Chin. Geogra. Sci. 2021 Vol. 31 No. 1 pp. 1−13
https://doi.org/10.1007/s11769-021-1173-0

Springer Science Press
www.springerlink.com/content/1002-0063

Applying the Hidden Markov Model to Analyze Urban Mobility Patterns: An Interdisciplinary Approach

LOO Becky P Y[1], ZHANG Feiyang[1], HSIAO Janet H[2], CHAN Antoni B[3], LAN Hui[3]

(1. *Department of Geography*, *the University of Hong Kong*, *Hong Kong* 999077, *China*; 2. *Department of Psychology*, *the University of Hong Kong*, *Hong Kong* 999077, *China*; 3. *Department of Computer Science*, *the City University of Hong Kong*, *Hong Kong* 999077, *China*)

Abstract: With the emergence of the Internet of Things (IoT), there has been a proliferation of urban studies using big data. Yet, another type of urban research innovations that involve interdisciplinary thinking and methods remains underdeveloped. This paper represents an attempt to adopt a Hidden Markov Model (HMM) toolbox developed in Computer Science for the analysis of eye movement patterns in Psychology to answer urban mobility questions in Geography. The main idea is that both people's eye movements and travel behavior follow the stop-travel-stop pattern, which can be summarized using HMM. Methodological challenges were addressed by adjusting the HMM to analyze territory-wide travel survey data in Hong Kong, China. By using the adjusted toolbox to identify the activity-travel patterns of working adults in Hong Kong, two distinctive groups of balanced (38.4%) and work-oriented (61.6%) lifestyles were identified. With some notable exceptions, working adults living in the urban core were having a more work-oriented lifestyle. Those with a balanced lifestyle were having a relatively compact zone of non-work activities around their homes but a relatively long commuting distance. Furthermore, working females tend to spend more time at home than their counterparts, regardless of their marital status and lifestyle. Overall, this interdisciplinary research demonstrates an attempt to integrate spatial, temporal, and sequential information for understanding people's behavior in urban mobility research.

Keywords: activity-travel pattern; urban mobility; activity sequences; cluster analysis; Hidden Markov Model

Citation: LOO Becky P Y, ZHANG Feiyang, HSIAO Janet H, CHAN Antoni B, LAN Hui, 2021. Applying the Hidden Markov Model to Analyze Urban Mobility Patterns: An Interdisciplinary Approach. *Chinese Geographical Science*, 31(1): 1−13. https://doi.org/10.1007/s11769-021-1173-0

1 Introduction

The 21st Century has witnessed a proliferation of urban study innovations based on geospatial data, which are becoming both more accurate and more readily available. The fact that daily activities in an 'e-society' are captured by the Internet of Things (IoT), ranging from transit smart cards, location-based sensors to smartphones, suggest that 'digital traces' of the urban population are generated continuously (Loo and Tang, 2018). The use of cloud computing, together with the rapidly declining cost of computer storage and capabilities, means that these 'digital traces' are stored rather than discarded shortly due to capacity and cost concerns. Companies, such as public transport companies, Amazon.com, Google and Netflix, and governments are realizing that IoT data are very valuable for commercial, policy and planning purposes. The demand for open data in different parts of the world has further triggered enormous opportunities for researchers to examine big data at the disaggregate level. This group of research innovations are primarily driven by new types and huge

Received date: 2020-05-29; accepted date: 2020-09-22
Corresponding author: LOO Becky P Y. E-mail: bpyloo@hku.hk

volumes of IoT data. In relation, they are strongly tied to the collection and availability of big data, which are highly uneven across the world and among different cities within the same country (e.g.Becker et al., 2011; Long et al., 2012; Ferreira et al., 2013; Loo and Wang, 2017; Zhang et al., 2017; Zhou et al., 2018; Xu et al., 2019; Zhao and Hu, 2019; Li et al., 2020; Xu et al., 2020).

However, another group of urban study innovations in the 21st Century remains underdeveloped. They relate not so much to new types and huge volumes of IoT data, but to innovative ideas and methods, often from other academic disciplines, to take a fresh look at existing urban problems that are becoming more acute in modern society. While the value of interdisciplinary research has been increasingly recognized (Lawrence, 2004; Sallis et al., 2004; Pugh et al., 2012; European Commission, 2019; Australian Research Council, 2020; National Science Foundation, 2020), it requires much intellectual effort to ensure that the creativity is indeed meaningful and appropriate. Despite the difficulties, the promise of major breakthroughs can be huge, especially for relatively mature fields or disciplines where methodological advancements along the traditional pathways are getting more difficult in nature, and fewer in number. In this study, we chose Hong Kong, China as a case study as detailed household travel survey data are available and have been analyzed using different conventional methods (e.g., Lam et al., 2013; Loo and Lam, 2013). By investigating the urban mobility pattern from a new perspective, meaningful lessons about the urban spatial structure and travel behavior can be gained. The method can be applied to any city with disaggregate travel survey data. This paper aims to encourage more open interdisciplinary thinking in addressing topics which have be examined for a long time using conventional methods. Through sharing this process, other interdisciplinary teams may learn about the advantages and expected challenges.

2 Urban Mobility Research

In urban mobility research, a fundamental question is how people allocate their daily activities across space and time. As the urban population continues to grow, it is crucial for local governments and transport service providers to understand the dynamics behind in order to better manage and accommodate people's travel demand. Moreover, different data sources and data analytics methods have been used to analyze key aspects of transport inequity, such as gender, age and low-income families (Hanson and Hanson, 1980; Offer and Schneider, 2011; Loo and Lam, 2013). In particular, there is a vast literature about travel behavior research following the conceptual framework of Hagerstrand's time geography (Hagerstrand, 1970; Hanson and Hanson, 1980; Miller, 1991; Kwan, 1998; Buliung and Kanaroglou, 2006; Buliung and Remmel, 2008; Shaw et al., 2008; Wilson, 2008; Lam et al., 2013; Choi et al., 2014; Babb et al., 2017; Fransen et al., 2018). Researchers have used tools such as the activity space and space-time prisms (Buliung and Kanaroglou, 2006) to visualize mobility patterns. More complex mathematical functions have also been developed. González et al. (2008) analyzed the trajectories of 100 000 mobile phone users, and came up with a truncated power-law function that approximates the movement pattern of all mobile phone users. They identified a high degree of spatial and temporal regularities in the activity-travel patterns, meaning that people tend to visit a certain set of locations repeatedly. Their research, however, did not attempt to categorize the stop locations into different types of places (such as home, work, *etc*.).

Besides, some researchers have clustered individual activity-travel patterns into groups, and then compared the differences among them. Based on the trajectory data of 100 mobile phone users, Eagle and Pentland (2009) applied a method named 'Eigen behaviors' to model the rhyme of people's daily routine. They labeled the users' location status as 'home', 'work', 'elsewhere', 'no signal' or 'off'. The study greatly simplified the spatial information by displaying labels without actual spatial locations. Similarly, Jiang et al. (2012) used principle component analysis and eigen decom position to identify and depict mobility patterns of individual citizens, which are then clustered into different groups by the K-means clustering algorithms. Their research also highlights the temporal regularities of daily activities. A major breakthrough is that it applied powerful data mining and statistical learning methods to traditional travel surveys, making it possible to link group travel patterns with the underlying socio-economic information. Taking a further step, some researchers utilized the sequence alignment methods (SAM) to cluster people's activity-travel pattern (Wilson, 1998; Joh et al., 2002; Kwan et al., 2014; Liu et al., 2015; Cho et al., 2019). Their methods are more comprehensive by focusing on the sequential order of activities and trips, but they also tend to simplify the spatial information. For instance, Liu et al. (2015) simplified the location information to a series of location identifiers (IDs) and did not consider the spatial relationship (such as distance) between locations when generating the clusters. Cho et al. (2019) only considered movement direction of trips. To sum up, the existing clustering methods tend to either overlook or greatly simplify the spatial dimension of urban mobility, which limited the potential value of understanding the spatiality in urban mobility research.

3 The Interdisciplinary Idea and Research Hypotheses

In many ways, activity-travel patterns of individual citizens are similar to our eye fixation and movement pattern when we try to recognize other people's faces. Our eyes usually scan through different parts of a person's face (Chuk et al., 2014), just like citizens traveling in a city. Specifically, individual citizen's activities and travel behavior resemble the eye fixation and movement pattern on a face in that it follows a stop-travel-stop pattern. In both cases, the subject will spend some time staying at each 'stop' and then move to the next 'stop', and the prediction of the next 'stop' depends on the current 'stop', making it applicable to use the Hidden Markov Model (HMM) to estimate the patterns. Many studies have already used HMMs to summarize people's activity-travel pattern (Eagle and Pentland, 2006; Alvarez-Garcia et al., 2010; Lv et al., 2014, 2017; Han and Sohn, 2016; Zhu et al., 2018; Shi et al., 2019). However, most of them have only applied the method to GPS dataset or cellular network dataset where socio-demographic information is typically missing. Our study differs from previous ones in that we apply HMM to city-scale travel survey data, making it possible to generate more insights by comparing the model results between people with different socio-demographic characteristics, as well as using the trip purpose information to enrich the results.

In the applications of face recognition, a MATLAB toolbox was developed using HMM to estimate and compare the participants' eye movement patterns, namely the Eye Movement analysis with HMMs (EMHMM) (Chuk et al., 2014). In Psychology, the original applications were to examine the relationship between eye movement patterns and cognitive functions in different age groups or populations (Chuk et al., 2014, 2017; Chan et al., 2018; Zhang et al., 2019). In this research, we aim to adopt and further develop the EMHMM toolbox in psychological experiments to answer research questions about real-life urban mobility. A key step in all interdisciplinary research is the formulation of meaningful research questions relevant to the subject field and arguably better analyzed by the new methodological tool. In other words, there should be additional insights to be gained from the application of the new methodological tools from another discipline (in this case, the EMHMM toolbox), not readily gained using the more conventional methods of the field (such as the mapping of space-time trajectories). In relation, the following research questions are asked.

3.1 Research Question 1 (RQ1): Can the application of HMM reveal new insights about people's activity-travel pattern?

How do the activity space, the sequence of activities, the direction and duration of trips vary in different parts of a city? Instead of using all stop locations to depict the daily activity space or identify the primary region of interest (ROI) (as in González et al., 2008), the HMM method allows us to discover ROIs and transition probabilities among the ROIs in a data-driven fashion according to the visiting frequency and sequence of stop locations. With the discovered ROIs based on the underlying data nature rather than pre-identified trip purposes, we are able to examine the relationship between activity-travel pattern and trip purposes. In other words, this method does not require trip purpose information to define ROIs. It also allows us to compare activity-travel patterns more systematically, despite people's different home locations.

3.2 Research Question 2 (RQ2): Can we generalize about the spatial patterns of urban mobility based on the nature of stop-travel-stop patterns of people in a day?

Based on the results above, it is believed that there are underlying common patterns among individual activity-travel patterns that will make a typology of urban mobility possible. Through the clustering algorithm of the EMHMM method (Coviello et al., 2014), new insights about people's urban mobility patterns can be gained. While space-time trajectories can present the activity-travel pattern of an individual or small groups like a family well, the diagram can become very 'messy' and extremely crowded when the sample size is large (say, more than 100). Moreover, it is difficult to group these activity-travel patterns analytically because the home locations vary among the surveyed subjects.

3.3 Research Question 3 (RQ3): To what extent do people of different lifestyles living near urban core (CBD) spend more time at home, when compared to those who live farther away from the CBD?

Making use of HMM, we can integrate both spatial and temporal information by mapping out the ROIs, while at the same time showing the time spent by the subject at each ROI, as well as the probability of the subject moving from one region to another region. Hence, it is possible for us to focus more on the time dimension and examine whether the argument that people living in suburbs have less time to spend at home due to their longer commutes. Conversely, would living closer to the CBD mean longer time spent at home as well as for other activities?

3.4 Research Question 4 (RQ4): Is there any gender difference in terms of time spent at home for working adults living together?

Finally, the EMHMM toolbox allows us to consider subgroups with the urban population and gain insights into transport equity and other issues. Based on previous research (Hanson and Hanson, 1980; Buliung and Kanaroglou, 2006; Loo and Lam, 2013), the use of the EMHMM method is used to test whether women tend to spend more time at home than men, even if both of them are working. Moreover, we want to examine whether the gender difference varies across different parts of the city.

4 Methodology

4.1 Data and analytical framework

The major data source is the Travel Characteristic Survey (TCS 2011) in Hong Kong, China. Specifically, this study only uses data from the Household Interview Survey (HIS), the main part of the TCS 2011, conducted between September 2011 and January 2012. Data from 35 401 households were collected. Specifically, the HIS comprises of the household member (HM) database and mechanized trip (TP) database. It is important to note that the location information in these databases are provided in the form of street blocks or village clusters (SBVC), a geographical unit used by the Planning Department in Hong Kong. Therefore, the geographic coordinates of the centroids of the SBVCs are taken as origins and destinations of trips. For the purpose of this study, the HM and TP databases were integrated using Python in order to link individual attributes with the trip information. Household member expansion factors in the HM database are used at a later stage to estimate the number of surveyed people in each group, as well as the (weighted) average time spent at home and its immediate neighborhood in each administrative district. Moreover, TCS 2011 does not provide direct information about the marital status of survey participants. We can only identify working men and women living under the same roof. From the subset, we then find out the marital status by the reported 'relationship of household members' with the 'head of household'. If one person was reported as the household head and another person was reported to be his/her spouse, they are considered married working couples. Otherwise, we consider the relationship of the surveyed subjects (of different genders) living under the same roof as 'uncertain'. For instance, this may be the case for brother-and-sister or mother-and-son living together. Fig. 1 summarizes the relationship of the four research questions and the flow of data analysis.

4.2 Methodological challenges and advancements

Since the EMHMM toolbox was originally designed for estimating the pattern of eye movement, several conceptual and methodological challenges need to be overcome before it can be used for analyzing individual activity-travel patterns. These methodological challenges are listed below and discussed in relation to the research questions.

4.2.1 From Repeated Experiments of Each Subject in a Laboratory Setting to Single-day Travel Records of Individuals in the Real World

In the original eye movement analysis of face recognition, each participant undergone multiple experiments to provide sufficient data to generate an HMM. In TCS 2011, since each surveyed subject only provided trip in-

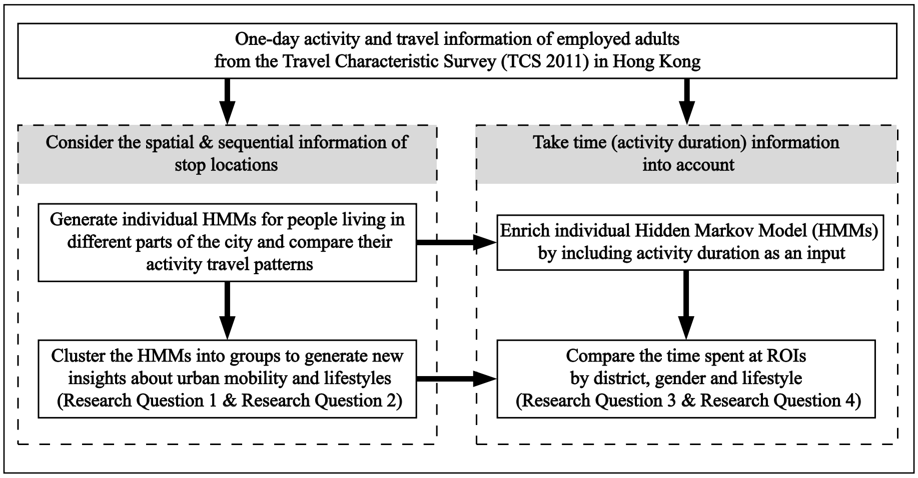

Fig. 1 An analytical framework showing the steps of data analysis and their relationships with the underlying research questions

formation on one day, it is insufficient to generate a reliable HMM. We were able to solve this problem by pre-processing the data to group them according to geographical units, in this case SBVC in Hong Kong, and then consider each geographical unit as one subject. Using this method, each subject will have multiple 'trials' that make it possible to generate a reliable HMM. In other words, the unit of analysis is spatial with one SBVC being considered as one subject. Only SBVCs having at least one working adult subject were selected. An SBVC containing only one surveyed subject was merged with the nearest SBVC. Also, the TCS 2011 database does not provide specific spatial coordinate information. Instead, the IDs of the SBVC of stop locations were provided. After grouping the data, the coordinate information of the home location was appended to the table by linking the IDs of the SBVC and the centroid coordinates of the SBVC derived from the SBVC boundary Shapefiles.

4.2.2 From the face to the city

In the original eye movement of face recognition research, the first eye fixation location of a participant may start at any point on the human face. In reality, most individuals would actually start their day at the home locations, which are fixed for the individuals but spatially scattered at the city level. This issue is addressed by pre-processing data in a way that always place the home location of each individual at the center of the canvas. In the eye movement analysis, the location of spatial features (such as eye, nose and mouth) are also relatively fixed on a human face. Different from that, destinations are much more variable across a city; and it is hard to judge the nature of place or activity type by spatial locations only. Therefore, a function was added to the EMHMM toolbox to illustrate the proportion of different place types (home, work, and others) at each ROI. Furthermore, the TCS 2011 database only provides trip information, while the modified EMHMM toolbox needs stop locations as its input. Therefore, Python scripts were developed to transform the original trip database into a stop location database, in which the activity duration, i.e. the time spent at each stop location, is calculated based on the arrival and departure time. After deriving the stop location database, we use the actual geographical coordinates of stop locations to compare with the home coordinates of each individual subject. In this way, we get the spatial deviation (both direction and distance) from the home location and use that information as the X, Y coordinates for the toolbox input.

5 Results

5.1 Home as the anchor (RQ1)

Using home locations as the 'anchor points' to visualize the activity-travel pattern of individual citizens makes the urban mobility pattern between individual citizens more comparable. Different from previous methods, here we recognize the importance of the place of residence (home location) by considering it as the center of a person's daily activities. After merging, there are 1494 SBVCs that have more than one working adult who participated in TCS 2011. Fig. 2 shows a sample individual HMM for one SBVC located relatively far away from the Central Business District (CBD) of Hong Kong, China. The idea of home location being an anchor point

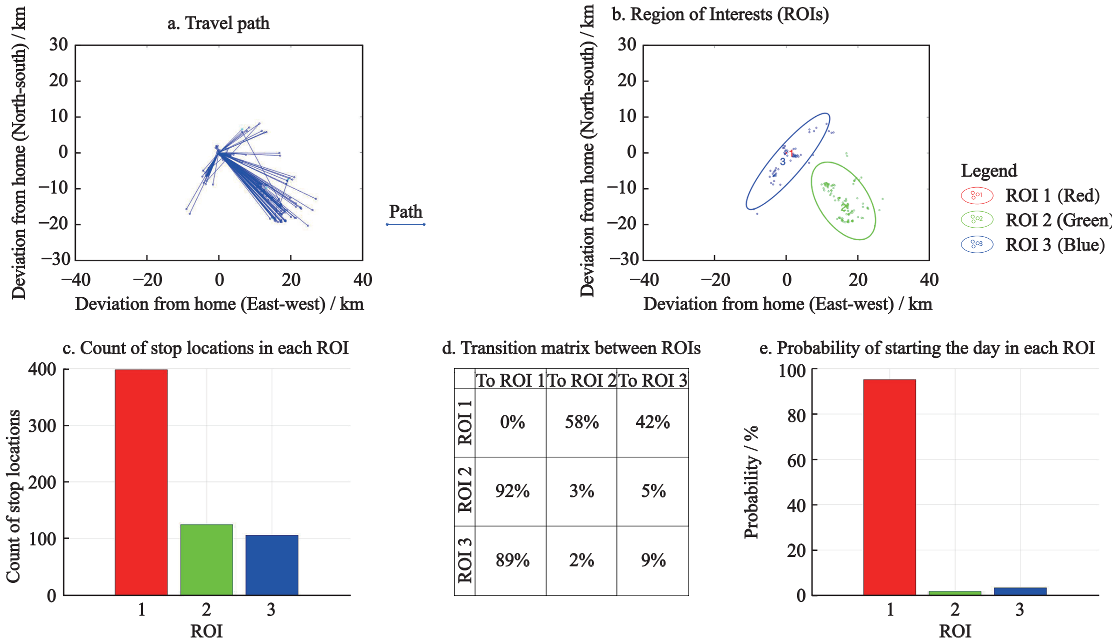

Fig. 2 A sample of an individual Hidden Markov Model (HMM) for one Street Block & Village Cluster (SBVC) far away from Central Business District (CBD) in Hong Kong, China. ROI stands for Region of Interest

of activities is clearly illustrated by Fig. 2a. Moreover, as is shown in Fig. 2b, the toolbox identified three regions of interest (ROIs) for the subject, which are depicted as ellipses colored in red, green, and blue. The ROIs are automatically ranked in a way that they represent the most likely travel path of individuals. Fig. 2c shows the stop location counts in each of the three ROIs, with the red ROI being the most frequently visited region, while the green ROI is the second frequently visited one and so on.

Fig. 3 shows the home and stop locations, as well as the activity space, in the form of one standard deviation

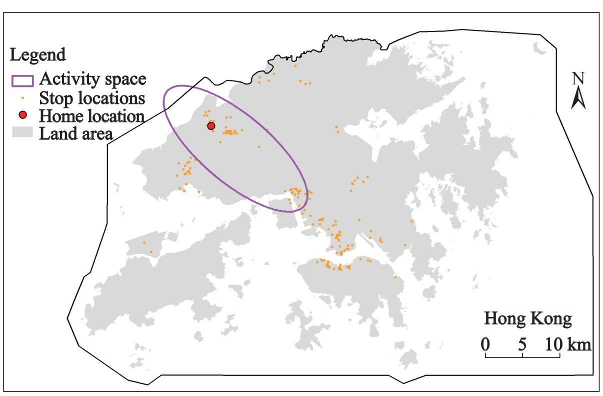

Fig. 3 Home location, stop locations, and activity space (in the form of standard deviation ellipse) for one Street Block & Village Cluster (SBVC) far away from Central Business District (CBD) in Hong Kong, China

ellipse for the street block far away from the CBD. In comparison, it is clear that the three ROIs in Fig. 2 provide a much better coverage of the stop locations than the activity space (In Fig. 3, a large number of stop locations are not covered by the standard deviation ellipse). Moreover, the HMM method provides more information about the most likely travel path of the subjects in the form of a transition matrix. The transition matrix in Fig. 2d shows the probability of a subject moving from one ROI to another. In this example, if the subject is currently in the red ROI, it has a 58% chance of traveling to the green ROI and 42% of chance traveling to the blue ROI. Fig. 2e shows the probability of the location of the subject at the start of the survey day. It can be observed that our method can provide more information (such as the most likely travel path) about the activity-travel pattern than the activity space.

5.2 Spatial typology of urban mobility types (RQ2)

Next, we use the EMHMM toolbox to identify common activity-travel patterns. After many trials, four clusters, as shown in Fig. 4, having the most distinctive patterns and highest value of interpretation are found. Groups 1 and 4 are visually different from Groups 2 and 3 in terms of the relative spatial distribution of the green and blue ROIs. The major difference between Group 1 and Group 4 is that the latter has a smaller blue ROI. On the

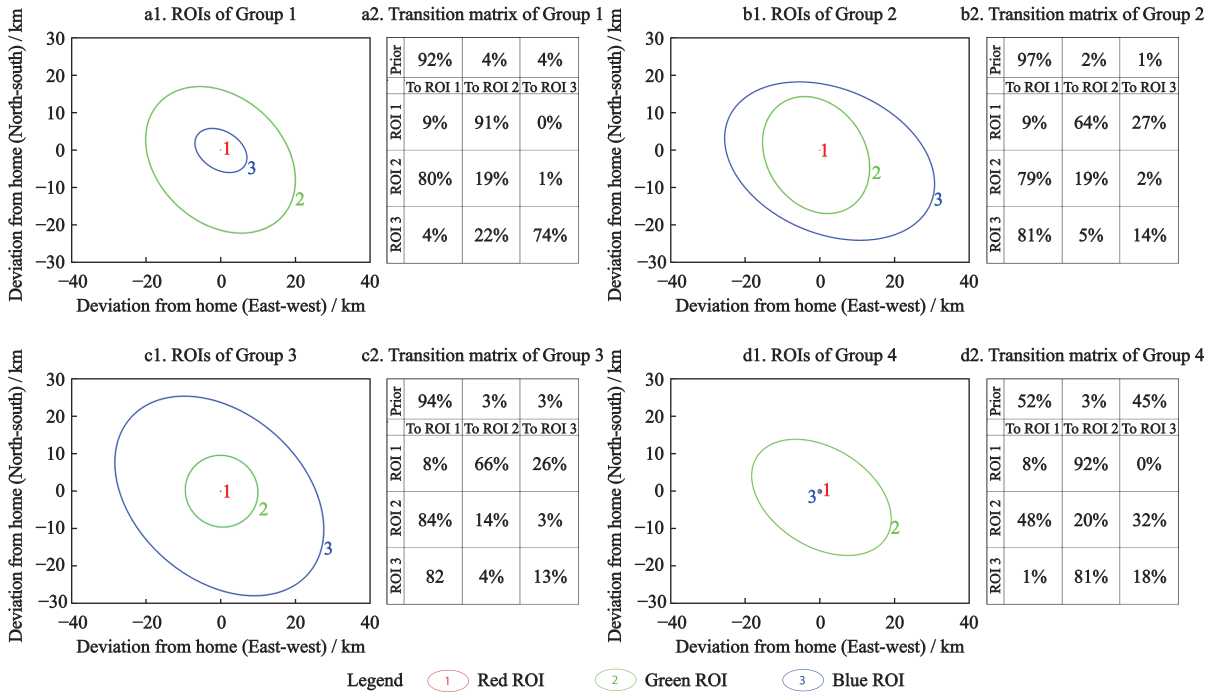

Fig. 4 Clustering results of all individual Hidden Markov Models (HMMs) in Hong Kong, China. ROI stands for Region of Interest. The first rows of the transition matrixes refer to the probabilities of the individual having his/her first stop location in each ROI

other hand, the major difference between Group 2 and Group 3 is that the latter has a smaller green ROI and a larger blue ROI than the former. From Fig. 4, there are also clear differences in the mostly likely travel paths (the values in the row named 'prior' shows the likelihood of starting at a certain ROI, and the value in matrix shows the probability of traveling from one ROI to another ROI). After starting their day at the red ROI, subjects in Group 1 and Group 4 have more than 90% of chance to travel to the green ROI, with almost 0% of chance of traveling to the blue ROI. On the contrary, after starting their day at the red ROI, Group 2 and Group 3 are having about 64% of chance to travel to the green ROI, and about 26% of chance to travel to blue ROI. Moreover, Group 4 has about 52% of chance to start their day at the red ROI, and 45% at the blue ROI, while the other three groups have more than 90% of chance to start their day at the red ROI.

Till now, no trip purpose information from the TCS 2011 dataset was incorporated. To get a better understanding of the new insights obtained, Table 1 shows the percentage of the types of places in each ROI of each group, using the trip purpose information available in TCS 2011 dataset. A few technical points are worth mentioning. First, the three ROIs have been generated independently for each group. Second, the three ROIs are generated solely based on the HMM model of stop-travel-stop pattern, that is, without defining the nature of the stop (e.g. home, work or others) a priori. Despite the above, there are noticeable similarities, especially in relation to the red ROIs. Specifically, more than 99% of stop locations in the red ROIs are labeled as 'home' for

Table 1 The nature of stops within each Region of Interest (ROI) / %

Group	ROI	Nature of Place		
		Home	Work	Others
1	Red	99.7	0.1	0.2
	Green	0	82.9	17.1
	Blue	0	44.6	55.4
2	Red	99.5	0.1	0.4
	Green	0	84.7	15.3
	Blue	0	85.8	14.2
3	Red	99.5	0.1	0.4
	Green	0	80.8	19.2
	Blue	0	82.1	17.9
4	Red	99.6	0.1	0.3
	Green	0	84.8	15.2
	Blue	0	24.4	75.6

Notes: ROI stands for Region of Interest. Percentages in table represent the shares of certain types of stop locations in the corresponding Region of Interests (ROIs) all four groups. This means that the red ROI typically refers to the home location. While about 99% of the red ROIs represent home locations across the four groups, the nature of the green and blue ROIs is much more variable. For the green ROIs, 80.8% (Group 3) to 84.8% (Group 4) (Table 1) were work locations. In other words, the green ROIs generally represent work locations but there are variations across groups. These shares can be considered as probabilities of the ROIs being associated with certain activities (such as work or leisure). For the blue ROIs, the situation is even more contrasting across groups with 14.2% (Group 2) to 75.6% (Group 4) being locations other than home and work place. For Groups 2 and 3, blue ROIs actually also refer predominantly to work-related locations (85.8% and 82.1% respectively). In other words, people in Groups 2 and 3 have their daily mobility primarily determined by various work activities. Inspired by the well-known term work-life balance, which differentiates work activities and other non-work activities such as leisure or family activities (Gregory and Milner, 2009; Lin et al., 2009; Jones et al., 2012), Group 2 and Group 3 are therefore labelled as having a 'work-oriented lifestyle' with both the green and blue ROIs primarily work-related. In contrast, working adults in Group 4 have a much more balanced daily routine represented by the home, work and other locations. This applies also but to a smaller extent to Group 1. Hence, these two groups are labelled as having a 'balanced lifestyle'. Spatially, the results suggest that, with home at the center, people with a balanced lifestyle (Groups 1 and 4) were having a relatively compact zone of other (non-work) activities around their homes but a relatively long commuting distance (as shown by the large extent of green ROIs around the center). This is particularly clear in the case of Group 4 with a very compact blue ROI. For the cluster of work-oriented lifestyle, these people were having the second highest frequent work stops near their homes, as shown by the more compact green ROIs around 10 km of their homes. Yet, they made work-related stops far away from home, as shown by the blue ROI being scattered and covering a bigger spatial extent of up to 40 km in different directions. Most stops in their daily life (including both the green and blue ROIs) were related to work.

Table 2 summarizes the number of SBVCs and the number of unweighted and weighted sample size in each group. When the weighted sample size is considered, Groups 2 and 3 together represented about 61.6% of the surveyed working adults. In other words, more than half of Hong Kong's working adults tend to have a work-oriented lifestyle, with their daily mobility pattern largely determined by work and work-related activities. In contrast, about 38.4% of the surveyed working adults belonged to Groups 1 and 4, suggesting that they are having a relatively more balanced lifestyle with a more or less non-work related ROI near their home locations.

Fig. 5 maps the geographical locations of the SBVCs belonging to each group. The dispersed patterns suggest that the methodology is powerful in differentiating different neighborhoods beyond the simple district divisions in Hong Kong. Spatially, the general pattern is for the street blocks closer to the urban core, that is, Hong Kong Island and Kowloon peninsula, to be in the work-oriented lifestyle of Groups 2 and 3. Some clear exceptions are found in the Stanley, the Mid-levels, and Chai Wan on Hong Kong Island, and Kowloon Tong, Hung Hom and Tai Kok Tsui on the Kowloon peninsula. Quite on the contrary, people living further away from the urban core tend to be in the balanced lifestyle cluster of Groups 1 and 4. Some clear exceptions are found around Kwun Tong, Shatin and Tsuen Wan.

5.3 The gender-cum-spatial dimension of activity-travel patterns (RQ3 and RQ4)

To answer RQ3 regarding the difference in the duration of time spent at home between working adults of different lifestyles who live near or far away from CBD, we

Table 2 Relative importance of the four activity-travel groups across Hong Kong, China

Type of data statistics	Group 1	Group 2	Group 3	Group 4	Total
Number of SBVCs	606	375	357	156	1494
Unweighted sample size	10402 (26.4%)	20712 (52.6%)	3676 (9.3%)	4571 (11.6%)	39361 (100%)
Weighted sample size	707977 (26.7%)	1377322 (52.0%)	254654 (9.6%)	308740 (11.7%)	2648693 (100%)

Notes: SBVCs stand for Street Blocks and Village Clusters; values in parenthesis are the proportion of the group in total samples

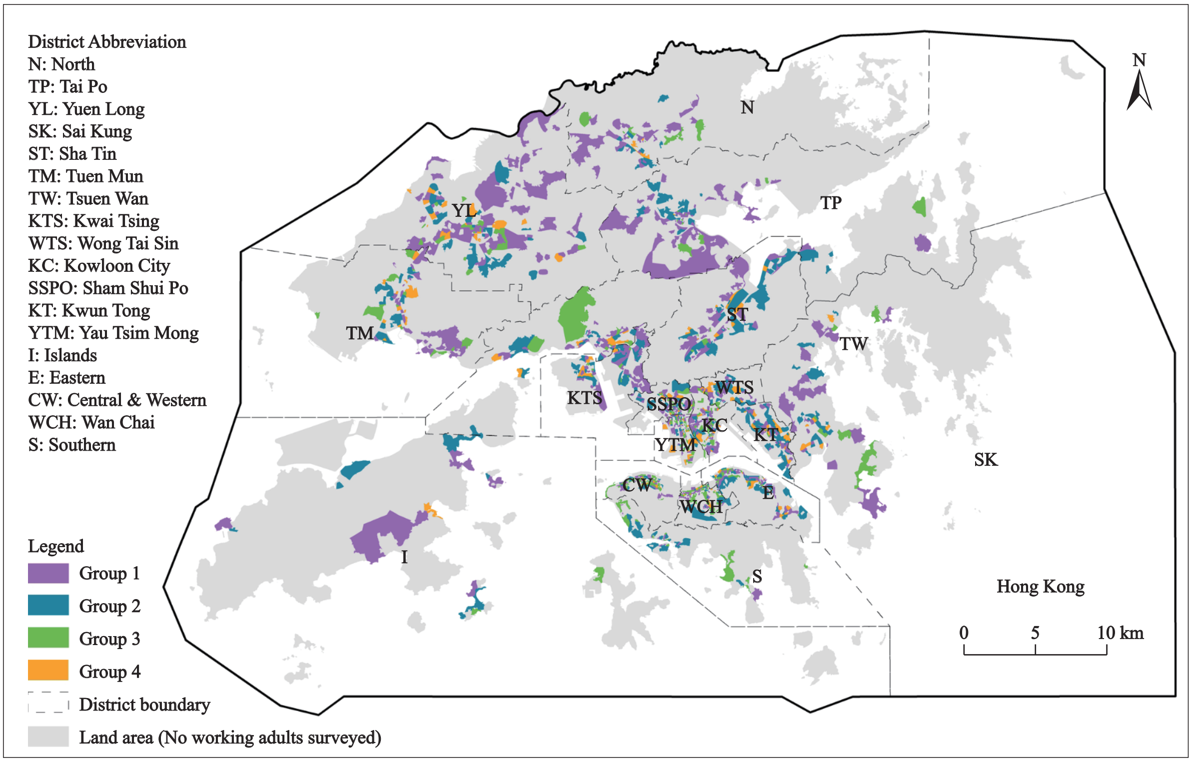

Fig. 5 Geographical distribution of activity-travel groups in Hong Kong, China

incorporate the time information (time spent at each stop location) in the analysis. To get a broader understanding of the time usage difference, Table 3 lists the means and standard deviations (s.d.) of the duration spent at each ROI by district. The top three values for each ROI are marked with asterisks. Focusing on the red ROI (notably homes), all top values were found in urban core areas, with individual workers spending an average of 6.89 h at home in Wan Chai. The figure was the lowest (6.13 h) for working adults living on Islands. Furthermore, Fig. 6 presents a scatterplot of average time spent at the red ROI by district. There is a negative relationship between the distance of home location to the urban core and time spent at home. The correlation coefficient (−0.85) is negative, strong and statistically significant at 0.05 level. In other words, people living near urban core tend to spend more time at the home location and its immediate neighborhood than those who live in peripheral areas. Time spent with family members at home has proven to be important for the quality of life for married couples (Greenhaus et al., 2003; Offer and Schneider, 2011). With more time to spend locally, people can also have a higher chance to interact with their neighbors, which can in turn cultivate social capital and contribute to the wellbeing of people (Loo et al., 2017), as well as the long-term success of communities (Putnam, 1993).

Fig. 7 illustrates the gender difference in time spent at the red ROI (RQ4). The districts are ordered by distance from the CBD of the city. Generally, employed women do spend more time at home compared to employed men regardless of where they live. The time-spent-at-home gap between different genders are bigger in districts located closer to the urban core, while the difference is smaller in districts located relatively far from the CBD, such as Tuen Mun, Yuen Long, and the North. To check if there is any statistically significant difference between the time spent at the red ROI by males and females at the SBVC level in Hong Kong, a *t*-test was conducted. The resulting P-value < 0.01 suggests that the difference is statistically significant. The gender gap of time spent at home identified can be explained by the gender difference in household responsibilities. Women are still generally expected to take more domestic responsibilities than men even when both of them are working (Hanson and Hanson, 1980; Coltrane, 2000; Offer and Schneider, 2011). The difference further widens when they have children (Loo and Lam, 2013; Jolly et al., 2014). Table 4 shows the weighted average time spent of married couples at the

Table 3 Time spent at each Region of Interest (ROI) in different districts of Hong Kong, China /h

Type of districts	Districts		Red ROI	Green ROI	Blue ROI
Urban Core Districts	Territory-wide	Mean	6.49	8.96	6.70
		s.d.	0.56	1.57	4.56
	Central & Western	Mean	6.74*	8.99	5.21
		s.d.	0.72	2.03	5.13
	Wan Chai	Mean	6.89*	8.36	5.31
		s.d.	1.00	2.59	4.96
	Yau Tsim Mong	Mean	6.69*	8.85	5.49
		s.d.	0.79	2.12	4.91
	Kowloon City	Mean	6.62	8.59	5.09
		s.d.	0.74	1.88	4.73
	Sham Shui Po	Mean	6.69*	8.68	6.31
		s.d.	0.55	1.64	4.88
	Eastern	Mean	6.63	8.64	7.08
		s.d.	0.53	1.57	4.61
	Southern	Mean	6.42	9.11	7.48
		s.d.	0.45	1.20	4.29
	Wong Tai Sin	Mean	6.58	8.73	5.93
		s.d.	0.42	1.22	4.37
	Kwun Tong	Mean	6.55	8.91	6.24
		s.d.	0.37	1.05	3.78
	Kwai Tsing	Mean	6.47	9.06	6.78
		s.d.	0.49	1.72	4.45
Other Districts	Sha Tin	Mean	6.49	8.80	7.31
		s.d.	0.42	1.29	4.52
	Tsuen Wan	Mean	6.45	9.37*	7.92*
		s.d.	0.45	1.92	4.34
	Sai Kung	Mean	6.37	9.38*	7.75*
		s.d.	0.43	0.98	3.97
	Islands	Mean	6.13	9.10	8.07*
		s.d.	0.69	1.13	3.57
	Tai Po	Mean	6.43	8.68	6.29
		s.d.	0.57	1.48	5.08
	Tuen Mun	Mean	6.29	9.26*	7.13
		s.d.	0.43	1.19	3.71
	Yuen Long	Mean	6.25	9.44*	7.19
		s.d.	0.56	1.86	4.73
	North	Mean	6.30	8.90	5.97
		s.d.	0.51	1.24	5.21

Notes: * Top three ranks (There are ties at the third rank. Hence, four values are marked); s.d., standard deviations

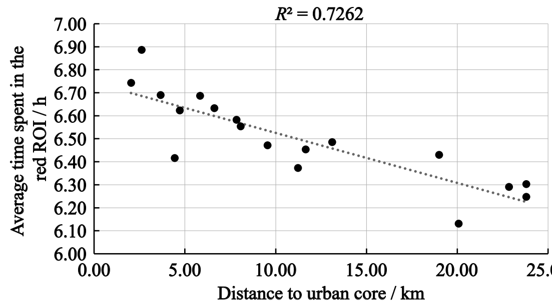

Fig. 6 Relationship of the average distance from home to urban core and the average time spent in the red Region of Interest (ROI) in 18 districts of Hong Kong, China

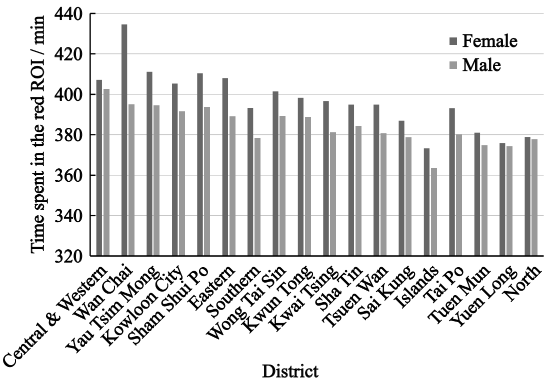

Fig. 7 Average time spent in the Red Region of Interest (ROI) for subjects across 18 districts in Hong Kong, China /min

Table 4 Weighted average time spent at home by married couples with different lifestyles /h

Districts	Work-oriented lifestyle		Balanced lifestyle	
	Female	Male	Female	Male
Territory-wide	6.58	6.39	6.63	6.40
Central & Western#	6.77	6.60	6.66	6.73
Wan Chai#	7.37	6.72	7.06	6.54
Yau Tsim Mong#	7.12	6.63	6.90	6.41
Kowloon City#	6.63	6.52	7.00	6.47
Sham Shui Po#	6.80	6.59	7.01	6.55
Eastern#	6.75	6.53	6.86	6.52
Southern#	6.57	6.30	6.39	6.39
Wong Tai Sin#	6.63	6.39	6.77	6.56
Kwun Tong#	6.65	6.49	6.69	6.47
Kwai Tsing#	6.58	6.45	6.69	6.17
Sha Tin	6.53	6.35	6.62	6.49
Tsuen Wan	6.51	6.24	6.68	6.44
Sai Kung	6.44	6.33	6.42	6.22
Islands	6.57	6.21	4.92	5.21
Tai Po	6.72	6.32	6.46	6.38
Tuen Mun	6.36	6.28	6.34	6.19
Yuen Long	6.28	6.19	6.23	6.31
North	6.22	6.29	6.31	6.28

Notes: # urban core districts

red ROI by lifestyle cluster. Moreover, the gender difference within the same lifestyle group is also statistically significant ($P < 0.01$, $df = 390\ 840$ for the balanced lifestyle; $P < 0.01$, $df = 574\ 892$ for the work-oriented lifestyle). In other words, married women tend to spend more time at home than their counterparts regardless of their lifestyle.

6 Conclusions

This study represents an interdisciplinary attempt to adopt an HMM toolbox developed in Computer Science for the analysis of eye movement patterns in Psychology to answer urban mobility questions in Geography. With regard to RQ1, this application shows that HMM can reveal new insights about people's activity-travel pattern. The conceptual and technical difficulties of transferring eye fixation on a face to locations of activities people engaged in a city can be overcome. Furthermore, the application demonstrates that we can generalize about the spatial patterns of urban mobility based on the nature of stop-travel-stop patterns of people in a day (RQ2). By using the adjusted toolbox to identify the activity-travel patterns of working adults in Hong Kong, two distinctive groups of balanced (38.4%) and work-oriented (61.6%) lifestyles were identified. As a step further, the spatial dimension was explored (RQ3). With some notable exceptions, working adults living the urban core were having a more work-oriented lifestyle. Those with a balanced lifestyle were having a relatively compact zone of non-work activities around their homes but a relatively long commuting distance. Last but not least, there was clear gender difference in terms of time spent at home for working adults living together (RQ4). Working females tend to spend more time at home than their counterparts, regardless of their marital status and lifestyle.

Overall, this interdisciplinary research demonstrates an innovative attempt to integrate spatial, temporal, and sequential information for understanding people's behavior in urban mobility research. It has a number of advantages over traditional approaches. First and foremost, this interdisciplinary approach can take the spatial, sequential, and temporal information of the data into consideration, while traditional methods tend to focus only on either the spatial or the temporal dimension when depicting individual travel patterns. The current study made use of TCS 2011 dataset. Nonetheless, the analysis can be repeated for any other city or place where disaggregate mobility survey data and location information are available. By taking the sequence of stop locations as an input, this method identifies not only the major ROIs that the subject has visited but also the most likely travel paths between ROIs. By considering the home location as the anchor point, geo-visualization of people's travel direction and distance becomes easier. Moreover, this approach can depict ROIs that indicate the nature of stop locations according to the visiting frequency and sequence of stop locations (even without trip purpose/place nature) information, making it a promising method to analyze data without trip purpose information (such as GPS data or mobile phone data). In an era where more digital traces are available without detailed trip purpose information, the method presented in this paper may serve as a powerful tool to conduct big data analysis about people's activity-travel pattern. This method can also cluster individual travel patterns into groups, which provide meaningful implications about the travel direction, distance and most likely travel sequence among different ROIs. It highlights the spatial and sequential dimensions when compared to other clustering methods. With this approach, both the direction of movement and the actual X, Y coordinates in relation to home location are considered. One weakness of the method is that the patterns of different groups after clustering are sometimes not very distinctive, possibly because the method considers multiple dimensions at the same time. More in-depth studies about the relationships of ROIs and activity types are needed. Another limitation of this study is that only one day of activity-travel diary is available. In the future, if multiple-day activity-travel dairy is available, a more detailed analysis can be conducted. Nonetheless, it is hoped that this research has demonstrated the potential value and challenges of innovative and interdisciplinary thinking in making theoretical and methodological contributions to urban mobility research.

References

Alvarez-Garcia J A, Ortega J A, Gonzalez-Abril L et al., 2010. Trip destination prediction based on past GPS log using a hidden Markov model. *Expert Systems with Applications*, 37(12): 8166–8171. doi: 10.1016/j.eswa.2010.05.070

Australian Research Council, 2020. An overview of ARC grant opportunities. Available at https://www.arc.gov.au/overview-arc-grant-opportunities/linkage-program-supporting-industry-research-collaboration. 2020-01-05.

Babb C, Olaru D, Curtis C et al., 2017. Children's active travel, local activity spaces and wellbeing: a case study in Perth, WA. *Travel Behaviour and Society*, 9: 81–94. doi: 10.1016/j.tbs.2017.06.002

Becker R A, Caceres R, Hanson K et al., 2011. A tale of one city: using cellular network data for urban planning. *IEEE Pervasive Computing*, 10(4): 18–26. doi: 10.1109/MPRV.2011.44

Buliung R N, Kanaroglou P S, 2006. Urban form and household activity-travel behavior. *Growth and Change*, 37(2): 172–199. doi: 10.1111/j.1468-2257.2006.00314.x

Buliung R N, Remmel T K, 2008. Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. *Journal of Geographical Systems*, 10(2): 191–216. doi: 10.1007/s10109-008-0063-7

Chan C Y H, Chan A B, Lee T M C et al., 2018. Eye-movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic Bulletin & Review*, 25(6): 2200–2207. doi: 10.3758/s13423-017-1419-0

Cho S J, Janssens D, Joh C H et al., 2019. Space-time sequential similarity for identifying factors of activity-travel pattern segmentation: a measure of sequence alignment and path similarity. *Geographical Analysis*, 51(2): 203–220. doi: 10.1111/gean.12186

Choi J, do Lee W, Park W H et al., 2014. Analyzing changes in travel behavior in time and space using household travel surveys in Seoul Metropolitan Area over eight years. *Travel Behaviour and Society*, 1(1): 3–14. doi: 10.1016/j.tbs.2013.10.003

Chuk T, Chan A B, Hsiao J H, 2014. Understanding eye movements in face recognition using hidden Markov models. *Journal of Vision*, 14(11): 8. doi: 10.1167/14.11.8

Chuk T, Chan A B, Hsiao J H, 2017. Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling. *Vision Research*, 141: 204–216. doi: 10.1016/j.visres.2017.03.010

Coltrane S, 2000. Research on household labor: modeling and measuring the social embeddedness of routine family work. *Journal of Marriage and Family*, 62(4): 1208–1233. doi: 10.1111/j.1741-3737.2000.01208.x

Coviello E, Chan A B, Lanckriet G R G, 2014. Clustering hidden markov models with variational HEM. *The Journal of Machine Learning Research*, 15(1): 697–747. doi: 10.5555/2627435.2627457

Eagle N, Pentland A, 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4): 255–268. doi: 10.1007/s00779-005-0046-3

Eagle N, Pentland A S, 2009. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7): 1057–1066. doi: 10.1007/s00265-009-0739-0

European Commission, 2019. Cities as climate-resilient, connected multimodal nodes for smart and clean mobility: new approaches towards demonstrating and testing innovative solutions. Available at https://www.arc.gov.au/overview-arc-grant-opportunities. 2020-01-05.

Ferreira N, Poco J, Vo H T et al., 2013. Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12): 2149–2158. doi: 10.1109/TVCG.2013.226

Fransen K, Farber S, Deruyter G et al., 2018. A spatio-temporal accessibility measure for modelling activity participation in discretionary activities. *Travel Behaviour and Society*, 10: 10–20. doi: 10.1016/j.tbs.2017.09.002

González M C, Hidalgo C A, Barabási A L, 2008. Understanding individual human mobility patterns. *Nature*, 453(7196): 779–782. doi: 10.1038/nature06958

Greenhaus J H, Collins K M, Shaw J D, 2003. The relation between work-family balance and quality of life. *Journal of Vocational Behavior*, 63(3): 510–531. doi: 10.1016/S0001-8791(02)00042-8

Gregory A, Milner S, 2009. Editorial: work-life balance: a matter of choice? *Gender, Work & Organization*, 16(1): 1–13.

Hägerstraand T, 1970. What about people in regional science? *Papers in Regional Science*, 24(1): 7–24. doi: 10.1111/j.1435-5597.1970.tb01464.x

Han G, Sohn K, 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model. *Transportation Research Part B: Methodological*, 83: 121–135. doi: 10.1016/j.trb.2015.11.015

Hanson S, Hanson P, 1980. Gender and urban activity patterns in Uppsala, Sweden. *Geographical Review*, 70(3): 291–299. doi: 10.2307/214257

Jiang S, Ferreira J, González M C, 2012. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25(3): 478–510. doi: 10.1007/s10618-012-0264-z

Joh C H, Arentze T, Hofman F et al., 2002. Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological*, 36(5): 385–403. doi: 10.1016/S0191-2615(01)00009-1

Jolly S, Griffith K A, DeCastro R et al., 2014. Gender differences in time spent on parenting and domestic responsibilities by high-achieving young physician-researchers. *Annals of Internal Medicine*, 160(5): 344–353. doi: 10.7326/M13-0974

Jones F, Burke R J, Westman M, 2012. Work-life Balance: A Psychological Perspective. New York: Psychology Press.

Kwan M P, 1998. Space-time and integral measures of individual accessibility: a comparative analysis using a point - based framework. *Geographical Analysis*, 30(3): 191–216. doi: 10.1111/j.1538-4632.1998.tb00396.x

Kwan M P, Xiao N C, Ding G X, 2014. Assessing activity pattern similarity with multidimensional sequence alignment based on a multiobjective optimization evolutionary algorithm. *Geographical Analysis*, 46(3): 297–320. doi: 10.1111/gean.12040

Lam W W Y, Loo B P Y, Yao S J, 2013. Towards exposure-based time-space pedestrian crash analysis in facing the challenges of ageing societies in Asia. *Asian Geographer*, 30(2): 105–125. doi: 10.1080/10225706.2012.735436

Lawrence R J, 2004. Housing and health: from interdisciplinary principles to transdisciplinary research and practice. *Futures*, 36(4): 487–502. doi: 10.1016/j.futures.2003.10.001

Li S Y, Lyu D, Liu X P et al., 2020. The varying patterns of rail transit ridership and their relationships with fine-scale built environment factors: Big data analytics from Guangzhou. *Cities*, 99: 102580. doi: 10.1016/j.cities.2019.102580

Lin H Z, Lo H P, Chen X J, 2009. Lifestyle classifications with and without activity-travel patterns. *Transportation Research Part A: Policy and Practice*, 43(6): 626–638. doi: 10.1016/j.tra.2009.04.002

Liu F, Janssens D, Cui J X et al., 2015. Characterizing activity sequences using profile hidden markov models. *Expert Systems with Applications*, 42(13): 5705–5722. doi: 10.1016/j.eswa.2015.02.057

Long Ying, Zhang Yu, Cui Chengyin, 2012. Identifying commuting pattern of Beijing using bus smart card data. *Acta Geographica Sinica*, 67(10): 1339–1352. (in Chinese)

Loo B P Y, Lam W W Y, 2013. A multilevel investigation of differential individual mobility of working couples with children: A case study of Hong Kong. *Transportmetrica A: Transport Science*, 9(7): 629–652. doi: 10.1080/18128602.2011.643509

Loo B P Y, Lam W W Y, Mahendran R et al., 2017. How is the neighborhood environment related to the health of seniors living in Hong Kong, Singapore, and Tokyo? Some insights for promoting aging in place. *Annals of the American Association of Geographers*, 107(4): 812–828. doi: 10.1080/24694452.2016.1271306

Loo B P Y, Tang W S M, 2018. "Mapping" smart cities. *Journal of Urban Technology*, 26(2): 129–146. doi: 10.1080/10630732.2019.1576467

Loo B P Y, Wang B, 2017. Progress of e-development in China since 1998. *Telecommunications Policy*, 41(9): 731–742. doi: 10.1016/j.telpol.2017.03.001

Lv Q J, Mei Z S, Qiao Y Y et al., 2014. Hidden Markov model based user mobility analysis in LTE network. In Proceedings of 2014 International Symposium on Wireless Personal Multimedia Communications. Sydney, Australia: IEEE, 379 –384. doi: 10.1109/WPMC.2014.7014848

Lv Q J, Qiao Y Y, Ansari N et al., 2017. Big data driven hidden Markov model based individual mobility prediction at points of interest. *IEEE Transactions on Vehicular Technology*, 66(6): 5204–5216. doi: 10.1109/TVT.2016.2611654

Miller H J, 1991. Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information System*, 5(3): 287–301. doi: 10.1080/02693799108927856

National Science Foundation, 2020. Environmental convergence opportunities in chemical, bioengineering, environmental, and transport systems (ECO-CBET). Available at https://www.nsf.gov/pubs/2020/nsf20517/nsf20517.htm. 2020-01-05.

Offer S, Schneider B, 2011. Revisiting the gender gap in time-use patterns: multitasking and well-being among mothers and fathers in dual-earner families. *American Sociological Review*, 76(6): 809–833. doi: 10.1177/0003122411425170

Pugh T A M, MacKenzie A R, Whyatt J D et al., 2012. Effectiveness of green infrastructure for improvement of air quality in urban street canyons. *Environmental Science & Technology*, 46(14): 7692–7699. doi: 10.1021/es300826w

Putnam R D, 1993. The prosperous community: social capital and public life. *The American Prospect*, 13: 35–42.

Sallis J F, Frank L D, Saelens B E et al., 2004. Active transportation and physical activity: opportunities for collaboration on transportation and public health research. *Transportation Research Part A: Policy and Practice*, 38(4): 249–268. doi: 10.1016/j.tra.2003.11.003

Shaw S L, Yu H B, Bombom L S, 2008. A space-time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in GIS*, 12(4): 425–441. doi: 10.1111/j.1467-9671.2008.01114.x

Shi H Z, Li Y, Cao H C et al., 2019. Semantics-aware hidden Markov model for human mobility. IEEE Transactions on Knowledge and Data Engineering. doi: 10.1109/TKDE.2019.2937296

Wilson W C, 1998. Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A: Economy and Space*, 30(6): 1017–1038. doi: 10.1068/a301017

Wilson W C, 2008. Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation*, 35(4): 485–499. doi: 10.1007/s11116-008-9162-z

Xu Y Y, Jiang S, Li R Q et al., 2019. Unraveling environmental justice in ambient PM2. *5 exposure in Beijing: a big data approach. Computers, Environment and Urban Systems*, 75: 12–21. doi: 10.1016/j.compenvurbsys.2018.12.006

Xu Z, Liu Y H, Yen N Y et al., 2020. Crowdsourcing based description of urban emergency events using social media big data. *IEEE Transactions on Cloud Computing*, 8(2): 387–397. doi: 10.1109/TCC.2016.2517638

Zhang J X, Chan A B, Lau E Y Y et al., 2019. Individuals with insomnia misrecognize angry faces as fearful faces while missing the eyes: an eye-tracking study. *Sleep*, 42(2): zsy220. doi: 10.1093/sleep/zsy220

Zhang P, Zhou J P, Zhang T R, 2017. Quantifying and visualizing jobs-housing balance with big data: a case study of Shanghai. *Cities*, 66: 10–22. doi: 10.1016/j.cities.2017.03.004

Zhao P J, Hu H Y, 2019. Geographical patterns of traffic congestion in growing megacities: big data analytics from Beijing. *Cities*, 92: 164–174. doi: 10.1016/j.cities.2019.03.022

Zhou X G, Yeh A G O, Yue Y, 2018. Spatial variation of self-containment and jobs-housing balance in Shenzhen using cellphone big data. *Journal of Transport Geography*, 68: 102–108. doi: 10.1016/j.jtrangeo.2017.12.006

Zhu W Z, Zhang C, Yao S C et al., 2018. A spherical hidden Markov model for semantics-rich human mobility modeling. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI, 4613–4620.