadvantages in providing appropriate measures that can support effective investigations of individual activity patterns in an integrated space–time environment. For example, the Hausdorff distance and Fréchet distance only use the distance between point pairs from two curves to compute their similarity. Such point-based measures cannot capture detailed differences between two space–time paths, which contain important activity features such as activity time, duration, sequence and frequency. The longest common subsequence algorithm can analyze activity data on a nominal scale and identify the longest common subsequence of activities (e.g., an activity sequence of home-work-home) on the space–time paths which record individuals' daily activities.

However, other important characteristics of activities (e.g., spatial and temporal locations of a person's activities) are missing in the measure. The dynamic time warping algorithm can identify similar sequences which may vary in starting time and time duration by manipulating the time dimension, but this approach becomes inappropriate when the exact activity time and duration are essential factors in comparing individuals' activity patterns during a specific time period (e.g., a 30-min work trip made during the morning peak hours is deemed different from a 30-min wok trip made in the late morning in many activity studies). Without explicitly considering space and time in an integrated manner, these measures may have some weaknesses of identifying different spatiotemporal patterns among space–time paths.

Some of these measures have been reported in recent studies that compute the similarity among individual space–time paths to compare their patterns. These studies have focused on, for example, measures of selected characteristics of individual space–time paths such as distance between locations (e.g., Sinha and Mark, 2005; Andrienko et al., 2007; González et al., 2008), moving speeds and directions (e.g., Laube et al., 2005), as well as sequences of locations (e.g., Shoval and Isaacson, 2007) to examine whether individual space–time paths are similar to each other. In the meantime, these methods fall short of considering variables such as activity duration or activity time related to space–time constraints that people face in arranging their daily activities. In addition, due to the high computational complexity of these algorithms, their efficiency of computing path similarity measures often is severely hampered as the number of paths increases, which is a common situation in working with large individual-level spatiotemporal datasets. In this study, we propose a simple, yet robust, space–time GIS approach, which can serve as an alternative to those computational intensive algorithms for exploring human activity patterns.

3. Activity Pattern Analyst (APA)

In this study, we develop a space–time GIS to support exploratory analysis of activity data at the individual level and in a space–time context. It is based on a space–time GIS design with temporal dynamic segmentation (Yu, 2006; Shaw and Yu, 2009) and implemented as an extension in ArcScene, which is the 3D viewer of ArcGIS. We call this extension an Activity Patten Analyst (APA). This APA extension includes a previously developed function of generating space–time paths as well as a series of query and analysis functions for investigating hidden activity patterns in an individual-level activity diary dataset.

There are six major components in this APA extension (Fig. 2). The first component is *space–time path generation* which is used to build individual space–time paths from a given individual-level activity dataset. Such an input dataset can be derived from either activity diary data collected from a traditional questionnaire survey or activity data collected with global positioning system (GPS) tracking devices. If the locational information in a dataset is based on street addresses, a geocoding process is needed before the dataset can be used for the path generation function. This component creates a space–time GIS database that serves as the foundation of the other five components in this APA extension. The second component is a *space–time path filter*. It is used to extract subsets of individual space–time paths based on their non-spatial attributes (e.g., gender, age, education, occupation, income) as well as their spatial characteristics such as residential location. This component is implemented with various spatial and non-spatial query functions. A subset of individual space–time paths can be selected by applying a set of attribute and/or spatial selection functions. The third component is *space–time path segmentation* that is based on the concept of temporal dynamic segmentation (Yu, 2006; Shaw and Yu, 2009). Users can specify any time period and this function will dynamically extract both spatial and non-spatial data of the user-specified time period and attach the data to the corresponding segment on individual space–time paths. The above three components allow users to convert activity diary data and organize the data in a space–time GIS such that users can manage, query and visualize the data to interactively explore hidden spatiotemporal patterns of human activities.

The other three components in the APA extension are developed for pattern detection. The fourth component is *activity query*, which dynamically associates activity data of each individual to the corresponding locations on each individual's space–time path. For instance, Fig. 1 shows that an individual participated in a group meeting between 9:00 and 11:30 a.m. at the workplace. The duration of an activity and its location are dynamically mapped from the GIS database onto the individual's space–time path using the temporal dynamic segmentation method. This overcomes a shortcoming of the classical space–time path representation that shows only individual movements in space over time and does not provide explicit spatiotemporal information about the actual actions taken place along the space–time path. By displaying different types of activities in different colours, this activity query component can identify many distinct characteristics of human activities on individual space–time paths and thus facilitate interactive visualization of individual and/or group activity patterns.

The fifth component is *activity distribution/density analysis*, which dynamically retrieves the spatial distribution of all individuals at any user-specified time point. This component is used to facilitate analysis of aggregate pattern of human activities both in space and across time. Based on the point distribution of individual activity locations at a given time, we can apply a kernel density estimation to derive an aggregate activity distribution surface. The search radius used in this analysis is specified by users based on their knowledge of the specific problem domain. By comparing aggregate activity distribution surfaces between different time points, we can explore the change patterns as well as find the locations that gained or lost activities during the time period.

The sixth component is *space–time path clustering analysis*. It provides a space–time path based clustering method that groups space–time paths of similar geometry into the same cluster. This method provides an exploratory analysis approach which can help researchers identify useful activity patterns hidden in a large individual-level activity diary dataset. As it is difficult to compare the similarity of space–time paths represented as 3D features in the space–time system (Fig. 3a), standardized paths are used as the input to the clustering analysis in this study. We first transform each individual space–time path in the 3D space–time GIS into a 2D path in a time–distance plane coordinate system. As shown in Fig. 3b, all paths are adjusted to their residential locations so that they all start from the origin point of the time–distance coordinate

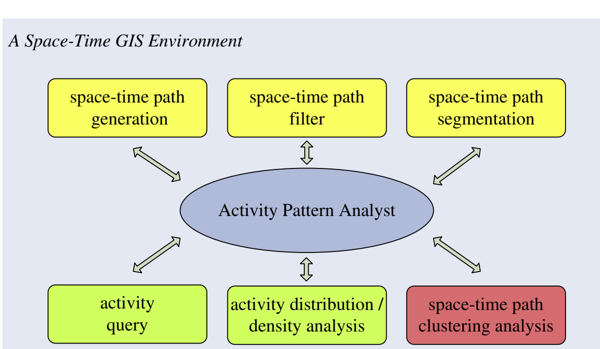


Fig. 2. Components of the Activity Pattern Analyst (APA) extension.

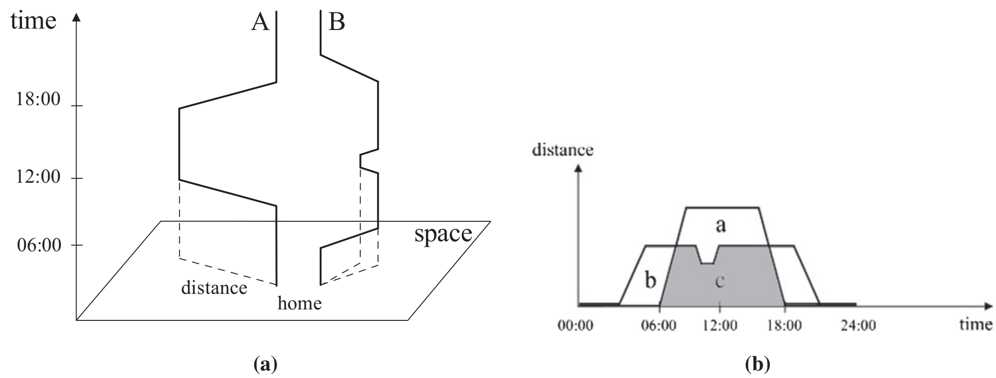


Fig. 3. Space–time path based similarity measure.

system. Therefore, each transformed path tracks how far a person moved away from his/her residential location for various daily activities over the study period. Although this transformation drops absolute activity locations and movement directions in geographic space, the path shape and other important features of individual space–time paths such as activity time, duration, sequence, frequency, and type are preserved. Moreover, as the residential location is the most important reference location for people to arrange their daily activities, the transformed paths in the time–distance coordinate system indicate the spatiotemporal characteristics of how people conduct their activities with reference to their home location.

Based on the transformed paths, we develop an area-based similarity measure to compare the shapes between two paths and compute the overlapping areas between each pair of paths along the time dimension. It is inspired by a similarity theory originated in psychology, which treats objects as feature collections and measures similarity among the objects as a feature-matching process (Tversky, 1977) that can be expressed by the equation below:

$$D = \frac{c}{a + b - c} \tag{1}$$

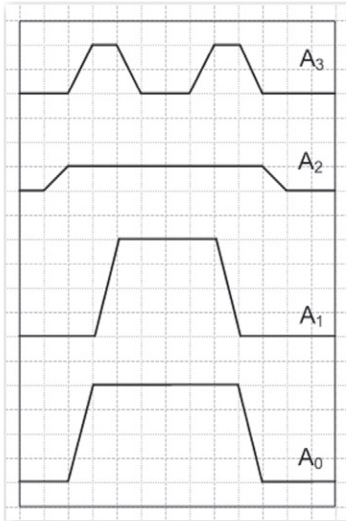
where a is the feature set of object A; b is the feature set of object B; c is the overlapping feature set between A and B; the similarity measure D is the ratio of the overlapping features to the whole features that belong to A and B. We therefore can define the similarity measure D for each pair of paths with the same equation, where a is the entire area between path A and the time axis (i.e., the horizontal axis) in the time–distance plane coordinate system (Fig. 3b); b is the entire area between path B and the time axis in the time–distance plane coordinate system; and c is the overlapping area of a and b. If two space–time paths have an identical shape, the similarity measure $D = 1$ since $a = b = c$ in this case. On the other hand, $D = 0$ if two space–time paths do not share any overlapping area.

Using this definition, the similarity measure between any pair of paths can be computed and a similarity matrix for all pairs of paths can be constructed to perform a cluster analysis. For instance, Fig. 4a shows four individual space–time paths of A_0, A_1, A_2 and A_3, each of which represents a space–time path transformed into the time–distance plane coordinate system. In order to clearly show the shape of each space–time path, we organize them as separate, rather than overlapping, areas in Fig. 4a. Based on the similarity measure D discussed above, we compute a similarity index for each pair of paths and then construct a similarity matrix (Fig. 4b). The larger the index, the more similar in shape the corresponding pair of space–time paths. Different from those point-based computational methods (e.g., Hausdorff distance and Fréchet distance), this area-based path similarity measure takes path shape into account and is able to incorporate the crucial characteristics of individual space–time paths, such as activity time, duration, sequence, frequency as well as the activity distance in reference to the residential location, to investigate spatiotemporal patterns of human activities. In addition, this measure is not computational intensive and easy to implement in a space–time GIS environment, which makes it a good alternative method for large individual-level spatiotemporal datasets.

In this study, we propose a space–time path based multi-level clustering method based on an agglomerative hierarchical clustering method with average linkage (Han and Kamber, 2006). As an exploratory analysis, this method allows users to use their domain knowledge and choose how many levels of clustering analysis to be executed and how many clusters to be generated at each level during the analysis process. In particular, according to the agglomerative hierarchical clustering method, each cluster contains one individual space–time path initially. The index between every two clusters is calculated according to the area-based similarity measure and a similarity matrix is constructed accordingly. Two clusters with the largest index in the similarity matrix are aggregated into a new cluster and the similarity matrix is correspondingly recalculated. This process repeats until it reaches the user specified number of clusters. Due to the definition of this area-based similarity measure, the clustering method is sensitive to outliers in the set of paths and tends to form several small clusters, each of which contains only a few individual space–time paths that have the most unique patterns, and place the remaining individual space–time paths in a large catch-all cluster. Therefore we can easily remove the outliers and repeat this clustering method to the catch-all cluster at each level in order to find sub-clusters embedded in the catch-all cluster at the next level. As the process reaches deeper levels and outliers are removed from the set step by step, clusters of individual space–time paths with distinctive patterns start to emerge. Each cluster usually contains a significant number of paths similar in shape, which exhibit representative activity patterns. One special case that this space–time path based multi-level clustering method must address is when both space–time paths are perfect vertical lines in the 3D space–time GIS (i.e., both individuals stayed at a fixed location throughout the entire study period). In this case, all three terms (i.e., a, b and c) in Eq. (1) have a value of 0. The equation of computing similarity measure D therefore becomes invalid. Our study takes care of this special case by checking all space–time paths and placing such space–time paths into a special cluster.

4. A case study of activity diary data collected in Beijing, China

Geovisualization is a powerful tool for exploratory data analysis of large spatiotemporal datasets. Researchers often use graphical representations to display intricate patterns in large datasets in



(a)

	A_0	A_1	A_2	A_3
A_0		0.714	0.265	0.286
A_1			0.247	0.098
A_2				0.417
A_3				

(b)

Fig. 4. Space–time path based similarity matrix.

ways that stimulate visual thinking and suggest possible hypotheses for further analysis (Hearnshaw and Unwin, 1994). This section presents a case study using activity diary data collected in Beijing, China to demonstrate the effectiveness of uncovering hidden spatiotemporal activity patterns with the functions developed in the Activity Pattern Analyst extension.

4.1. Data

Our research collaborators at Peking University in China conducted an Urban Residents Activity Survey in the Beijing metropolitan area in October and November of 2007. They used a stratified sampling strategy of "residential area – household – individual" to select ten residential areas with different socioeconomic/demographic characteristics in the Beijing metropolitan area. Within each selected residential area, 60 households were chosen based on a systematic sampling design. This survey collected a 48-h retrospective activity diary (Sunday and the next Monday) with face-to-face interviews to record all activities performed by individuals of at least 16 years old in the sampled households. Among the 600 participating households in this survey, 520 households (with a total of 1107 individuals) returned complete and valid responses. The dataset consists of a total of 19691 stationary activities (i.e., excluding travel) that are divided into five categories – physiological needs (e.g., sleep, have meals), household duties, work/school, leisure, and all other activities. This activity diary survey also collected data regarding the location and time associated with each activity, along with socioeconomic/demographic characteristics of all respondents.

4.2. Activity patterns in space and time

Visualizing spatiotemporal activity patterns of individuals with different socioeconomic/demographic characteristics can help us better understand the needs and constraints of different population groups. This study first uses the "space–time path generation" function to build individual space–time paths from the activity diary dataset. It then applies the "space–time path segmentation" and the "space–time path filter" functions to extract subsets of the data based on different hypotheses assumed with the different neighbourhoods and population groups in Beijing. Followed by the "activity query" function to associate activities with individual space–time paths, it allows us to interactively explore and investigate spatiotemporal activity patterns and their interactions with the built environment.

Fig. 5 shows an example of visualizing 1107 individual space–time paths with 9701 colour-coded activities on a Monday. It shows a general rhythm of human activities on a week day in Beijing, China. Most people get up in the morning after a sleep (line segments in purple colour[1] in the lower part of space–time paths) and then go to work/school (line segments in red colour). After about a 1-h lunch break (short line segments in purple colour in the middle part of space–time paths), they continue to work/attend school until they leave their workplace/school in late afternoon. After that, activities become diversified. Some people stay at work while others carry out household work (line segments in gold colour) or leisure activities (line segments in green colour).

Fig. 5 also shows variations in activity patterns among individuals. It therefore is important to be able to interactively select subgroups of people with different demographic, socioeconomic, and/or spatial characteristics to better understand what might be causing these variations. Fig. 6 illustrates the activity patterns of residents in two different neighbourhoods selected from the dataset presented in Fig. 5. Neighbourhood A (Fig. 6a) consists of residents who are mainly government employees, while residents in Neighbourhood B (Fig. 6b) are mainly low- or middle-income people. We can clearly see the differences between these two neighbourhoods. Residents in Neighbourhood A have a regular activity rhythm of 8-h work during the day and leisure activities in the evening. Residents in Neighbourhood B exhibit a more complex activity pattern with evening work hours. With the interactive visual exploratory analysis functions available in the APA extension, researchers can investigate activity patterns based on socioeconomic/demographic characteristics of different population groups or various space–time characteristics such as location, time, duration, sequence, frequency, and type of activity. These exploratory visual analysis functions developed from Hägerstrand's time-geographic concepts offer useful complements to conventional statistical analysis methods.

[1] For interpretation of color in Figs. 1,2, and 5–10, the reader is referred to the web version of this article.

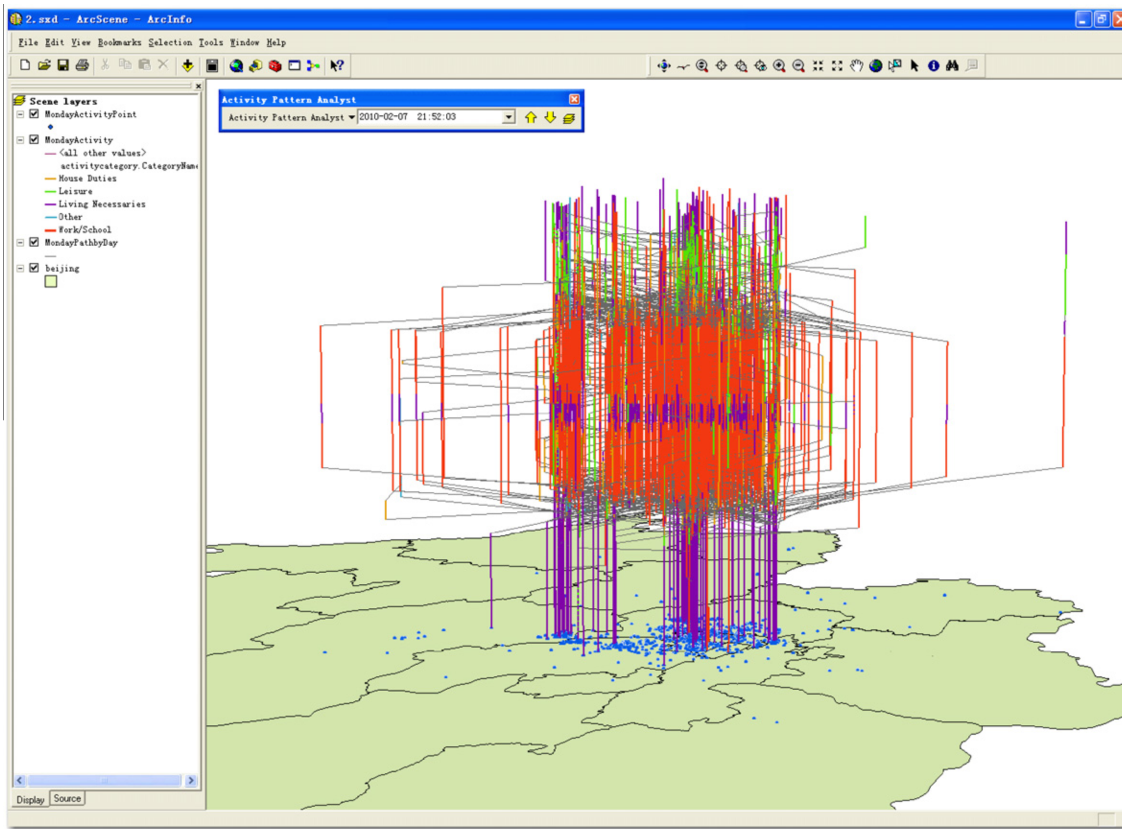
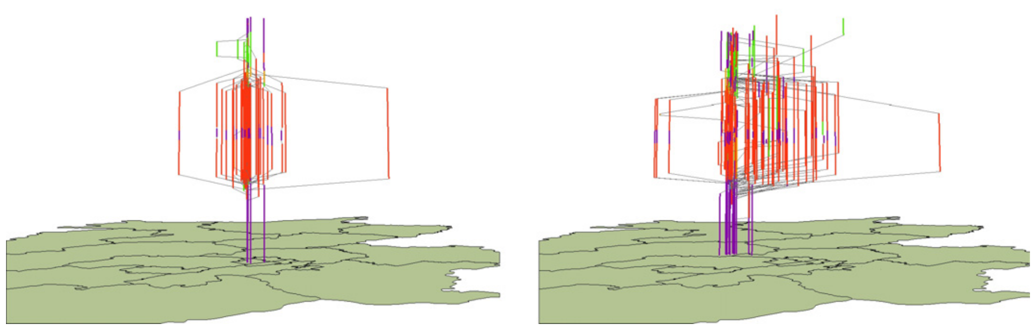


Fig. 5. Visualization of activity types in space and time.



(a) Neighbourhood A **(b) Neighbourhood B**

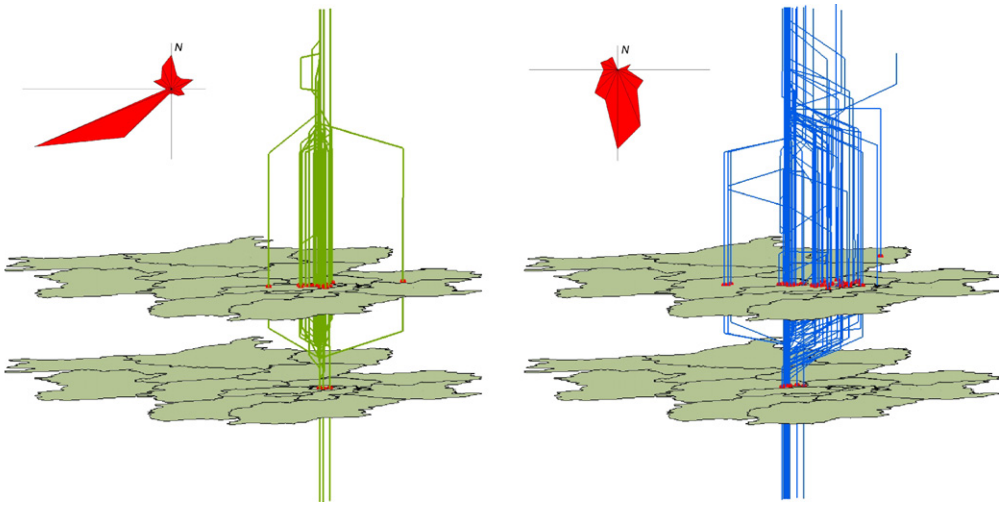
Fig. 6. Spatiotemporal distributions of activity types for Neighbourhood A and Neighbourhood B.

4.3. Activity distribution and density pattern

Spatial distributions of individual activities over time can be used to study aggregate patterns of human activities. The "activity distribution/density" function in APA is developed for exploring activity distribution and density patterns across space and time. This function dynamically interpolates locations of individual activities at any user-specified time point as we move the GIS layers along the space–time paths (Fig. 7). We therefore can compare individual activity locations at different time points to assess changing activity distribution patterns over time.

Fig. 7a shows the activity distribution of 93 individuals in Neighbourhood A and Fig. 7b illustrates the distribution pattern of 91 individuals in Neighbourhood B. These two neighbourhoods have roughly the same number of sampled individuals; however, they have very different spatial distributions with respect to travel directions and distances on a work day. As an example, we use the "activity distribution" function to derive activity locations at 5:30 a.m. and 10:30 a.m. respectively (see red dots where individual space–time paths intersect with the base GIS map layer in Fig. 7a and b). A comparison of these two distribution patterns clearly indicates the residential locations versus daytime activity locations. Furthermore, we use rose diagrams in Fig. 7a and b to illustrate directional patterns of home-to-work/school trips in Neighbourhoods A and B, respectively. In these rose diagrams, directions are equally divided into 16 classes. The value of each class represents the total number of individuals whose home-to-work direction matches with the directional range of that class. Users of the APA extension can conduct analyses based on any demographic/socioeconomic characteristics of individuals and/or attributes of residential neighbourhoods and workplaces to gain additional insights of the hidden activity patterns.

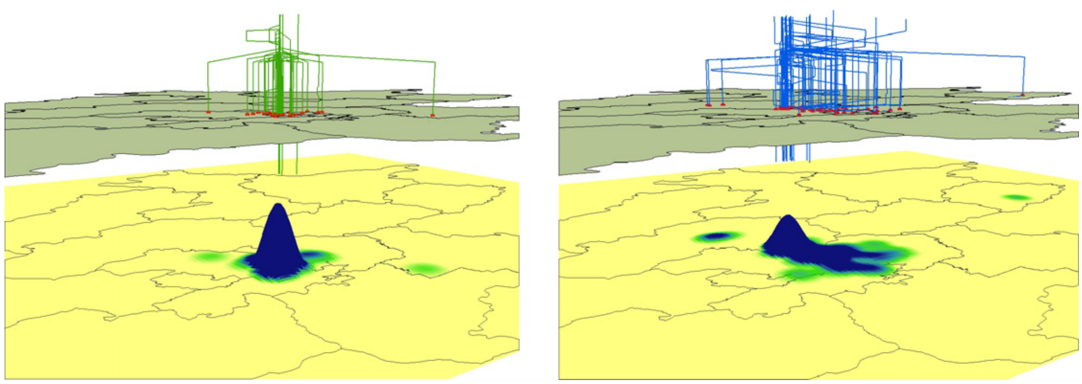
In addition to representation of activity distributions as a collection of points, we can use kernel density surfaces to show aggre-



(a) Neighbourhood A (b) Neighbourhood B

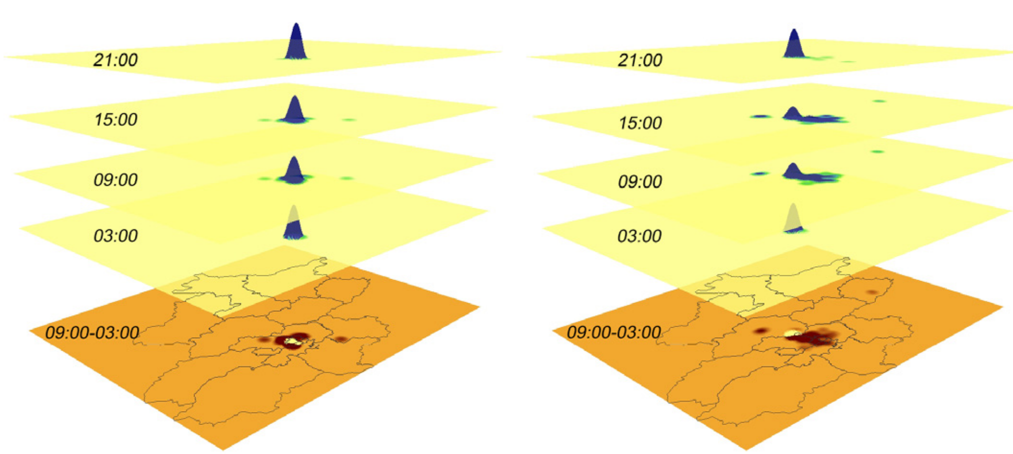
Fig. 7. Neighbourhood A and Neighbourhood B shown as a side view.

gate activity distribution patterns across space and time. For example, the same two neighbourhoods presented in Fig. 7 are used to generate density surfaces at 9:00 a.m. using the "activity distribution" function in the APA extension. Fig. 8a indicates a high concentration of activities while Fig. 8b shows a more spread pattern. When the spatial distributions of human activities are displayed continuously in a time sequence such as every hour, we can generate an animated display of changing aggregate activity distri-



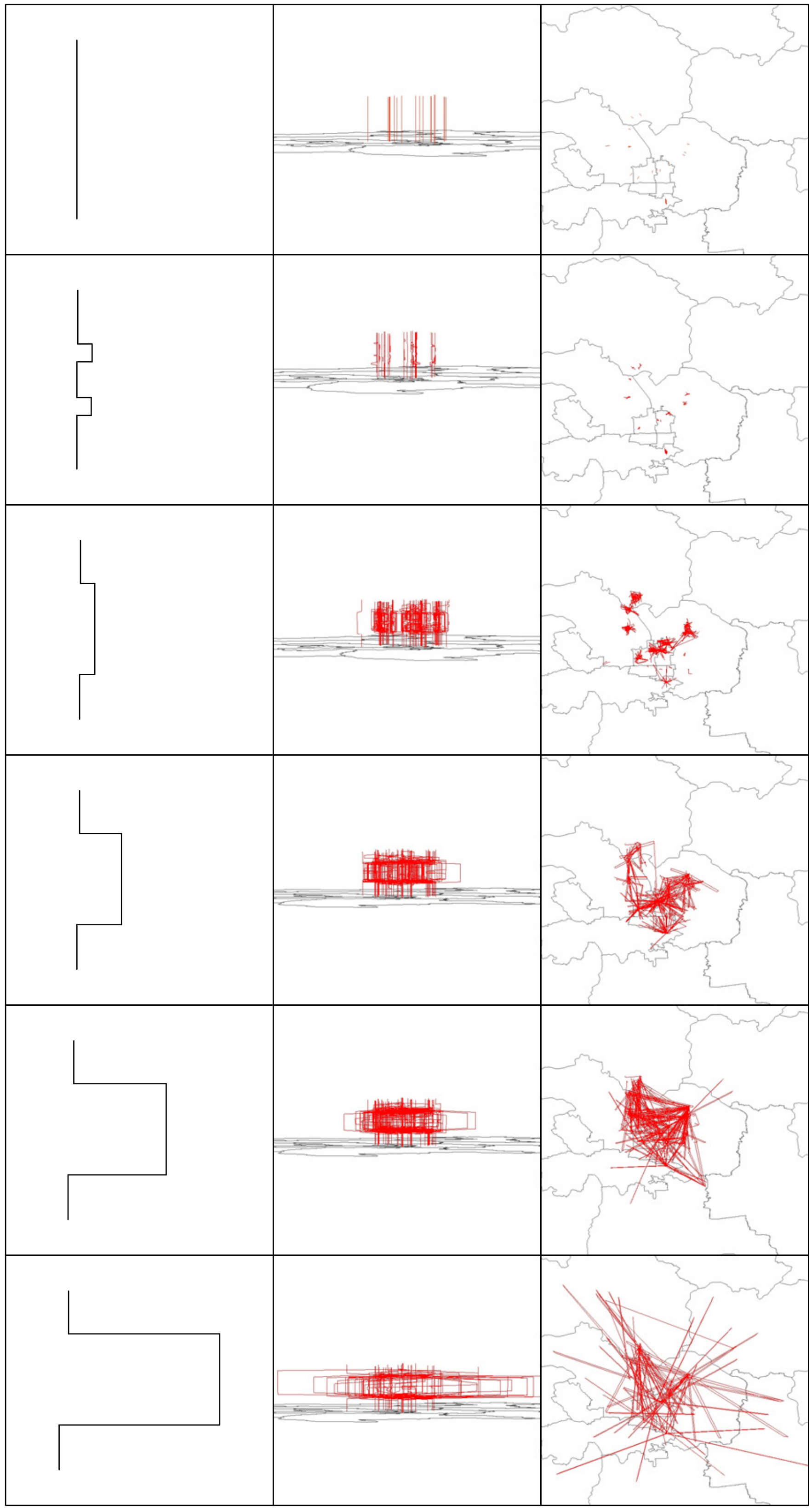
(a) Neighbourhood A (b) Neighbourhood B

Fig. 8. Activity density patterns of Neighbourhood A and Neighbourhood B.



(a) Neighbourhood A (b) Neighbourhood B

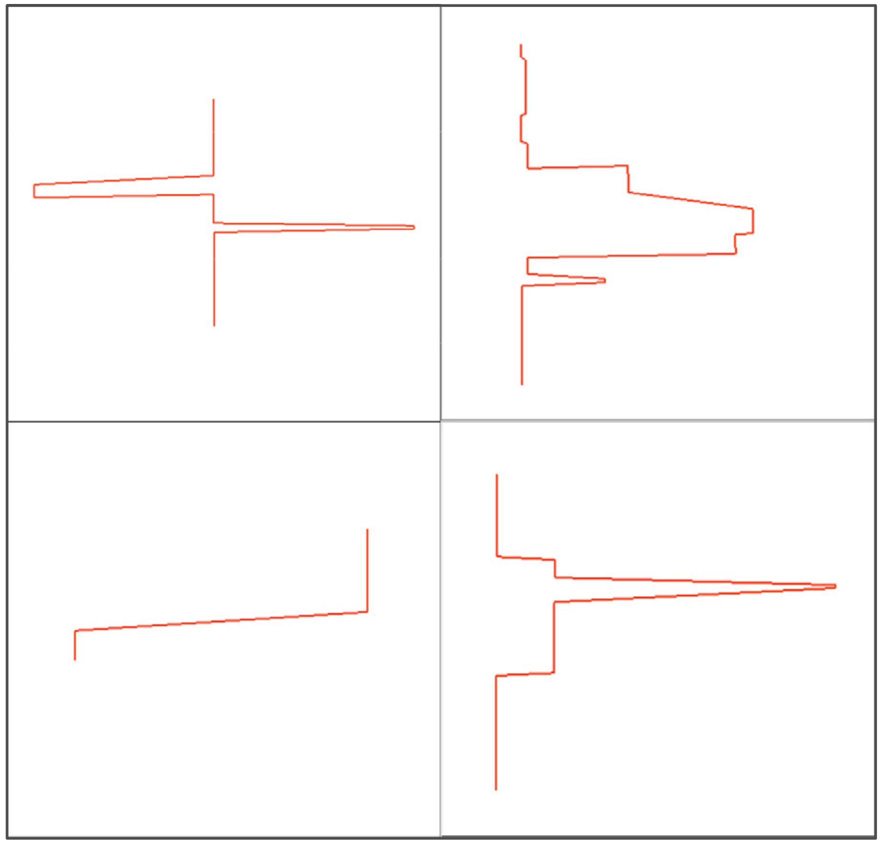
Fig. 9. Animated activity density patterns of Neighbourhood A and Neighbourhood B.



(a) Clustered individual space-time paths

Fig. 10. Space–time path based clustering patterns.

bution patterns across space and time (Fig. 9). This animated display function is useful for researchers to visualize complex activity patterns hidden in the spatiotemporal dataset. Moreover, since the GIS density surface is a raster layer and each cell in the layer has a



(b) Examples of outlier individual space-time paths

Fig. 10 (*continued*)

density value, we can use raster GIS algebraic operations to compare the density surfaces. For example, we could subtract the density surface at 9:00 a.m. from the density surface at 3:00 a.m. in Fis. 9a and b. The output raster GIS layer indicates which locations gained/lost activities between 3:00 a.m. and 9:00 a.m.

4.4. Space–time path clustering pattern

Fig. 10 demonstrates the "space–time path clustering analysis" function. For example, we apply the space–time path based multi-level clustering analysis to derive six clusters (i.e. six rows in Fig. 10a) as well as several outliers (e.g., four individual space–time paths in Fig. 10b) based on the data presented in Fig. 5. In Fig. 10a, the middle column shows the result of each of the six clusters as a side view. In order to illustrate the spatiotemporal pattern of each cluster, a representative sketch is provided in the left column to show the distinct characteristic of each cluster. The right column displays each cluster as a bird's eye view. The clustering analysis results clearly indicate that we can identify similarity within each cluster and differences among different clusters. The first cluster represents a group of individuals who stayed at home on the survey day. This is the special case discussed in the previous section. The second cluster shows a population group who took multiple short trips on the survey day. The remaining four clusters suggest four other population groups who all had a regular work schedule but travelled different distances to their workplaces. In the meantime, Fig. 10b shows some examples of individual space–time paths that possess unique activity/travel patterns, compared to those represented in Fig. 10a. They are examples of the outliers captured by our space–time path based multi-level clustering method. The upper-left one represents an individual who took long and fast trips with short stays at the destinations on the survey day. The upper-right one represents an individual who was busy travelling from one place to another place on the survey day. The lower-left path represents an individual who travelled to another place without coming back home at night. The lower-right path represents an individual who went to office first and then made a trip during the work hours. The above examples illustrate that the space–time path based multi-level clustering method can help researchers uncover hidden spatiotemporal activity/travel patterns within any user-specified time period and identify relevant characteristics (e.g., commuting distance or trip timing) that distinguish the clusters. Furthermore, this method can effectively and efficiently identify outlier individuals whose space–time paths exhibit peculiar activity/travel patterns.

We also can examine the socioeconomic/demographic characteristics of individuals in each of the six clusters in Fig. 10a to gain further insights into these clusters. For instance, the first cluster consists of mainly grandparents who help take care of grandchildren at home or people who work at home. The second cluster reflects mainly housewives who make multiple short trips to neighbourhood markets, drop off/pick up school children, or visit friends in nearby neighbourhoods. The other four clusters indicate people who have regular work hours but different commuting distances. The patterns derived from the space–time path based multi-level clustering method suggest that most people in Beijing have a relatively regular spatiotemporal activity/travel pattern on a weekday. Moreover, the spatiotemporal exploratory analysis functions in the APA extension allow us to examine how these patterns differ among different groups of people and the distinct characteristics associated with these different patterns and different population groups.

5. Conclusion

This study develops an ArcGIS extension of Activity Pattern Analyst that operationalizes Hägerstrand's time-geographic concepts with a set of spatiotemporal exploratory analysis functions for examining individual activity/travel patterns. With an integrated space–time system of representing individual activities, this extension offers useful functions for transportation researchers to investigate people's everyday lives as a process over time rather than aggregate snapshots. This paper illustrates how we can organize individual activity/travel diary data into a space–time GIS and apply the APA functions to uncover various interesting spatiotemporal patterns hidden in an activity diary survey dataset collected in Beijing, China in 2007. As demonstrated in the case study, the revealed patterns can help researchers gain insights of the activity diary dataset and suggest possible hypotheses of explaining the observed patterns. Such a space–time GIS implementation can be a very useful complement to the conventional statistical and mathematical approaches of developing activity-based travel demand models. In addition, the APA extension presented in this paper is applicable to many other studies involving spatiotemporal tracking data at the individual level such as GPS tracking data of vehicles (i.e., automobiles, trains, airplanes, ships, etc.), mobile phone tracking data of individual movements, or radio frequency identification (RFID) tracking data of freight shipments.

A space–time path based multi-level clustering method is presented and implemented in this study to explore hidden patterns in large individual-level spatiotemporal datasets. A specific index is developed to measure the similarity of two space–time paths after they are transformed into a 2D time–distance plane coordinate system. As the paths possess a continuous representation of time, this index incorporates the process of activity location changes in its structure. Using the standardized paths which are adjusted to their residential locations, we are able to use this index to capture the differences among people in a Beijing survey dataset regarding their spatial and temporal arrangements of daily activities. Based on this similarity index, the proposed space–time path based multi-level clustering method can effectively identify the paths that share similar activity patterns as well as the outlier paths that possess unique activity patterns, in terms of where, when, and how long the activities occur. The case study presented in this paper demonstrates the effectiveness of this space–time path based clustering analysis method in revealing interesting activity patterns hidden in an activity diary dataset.

The space–time GIS approach offers an integrated space–time analysis environment that can effectively represent and organize individual-level activity dairy data in the form of space–time paths. It opens up many new opportunities for researchers to analyze the data and investigate spatiotemporal patterns of human activities. With these opportunities, we also face challenges of developing effective and useful analysis functions that can further our understanding of human activities in a space–time context. As an initial attempt to perform clustering analysis on space–time paths, this study presents one similarity measure only and applies it to the transformed paths for clustering analysis. We plan to enhance this Activity Pattern Analyst extension in at least two directions. First of all, we will develop different similarity measures with each measure aiming at distinguishing particular types of cluster patterns effectively and efficiently (e.g., activity frequency, sequence, timing, location choice, etc.). Second, although this study demonstrates that we can analyze space–time paths in an integrated space–time GIS, it remains a major research challenge of developing a robust time object that can interact with the spatial objects dynamically in a temporal GIS. Progress in these directions will provide more powerful data management, analysis and visualization functions to help transportation researchers study human activity and interaction patterns in a space–time context.

Acknowledgements

This research is supported by US National Science Foundation (BCS-0616724) and the National Natural Science Foundation of China (40871184, 40671058).

References

Agrawal, R., Lin, K.I., Sawhney, H.S., Shim, K., 1995. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In: Proceedings of the 21st International Conference on Very Data Bases, Zürich, Switzerland, pp. 490–501.

Alt, H., Efrat, A., Rote, G., Wenk, C., 2003. Matching planar maps. Journal of Algorithms 49 (2), 262–283.

Alt, H., Godau, M., 1995. Computing the Fréchet distance between two polygonal curves. International Journal of Computational Geometry and Applications 5 (1–2), 75–91.

Anderson, J., 1971. Space–time budgets and activity studies in urban geography and planning. Environment and Planning 3 (4), 353–368.

Andrienko, G., Andrienko, N., Wrobel, S., 2007. Visual analytics tools for analysis of movement data. ACM SIGKDD Explorations Newsletter 9 (2), 38–46.

Bhat, C.R., Koppelman, F.S., 1999. A retrospective and prospective use of time-use research. Transportation 26, 119–139.

Brakatsoulas, S., Pfoser, D., Salas, R., Wenk, C., 2005. On map-matching vehicle tracking data. In: Proceedings of the 31st International Conference on Very Large Data Bases, Trondheim, Norway, pp. 853–864.

Buliung, R., Kanaroglou, P., 2004. On design and implementation of an object-relational spatial database for activity/travel behaviour research. Journal of Geographical Systems 6, 237–262.

Buliung, R., Kanaroglou, P., 2006. A GIS toolkit for exploring geographies of household activity/travel behaviour. Journal of Transport Geography 14 (1), 35–51.

Chai, Y., Liu, Z., Li, Z., et al., 2002. Time–Space Structure of Chinese Cities. Peking University Press, Beijing (in Chinese).

Chapin, F.S., 1974. Human Activity Patterns in the City: Things People Do in Time and in Space. John Wiley and Sons, New York.

Chen, L., Özsu, M.T., Oria, V., 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, pp. 491–502.

Cullen, I., Godson, V., Major, S., 1972. The structure of activity patterns. In: Wilson, A.G. (Ed.), Patterns and Processes in Urban and Regional Systems. Pion, London, pp. 281–296.

Ellegåd, K., 1999. A time-geographic approach to the study of everyday life of individuals – a challenge of complexity. GeoJournal 48, 167–175.

Forer, P., 1998. Geometric approaches to the nexus of time, space, and microprocess: implementing a practical model for mundane socio-spatial systems. In: Edenhofer, M.J., Golledge, R.G. (Eds.), Spatial and Temporal Reasoning in Geographic Information Systems. Oxford University Press, Oxford, England, pp. 171–190.

Frihida, A., Marceau, D., Thériault, M., 2002. Spatio-temporal object-oriented data model for disaggregate travel behavior. Transactions in GIS 6 (3), 277–294.

Gahegan, M., 2000. The case for inductive and visual techniques in the analysis of spatial data. Journal of Geographical Systems 2, 77–83.

González, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. Nature 453 (5), 779–782.

Golob, T.F., McNally, M.G., 1997. A model of activity participation and travel interactions between household heads. Transportation Research B 3 (3), 177–194.

Goulias, K.G., 2002. Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects. Transportation 29 (1), 31–48.

Gren, M., 2001. Time–geography matters. In: May, J., Thrift, N. (Eds.), Timespace. Geographies of Temporality. Routledge, London/New York, pp. 208–225.

Guo, D., Gahegan, M., MacEachren, A.M., Zhou, B., 2005. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. Cartography and Geographic Information Science 32, 113–132.

Hägerstrand, T., 1970. What about people in regional science? Papers of the Regional Science Association 24 (1), 7–21.

Hägerstrand, T., 1978. Survival and arena. In: Carlstein, T., Parkes, D., Thrift, N. (Eds.), Timing space and spacing time, Human Activity and Time Geography, vol. 2. John Wiley & Sons, New York, pp. 122–145.

Hägerstrand, T., 1989. Reflections on what about people in regional science? Papers of the Regional Science Association 66, 1–6.

Han, J.W., Kamber, M., 2006. Data Mining: Concepts and Techniques, second ed. Elsevier, Oxford, UK.

Hearnshaw, H.M., Unwin, D.J., 1994. Visualization in Geographic Information System. John Wiley and Sons, London.

Huisman, O., 2006. The Application of Time-Geographic Concepts to Urban Micro-process. PhD Thesis, School of Geography and Environmental Science, University of Auckland, New Zealand.

Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.A., 1993. Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (9), 850–863.

Jim, C.Y., Chen, W.Y., 2009. Leisure participation pattern of residents in a new Chinese city. Annals of the Association of American Geographers 99 (4), 657–673.

Joyce, M., Stewart, J., 1999. What can we learn from time-use data? Monthly Labor Review 122, 3–6.

Kang, H., Scott, D.M., 2008. An integrated spatio-temporal GIS toolkit for exploring intra-household interactions. Transportation 35, 253–268.

Kang, H., Scott, D.M., Doherty, S.T., 2009. Investigation of planning priority of joint activities in household activity-scheduling process. Transportation Research Record: Journal of the Transportation Research Board 2134, 82–88.

Kitamura, R., 1988. An evaluation of activity-based travel analysis. Transportation 15, 9–34.

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.P., Tsang, A.M., Switzer, P., Behar, J.V., Hern, S.C., Engelmann, W.H., 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. Journal of Exposure Analysis and Environmental Epidemiology 11 (3), 231–252.

Kraak, M.J., Huisman, O., 2009. Beyond exploratory visualization of space–time paths. In: Miller, H.J., Han, J.W. (Eds.), Geographic Data Mining and Knowledge Discovery, second ed. CRC Press, Boca Raton/London/New York, pp. 431–443.

Kwan, M.-P., 1998. Space–time and integral measures of individual accessibility: a comparative analysis using a point-based framework. Geographic Analysis 30 (3), 191–216.

Kwan, M.-P., Hong, X.D., 1998. Network-based constraints-oriented choice set formation using GIS. Geographical Systems 5, 139–162.

Kwan, M.-P., 1999a. Gender, the home-work link, and space–time patterns of non-employment activities. Economic Geography 74 (4), 370–394.

Kwan, M.-P., 1999b. Gender and individual access to urban opportunities: a study using space–time measure. The Professional Geographer 51 (2), 210–227.

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, 185–203.

Kwan, M.-P., 2002. Time, information technologies, and the geographies of everyday life. Urban Geography 23 (5), 471–482.

Kwan, M.-P., Lee, J., 2003. Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. In: Goodchild, M.F., Janelle, D.G. (Eds.), Spatially Integrated Social Science. Examples in Best Practice. Oxford University Press, Oxford, pp. 48–66.

Laube, P., Imfeld, S., Wiebel, R., 2005. Discovering relative motion patterns in groups of moving point objects. International Journal of Geographic Information Science 19, 639–668.

Lu, X., Pas, E.I., 1999. Socio-demographics, activity participation and travel behaviour. Transportation Research A 33, 1–18.

Michelson, W., 1973. Time budgets in environmental research: some introductory considerations. In: Preiser, W.F.E. (Ed.), Environment Design Research, Symposia and Workshops, 4th International EDRA Conference, vol. 11. Dowden, Hutchinson, Ross, Inc., Stroudsburg, PA.

Miller, H.J., 1991. Modelling accessibility using space–time prism concepts within geographic information systems. International Journal of Geographical Information Systems 5 (3), 287–301.

Miller, H.J., 1999. Measuring space–time accessibility benefits within transportation networks: basic theory and computational procedures. Geographical Analysis 31 (2), 187–212.

Miller, H.J., 2004. Activities in space and time. In: Hensher, D., Button, K., Haynes, K., Stopher, P. (Eds.), Handbook of Transport, Transport Geography and Spatial Systems, vol. 5. Elsevier Science, Amsterdam, pp. 647–660.

Miller, H.J., Bridwell, S.A., 2009. A field-based theory for time geography. Annals of the Association of American Geographers 99, 149–175.

Neutens, T., Schwanen, T., Witlox, F., De Maeyer, P., 2008a. My space or your space? Towards a measure of joint accessibility. Computers, Environment and Urban Systems 32, 331–342.

Neutens, T., Van De Weghe, N., Witlox, F., De Maeyer, P., 2008b. A three-dimensional network-based space–time prism. Journal of Geographical Systems 10, 89–107.

Neutens, T., Witlox, F., De Weghe, N.V., De Maeyer, P., 2007. Space–time opportunities for multiple agents: a constraint-based approach. International Journal of Geographical Information Science 21, 1061–1076.

Pendyala, R.M., Goulias, K.G., 2002. Time use and activity perspectives in travel behaviour research. Transportation 29 (1), 31–48.

Pipkin, J., 1995. Disaggregate models of travel behaviour. In: Hanson, S. (Ed.), The Geography of Urban Transportation, second ed. The Guilford Press, New York, pp. 188–218.

Pred, A., 1977. The choreography of existence. comments on Hägerstrand's time–geography and its usefulness. Economic Geography 53 (2), 207–221.

Ren, F., Kwan, M.-P., 2009. The impact of the internet on human activity-travel patterns: analysis of gender differences using multi-group structural equation models. Journal of Transport Geography 17, 440–450.

Robinson, J.P., Godbey, G., 1997. Time for Life: The Surprising Ways Americans Use Their Time. Pennsylvania State University Press, State College, PA.

Sakurai, Y., Yoshikawa, M., Faloutsos, C., 2005. FTW: fast similarity search under the time warping distance. In: Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Baltimore, Maryland, pp. 326–337.

Sankoff, D., Kruskal, J., 1983. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Addison-Wesley, MA.

Scott, D.M., 2006. Embracing activity analysis in transport geography: merits, challenges and research frontiers. Journal of Transport Geography 14, 389–392.

Shaw, S.-L., Wang, D., 2000. Handling disaggregate spatiotemporal travel data in GIS. GeoInformatica 4 (2), 161–178.

Shaw, S.-L., Yu, H., 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. Journal of Transport Geography 17, 141–149.

Shaw, S.-L., Yu, H., Bombom, L.S., 2008. A space–time GIS approach to exploring large individual-based spatiotemporal datasets. Transactions in GIS 12 (4), 425–441.

Shapcott, M., Steadman, P., 1978. Rhythms of urban activity. In: Carlstein, T., Button, K., Parkes, D., Thrift, N. (Eds.), Timing Space and Spacing Time, Human Activity and Time Geography, vol. 2. John Wiley & Sons, New York, pp. 49–74.

Shoval, N., Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. Annals of the Association of American Geographers 97, 282–297.

Sinha, G., Mark, D.M., 2005. Measuring similarity between geospatial lifelines in studies of environmental health. Journal of Geographical Systems 7, 115–136.

Stinson, L.L., 1999. Measuring how people spend their time: a time-use survey design. Monthly Labor Review 122, 12–19.

Szalai, A., 1972. The Use of Time. Mouton, The Hague, Netherlands.

Tversky, A., 1977. Features of similarity. Psychologica Review 84, 327–352.

Vrotsou, K., Ellegåd, K., Cooper, M., 2007. Everyday life discoveries: mining and visualizing activity patterns in social science diary data. In: 11th International Conference Information Visualization (IV'07). IEEE Computer Society.

Wang, D., Cheng, T., 2001. A spatio-temporal data model for activity-based transport demand modelling. International Journal of Geographical Information Science 15 (6), 561–585.

Weber, J., 2003. Individual accessibility and distance from major employment centers: an examination using space–time measures. Journal of Geographical Systems 5, 51–70.

Weber, J., Kwan, M.P., 2002. Bringing time back in: a study on the influence of travel time variations and facility opening hours on individual accessibility. The Professional Geographer 54, 226–240.

Yu, H., 2005. Temporal GIS Design of an Extended Time-Geographic Framework for Physical and Virtual Activities. PhD Thesis, Department of Geography, The University of Tennessee, Knoxville.

Yu, H., 2006. Spatio-temporal GIS design for exploring interactions of human activities. Cartography and Geographic Information Science 33 (1), 3–19.

Yu, H., 2007. Visualizing and analyzing activities in an integrated space–time environment: temporal GIS design and implementation. Transportation Research Record: Journal of the Transportation Research Board 2024, 54–62.

Yu, H., Shaw, S.-L., 2007. Revisiting Hägerstrand's time-geographic framework for individual activities in the age of instant access. In: Miller, H.J. (Ed.), Societies and Cities in the Age of Instant Access. Springer, Dordrecht, Netherlands, pp. 103–118.

Yu, H., Shaw, S.-L., 2008. Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. International Journal of Geographical Information Science 22, 409–430.