

RESEARCH ARTICLE

Open Access



Towards big data behavioral analysis: rethinking GPS trajectory mining approaches from geographic, semantic, and quantitative perspectives

Weixin Huang*  and Luying Wang

Abstract

The question regarding the actual usage of built environments is of immense importance in behavioral research. Yet traditional methods of collecting and analyzing data on movements and activities often lack needed accuracy and granularity. Thus, this article reviewed and summarized the applicability of emergent GPS trajectory mining approaches in the field of architecture from geographic, semantic, and quantitative perspectives, respectively. Accordingly, three experiments based on a case study using real GPS trajectory data from visitors to the Palace Museum in China were conducted to examine the usefulness and weakness of the aforementioned approaches. The findings revealed that although all three dimensions of the trajectory mining approaches had the potential to provide useful information for architectural and urban design, the higher the dimensionality in utilizing the data, the more effective the approach was in discovering generalizable knowledge of human behavioral pattern. Furthermore, the results suggested that to gain insights into the typological characteristics of human behaviors related to the built environments, the contribution of trajectory data alone was limited, hence, conventional field surveys and questionnaires which contain information on individual characteristics and spatial features should be used in conjunction. Future research and practical implications were outlined.

Keywords: Behavioral analysis, Clustering, GPS trajectory mining, Sequence alignment, The built environment, The Palace Museum

1 Introduction

The relationships between behavioral science and the built environment are of growing importance during recent decades, since awareness of rational allocation of spatial and environmental resources has risen. Mobility behavioral analysis specifically focuses on space utilization and users' spatial-temporal behavioral patterns, and contemporary mobility behavioral typology studies have an increasing field of interest on this issue. In the field of

architecture, most of the research emphasizes flow organization and functional structure in architectural spaces (Yang & Huang, 2019; Sheng et al., 2021), while in the behavior domain, individual characteristics and demands on the usage of spaces are worth more exploration (Kupper & Seyfried, 2020; Xu et al., 2020). For the recording of human mobility behavior, there are different data sources and methods, including activity diaries (Wilson, 2008), qualified observations (Peng & Maing, 2021), questionnaire surveys (Lee et al., 2011), spatial syntax (Chang & Penn, 1998; Yang et al., 2015; Sheng et al., 2021) and a variety of visualization attempts (Zhang et al., 2021). These methodologies are standardized nationally and internationally by organizations such as the International

*Correspondence: huangwx@mail.tsinghua.edu.cn

School of Architecture, Tsinghua University, 30, Shuangqing Road, Haidian District, Beijing 100084, China

Social Science Council (ISSC) and the Chinese Academy of Social Sciences (CASS) as behavioral measurement and interpretation tools and are also widely practiced in architecture and urban space investigations. Nevertheless, conventional qualitative methods are too limited and restricted to answer complicated questions about actual space usage in the built environment. The traditional questionnaire survey on space usage behavior, for example, such as the evaluation of flow organization and spatial layout, does not provide researchers with information concerning the real-time trajectories of the users or variations in behavioral decision-making when engaging with the environment. Self-report measures based on participants' feelings and memories are also prone to information omissions and impression mistakes. Qualified observations from a spectator's perspective could to some extent cover these deficiencies, but it is scarcely possible to monitor a user's spatial-temporal activities permanently with limited numbers of observers. Spatial-temporal behavioral analyses play a more and more important role in humanizing flow design, facilitating sustainable public resource allocation, and boosting post-occupancy evaluation. Data obtained following the traditional methodology, however, remains too sparse in resolution, and the majority of spatial behavioral studies still tend to follow qualitative approaches or focus on very narrow case studies. This has prompted an extensive debate on the scope and methodological innovation in research examining human behavior related to the built environment (Kwan et al., 2003; Hadjri & Crozier, 2009; Zhang et al., 2021).

Spatial-temporal behavioral studies in architectural and urban design contexts, despite the comprehensiveness of qualitative approaches, still need quantitative datasets to recognize the contradictions between the pre-designed spatial flows and the actual movement behavior of space users. Further, data collection techniques without interfering with individual activities are necessary to record individual movement trajectories over time objectively and accurately. Recent developments in space positioning technology, including Wi-Fi indoor positioning system (IPS), ultra-wideband (UWB) IPS, Bluetooth technology, mobile (cellular) phone positioning, global positioning system (GPS), along with multi-source big data including geo-tagged photos and surveillance videos are advancing surveying methods and digital datasets in spatial-temporal behavioral studies. The geographical information system (GIS) and relevant visualization methods also contribute to time and space behavioral studies (Ahas et al., 2008; Wu & Carson, 2008; Kwan et al., 2015; Xu et al., 2020). One of the emerging subjects in movement

behavioral studies is connected with GPS trajectory mining and spatial-temporal sequence analysis (Cho & Kang, 2019; Li et al., 2021; Tang et al., 2018; Yuan et al., 2017; Brum-Bastos et al., 2018). GPS trajectory mining has great potential for applications in space-time behavioral studies in buildings and urban spaces, though there are various restrictions and inaccuracies when applied in indoor spaces. Nevertheless, GPS trajectory data tends increasingly to complement conventional data sources and behavioral observations, as it is becoming crucial how humans utilize space and public resources.

GPS trajectory data is stored automatically in the memory files of handheld GPS locators, containing a huge amount of geographical information, which requires specialized analytic tools (e.g. GIS, Python programming, etc.) and data mining algorithms to obtain useful patterns and knowledge. The objective of this paper was to introduce and evaluate the applicability of the GPS trajectory mining approaches, which have been explored mainly in transportation and tourism studies but seldom studied in behavioral science. Three main research questions are as follows:

- (1) What are the potentials of trajectory mining approaches for uncovering useful behavioral patterns in architectural and urban design?
- (2) How do different trajectory mining approaches perform? Which approach is more suitable for exploring behavioral patterns in complex spatial-temporal trajectories?
- (3) With GPS trajectory data and mining techniques, do conventional data and methods such as questionnaires still matter?

To gain these insights and determine the potential worth of the various trajectory mining approaches, we illustrated our discussion with three analytical examples from the Palace Museum study projects with visitors' GPS trajectory data during the tour, and the advantages and limitations of each trajectory mining approaches were explored and discussed. The first example tested a hierarchical raw-data-based clustering method to identify the spatial similarity in the visitors' GPS trajectories from a geographic perspective. The second example examined a supervised sequence alignment approach to recognize similar trajectory sequences from a semantic perspective. The third example evaluated the metrics-based clustering approach, which used three spatial-temporal measures to represent movement features to cluster the trajectories, and the characteristics of clustering results were explored.

2 Trajectory pattern mining and behavioral science of the built environment

The study of human behavior in relation to the built environments and post-occupancy evaluation are quite intertwined, and these two research fields both encounter the same challenge: a lack of suitable technical tools for measuring and collecting data regarding the actual usage of the built environments. As Cooper et al. (1991) mentioned in a paper, “the question of measurement is difficult”, and “a data bank for the dissemination of information on accessibility and other aspects of environmental assessment” should be developed. To address this question, trajectory datasets, including those without high resolution as GPS, could play a significant role, because behavioral patterns could be identified from these trajectories, and provide reliable empirical evidence for post-occupancy evaluation and behavioral analysis related to the built environment.

More specifically, the GPS trajectory raw data contains basic information including time points, latitude, longitude, and altitude coordinates. Aside from trajectory visualization and basic spatial-temporal distribution, it is necessary to further mine the trajectory to obtain the available knowledge and implicit behavioral patterns. For example, identifying the frequent trajectory segments or location sequences in trajectories could provide insights on questions like which paths and spaces in a place have higher usage rates, and how to allocate public resources reasonably in line with the amount of people’s usage demand.

Take another example of stay behavior, identifying the stopover points or short stay areas from the trajectories could help identify the consistent stay behavior of people, so as to know which places or areas are more likely attractive for people to stay. Furthermore, the spatial and landscape features including spatial openness, spatial shape, interfacial textures, and the presence of greenery in these stay areas could be examined. In addition, analyzing the diverse patterns of people with diverse characteristics in stay behavior provides us the opportunity to further understand human decision-making mechanisms, and this knowledge obtained from the trajectory tends to be useful for the design of the architectural and urban spaces.

The exploration of trajectory data and its mining approaches for behavioral studies in built environments has been increasingly significant and pressing since the outbreak of the COVID-19 pandemic (Lu et al., 2021). In a special issue of *Architectural Journal*, there was a group discussion about the impact of COVID-19 on the contemporary city, new human habitat, and architectural design led by three Chinese academicians and 22 experts, scholars, and architects (Wang et al., 2020). In this paper,

a lack of corresponding empirical data and documentation on spaces formed by temporary gatherings, including bazaars and food markets was pointed out. Previous researchers have hardly examined the flow distribution, human behavioral characteristics, and the relationships between main flow lines during the usage of such spaces, resulting in difficulties in making rational judgments and accurate decisions when encountering unexpected events like the COVID-19 pandemic. When tracking the movement of patients to identify potentially infected persons, if provided with a more complete database on human spatial behavior, key associated areas and vulnerable populations could also be predicted and the virus transmission chain could be stopped in advance.

The demands for human trajectory data in green building and the building industry are also emerging, particularly in the context of sustainability and smart buildings (Moreno et al., 2016; Chen et al., 2019). Making predictions of human behavior or occupation and including these considerations in the design phase before construction or in the building control system can help with building energy saving and hence carbon reduction. With digital positioning data, the feasibility of such predictions has been confirmed, although relevant research results are limited in number (Jin et al., 2021). An occupancy prediction model was proposed by Huang et al. (2019) using a Bayesian method and proved to be effective based on the dwell time distribution extracted from the passenger flow data in a Chinese hub airport terminal.

Individuals’ trajectories not only reflect the places they visited, the routes they took, or the locations they stayed for a long period but also contain certain implicit information. For example, it is possible to identify people who may be companions based on the spatial-temporal similarity of trajectories, to discover the spatial functional associations based on frequent trajectory sequences, or to demonstrate the consistency of spatial-temporal behavioral patterns based on certain clustering features, even though their trajectory routes may be rather diverse. Further, we could discover how people make behavioral decisions regarding the utilization of spatial and temporal resources, and what factors may affect certain unconscious behavioral decisions.

All examples mentioned above demonstrate the importance of trajectory data in achieving an insightful understanding of human movement behavior in architectural or urban spaces. To the best of our knowledge, few studies, however, have addressed the aforementioned issues from the perspective of architectural disciplines involving the use of high-resolution trajectory data. While relating to all issues is beyond the scope of a single article, this study, therefore, tends to bring up a framework that structures an original synthesis of trajectory pattern

mining and behavioral research within the built environments, whose feasibility will be proved with a case study focusing on the evaluation of the applicability of various trajectory mining approaches.

3 Three dimensions of trajectory mining approaches

Early applications and analysis of GPS trajectory data originated from the field of transportation (Chakka et al., 2003) due to the application of GPS navigation systems in vehicles since the year 2001, producing massive vehicle and drivers' movement data. Two important tasks for trajectory data mining involve identifying similar paths and potential clusters in trajectories (Andrienko et al., 2008), hence, relevant approaches and algorithms are constantly being explored and innovated by numerous researchers and practitioners (Zhou et al., 2005, 2019; Lee et al., 2007; Dodge et al., 2012; Yuan et al., 2017; Cheng et al., 2021). Considering existing trajectory data analysis applications, the approaches of trajectory mining could be categorized into the following three types: raw-data-based, sequence-based, and metrics-based, interpreting the trajectory data in three diverse dimensions.

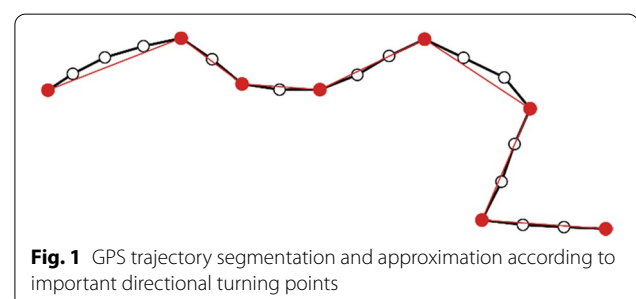
3.1 Raw-data-based approach

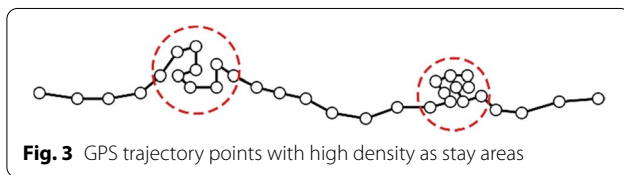
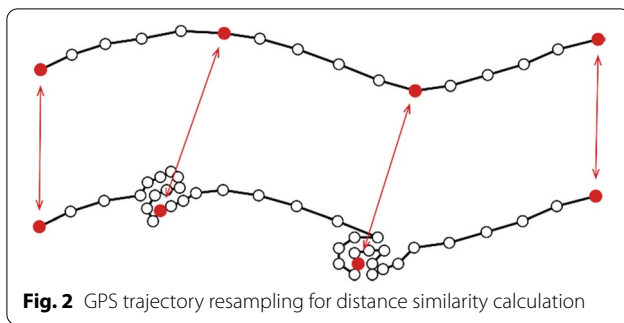
Raw GPS positioning data are usually of high resolution and massive data volume, and subject to occasional point jumps, which make it necessary to preprocess the trajectory data before analysis, otherwise, it would result in excessive memory consumption and unacceptable running time if we calculate the distance similarity for two trajectories directly. For this reason, trajectory data preprocessing techniques including compression, filtering, segmentation, and transformation were one of the most important research directions (Zheng & Zhou, 2011), aiming at efficient distance similarity calculation and fast clustering while retaining adequate geographical information (Lee et al., 2007; Tang et al., 2018). With the involvement and contribution of researchers in the field of computer science, relevant algorithms and techniques have been developed and exploited deeply (Lee & Krumm, 2011), including the Douglas-Peucker (DP) algorithm for trajectory compression (Douglas & Peucker, 1973; Hershberger & Snoeyink, 1992) and the Kalman filter for trajectory smoothing (Gelb, 1974). Regarding the situation that the trajectories are similar in spatial but inconsistent in time, certain algorithms modified from the DP algorithm such as a top-down time-ratio (TD-TR) algorithm (Meratnia & de By, 2004) were proposed, which took into consideration both geographic and temporal features of the trajectories.

Raw-data-based trajectory mining approaches mainly attempt to resolve three major issues: frequent path

search (Lee et al., 2007), trajectory route partition (Gaffney & Smyth, 1999), and stay area detection (Cho & Kang, 2019; Zhou et al., 2019). Regarding frequent path search, the basic analysis process is to segment the trajectories based on important directional turning points, where each segment represents an approximately linear path (see Fig. 1), and then calculate the similarity between segments based on certain features including distance and direction. After calculating the similarity matrix, various clustering algorithms including k-means, BIRCH, and DBSCAN could be used to search for similar trajectories. In the method framework of partition-and-group proposed by Lee et al. (2007), they argued that the density-based clustering method represented by DBSCAN is best suited for trajectory segment clustering because DBSCAN is particularly suitable for arbitrary shapes such as points and segments (Han et al., 2011).

Unlike frequent path search, trajectory route partition focuses on the similarities between complete trajectories rather than trajectory segments, which means that not only the spatial features of trajectories but also the temporal features must be taken into account. The duration of similar routes might vary greatly due to diverse movement speeds, which leads to trajectories with spatial similarities containing a widely different number of points, making the similarity calculation result inaccurate. The key to the trajectory route partition task is therefore to find similar spatial patterns while eliminating the interference of temporal factors, where the idea of Dynamic Time Warping (DTW) (Berndt & Clifford, 1994; Müller, 2007) could be useful, an algorithm first applied to speech recognition, and currently used in many areas including data mining and time series clustering (Senin, 2008). DTW finds the optimal non-linear alignment between two time series, and one point in a time series could be mapped to multiple points in the other time series. Regarding GPS trajectory, the problem of spatial-temporal inconsistency could also be tackled in a relatively straightforward way by resampling the trajectories and converting them into series with the same number of points (see Fig. 2). This less rigorous approach is feasible because GPS devices usually have a fixed sampling rate,





and the distance between each pair of sampling points is constrained without switching the transportation mode. Hence, the distance similarities are calculated only for the sampled points in order, which could greatly reduce the computational workload and improve the algorithmic efficiency.

Stay area detection is to identify various sets of points that appear in a small area consecutively for a period of time from movement trajectories, and it can be achieved by following two main concepts. One is to make full use of the spatial-temporal characteristics of the trajectory, calculate the average velocity of each trajectory segment, and mark a succession of points with an average velocity below a certain threshold as a stay area. The other approach is to treat all trajectories as a big dataset of points, and employ the density-based clustering algorithm (e.g. DBSCAN and OPTICS) to identify the points with high density as stay areas (see Fig. 3). This method, however, requires multiple iterations to adjust the parameters of DBSCAN regarding the neighborhood radius and the minimum number of points in each cluster, until an optimal stay area detection result is obtained.

3.2 Sequence-based approach

The trajectory sequence is an alternative representation of the trajectory, which is composed of symbols converted from the original trajectory using semantic information. Compared with raw-data-based approaches, basic spatial shapes of trajectories are no longer considered, instead, a higher semantic dimension is extended. This makes sequence-based trajectory mining approaches suitable for addressing issues in tourism studies and behavioral science because the trajectory data in early

tourism and activity research were mainly collected from travel or activity diaries (Wilson, 1998; Hallo et al., 2004), which contained sequence data of time, locations, and activities.

The question of how to analyze sequence data, however, was first raised in the field of molecular biology. To analyze the similarities between DNA or protein sequences, a sequence alignment method (SAM) was developed in the 1980s (Sankoff & Kruskal, 1983). SAM utilizes the edit distance such as the Levenshtein distance (Levenshtein, 1966) to measure the degree of similarity between two sequences. Edit distance or its variations (Dodge et al., 2012) are defined as the smallest number of elemental operations (e.g. insertions, deletions, and substitutions) required to convert one sequence into another. The more operations performed, the longer the distance between two sequences, hence, the smaller the similarity. The variations of SAM for GPS trajectory sequences could relate to artificially assigning different operation costs for different symbol pairs (Dodge et al., 2012), or representing GPS trajectories sequences using various coding methods, including meaningful functional locations, locations along with the time spent, and geographical polygonal area (Shoval et al., 2015).

Another common type of method for analyzing sequential data is association rule mining techniques such as PrefixSpan (Han et al., 2001), following the methodology of Apriori (Agrawal & Srikant, 1995), whose basic idea was to identify all of the strong association rules from the sequences given a minimum support threshold (the minimum occurrence frequency of the subsequences in the set of sequences) and a minimum confidence threshold (the minimum occurrence probability of the rule with a specific header subsequence).

A classical scenario using the Apriori method is to recognize patterns in sequences of events from an extensive database of customer transactions, for example, the rule {butter, bread} \Rightarrow {milk} means that if butter and bread are bought, customers also buy milk. With respect to trajectory sequences, certain fixed sequential patterns of associations between locations could be discovered using such approaches. In addition, the discovered rules could be used for predicting the next location of the following activity behaviors.

3.3 Metrics-based approach

Trajectory metrics refer to various measures, including total distance, total time, movement direction, movement speed, movement efficiency, etc., which are derived or calculated from the trajectory data to describe and characterize the trajectory. The selection of measures is mainly determined according to research questions and the characteristics of sample data (Gaffney & Smyth,

1999; Dodge et al., 2012; Ferrante et al., 2018; Grinberger & Shoval, 2019). These measures constitute a multidimensional feature matrix of the trajectory and could be analyzed using various algorithms, including dimension reduction, classification, clustering, etc., which are commonly used in the field of artificial intelligence and data mining. These algorithms, though, will not be elaborated on in the article. Compared with raw-data-based and sequence-based approaches, the metrics-based trajectory mining approach could identify more abstract behavioral and activity patterns beyond geographic or semantic patterns, which reflects human behavioral characteristics rather than geographic or location-based features.

In this article, we based our analyses on the aforementioned three dimensions of trajectory mining approaches to fully explore the possibilities of GPS trajectory data. While this study uses the approaches that generally follow the existing concepts and schemes for trajectory pattern mining, no known studies have implemented these three schemes on the same dataset to compare their differences in knowledge discovery and to demonstrate their applicable research questions, which is quite important when introducing a set of new methodologies into an emerging research field.

4 Case study: the Palace Museum

4.1 Data overview

The case study was based on GPS trajectories of 503 visitors collected in the Palace Museum in Beijing, China during 4 days of independent field surveys in the summer of 2020. The temporal sampling rate of the raw GPS data was approximately one fix per 3 s. An elaborate pre-processing procedure was conducted, including parsing the original format of GPS data and filtering data with multidimensional constraints including space and time. After the pre-processing, 405 valid trajectories with information including time, longitude, latitude, moving distance, and moving speed were obtained. Further data processing and conversion were performed accordingly for three types of trajectory mining approaches respectively, and these steps will be elaborated on in the corresponding sections. In addition, it should be noted that all data processing, analysis, mining, and visualization procedures conducted in this paper were implemented using Python programming (Kuhlman, 2009).

4.2 Raw-data-based approach: movement spatial similarity search from a geographic perspective

This experiment was intended to test the raw-data-based trajectory mining approach, and a basic hierarchical clustering method using Euclidean distance was employed. To improve the efficiency of the algorithm operation, the Ramer-Douglas-Peucker algorithm (Lee

et al., 2007) was applied following the pre-processing procedure for compressing and approximating the trajectories. An example of the approximate trajectory was drawn in Fig. 4, and the compressed trajectory represented the original one well concerning spatial shape and geographic information.

To reduce the interference of the time factor on the clustering process, all trajectories were equally divided to pick out the same number of points as samples. The distance matrix was then calculated between the corresponding sample points of each two trajectories. After that, a hierarchical agglomerative clustering algorithm was conducted, and the similarities between clusters were measured based on the complete linkage (Defays, 1977), i.e. calculating the distance between two most distant points in two clusters. Since this clustering process was unsupervised without knowing the optimal number of clusters, we evaluated the clustering quality using two common criteria, the average of the Sum of Squared Error (SSE) (Han et al., 2011) and the Silhouette Score (Rousseeuw, 1987) while varying the number of clusters, and determined the optimal results consequently.

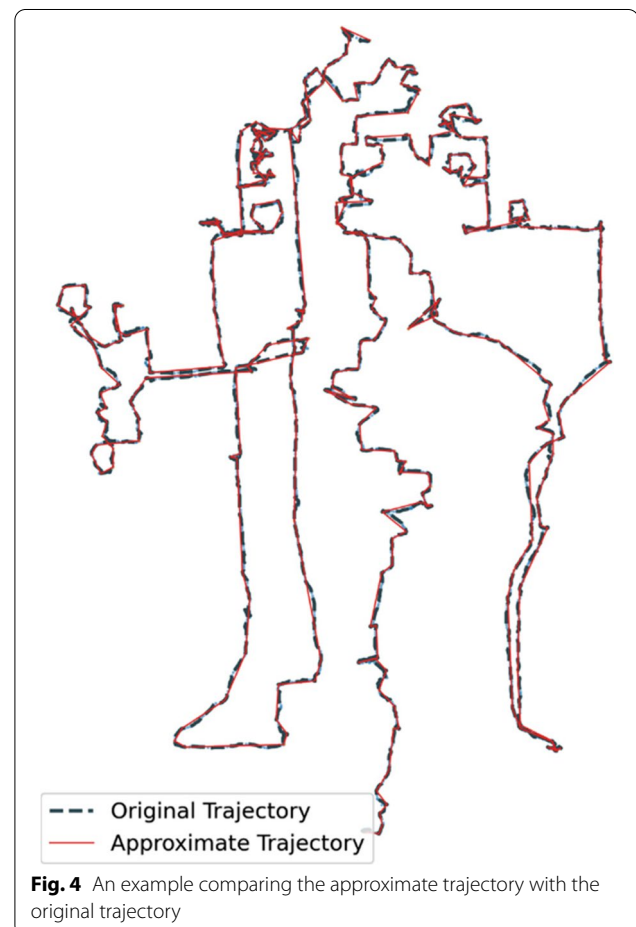


Fig. 4 An example comparing the approximate trajectory with the original trajectory

Figure 5 shows that the optimal number of clusters was 17, since a clustering result of good quality requires a smaller SSE and a larger Silhouette Score (greater than 0.3 is better). The clustering results with representative routes were visualized in Fig. 6. Seventeen clusters of

trajectories showed distinctions regarding route complexity, the number, and the order of the attractions visited. The trajectories in clusters 1, 2, 3, 4, and 14 were not quite complete, probably due to the data deficiency caused by equipment malfunction, indicating that the

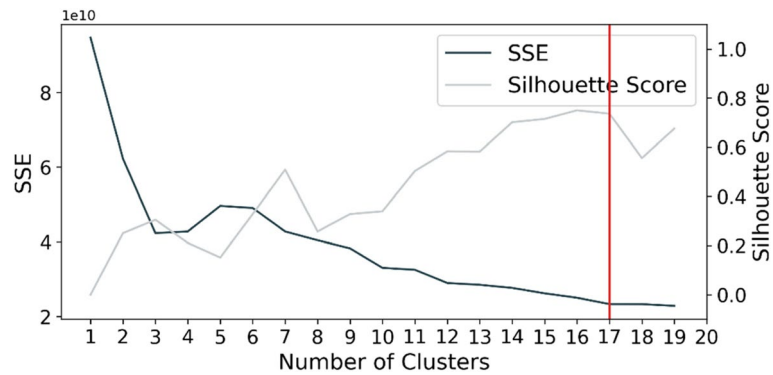


Fig. 5 Clustering results' validation with SSE and Silhouette Score

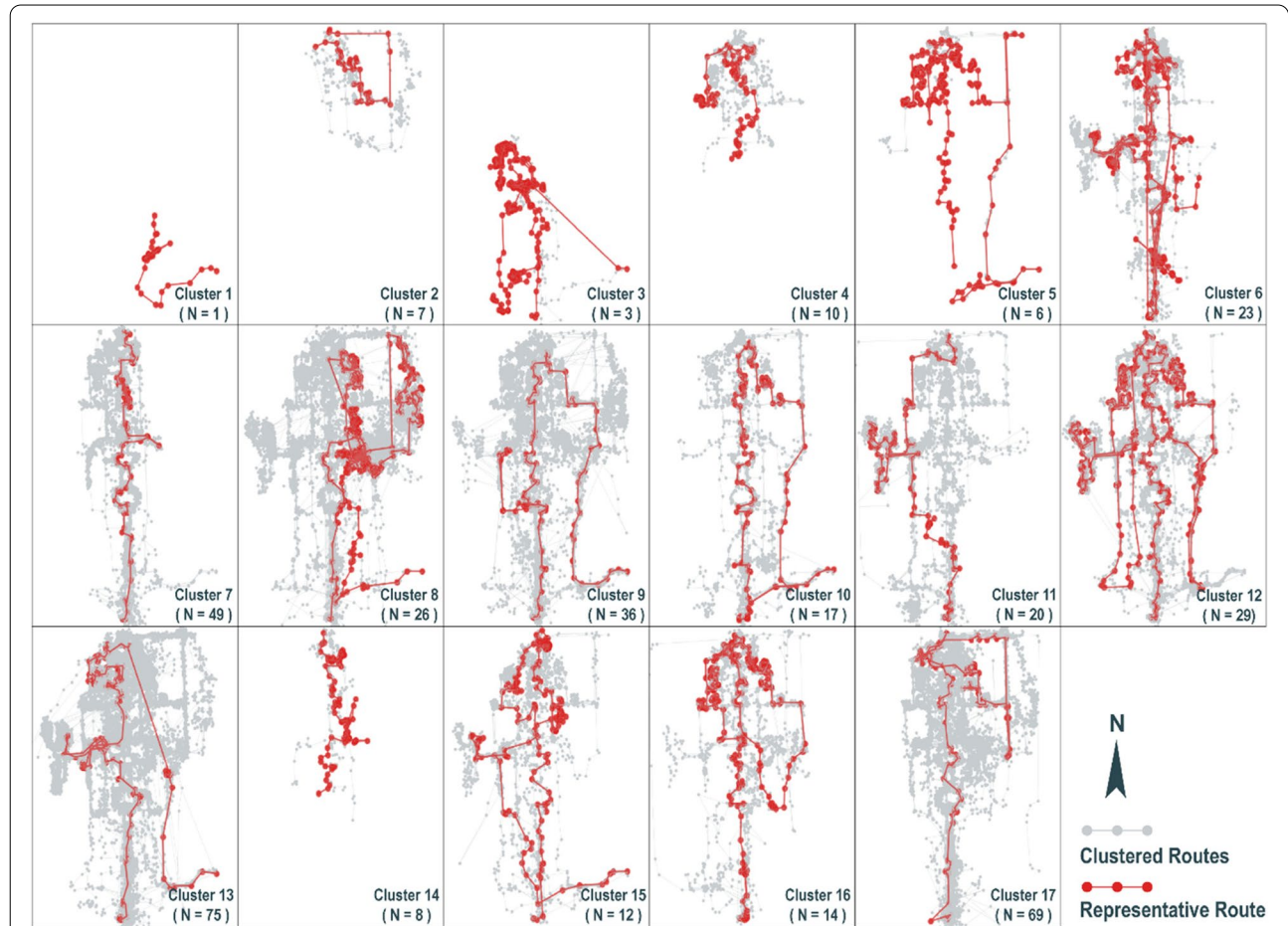


Fig. 6 Clustering results and representative routes of raw-data-based approach

outliers could be well recognized using this clustering method. Clusters 7, 13, and 17 represented three major route types respectively, and the representative trajectories of the three clusters showed that the order of route complexity is as follows: cluster 7 < cluster 17 < cluster 13. It is noted that the plotted representative trajectories were not always that representative when the trajectories in the clustering category were rather dispersed, for example from a geographic perspective, the most evident feature in clusters 8 and 9 was the northeastern part of the trajectories, but it did not appear in the representative route in cluster 9, and the westernmost area which featured in cluster 8 was not shown in the representative route, either. This is because the representative routes were calculated and selected via the program automatically, and characterized by a minimum sum of distances to the other trajectories within their clusters. It was therefore difficult to represent all features of a cluster with a single trajectory, particularly when the trajectories were complex.

The clustering results indicate that although the spatial similarities of trajectories were discovered using this approach, the information provided by such clustering results was vague for complex trajectories. Besides, although there were available methods to help determine the optimal number of clusters, the changes in sample size could strongly affect the stability of the clustering results, making it difficult to draw conclusions that could be generalized to other places. Another weakness of this approach was that the running time of the algorithm was rather lengthy. Nevertheless, the raw-data-based trajectory mining approach was shown to be effective for experimental exploration of newly collected unknown datasets, quick understanding of trajectories' components, and identification of the outliers that cannot be easily recognized in pre-processing.

4.3 Sequence-based approach: common routes identification from a semantic perspective

To differentiate from the previous experiment, this experiment used a supervised method for the test of sequence-based approaches, aiming at identifying the common routes chosen by the visitors from a semantic perspective. Following the pre-processing procedure, we converted all cleaned trajectories and ten recommended tour routes used as alignment references into sequences of attraction areas for pairwise alignment. Eight main attraction areas with a secondary exit were defined and coded according to the original zoning of the Palace Museum, as shown in Fig. 7. The recommended tour routes were classified into three categories (short, medium, and long routes) according to estimated time length, and relative information

including coded sequences, route description, and route code is presented in Table 1.

The sequence alignment method in this study was implemented with a Python package called Bio, which contains a specialized function to compute match scores between sequences, and the match score matrix was computed between each trajectory sequence and each pre-defined route sequence accordingly. The thermal distribution of all sequence alignment results is shown in Fig. 8. The brighter the color, the higher the matching score is. It should be noted that a 100% match does not mean that the two sequences are exactly the same, for example when aligning a short sequence with a long sequence, if the short sequence is the same as part of the long sequence, the matching score will also be high.

As shown in Fig. 8, route S1 and route M1 had the largest numbers of matching results with high matching scores, which covered both short and long trajectory sequences, indicating that these were two mostly frequent subsequences in all trajectory sequences. Route S1 corresponded to the sequence "ABCD", representing the complete central axis of the Palace Museum. With respect to a complex trajectory sequence, it is hard to match a route with a same sequence, but the trajectory likely contains the most important route of the central axis. Regarding route M1, its corresponding sequence "ABCFD" had an additional western area "F" relative to the central axis, indicating that it was a common visiting route in the Palace Museum chosen by a majority of visitors relative to the eastern route. The numbers of matching results for routes L1, L2, and M5 were rather small, in that routes L1 and L2 were too long to match identical trajectory sequences, and route M5, as seen in Table 1, was a less common route specially designed for visitors with particular objectives. This finding suggests that visitors with sufficient time budget would arrange their tour routes more freely rather than strictly following any recommended routes.

The alignment results demonstrate the effectiveness of the sequence-based approach, which well identified frequent subsequences that represented the core routes, and uncovered both the flexibility and fidelity characteristics of the trajectories. To address research questions of diverse granularity, the sequence coding could be altered to extend the applicability of the method. If other information, including visitors' tour duration and demographic characteristics, were analyzed for inclusion, the tendency of route selection under different initial conditions could be further explored and applied to tasks including trajectory prediction and route recommendation, which is significant for flow design and spatial functional planning.

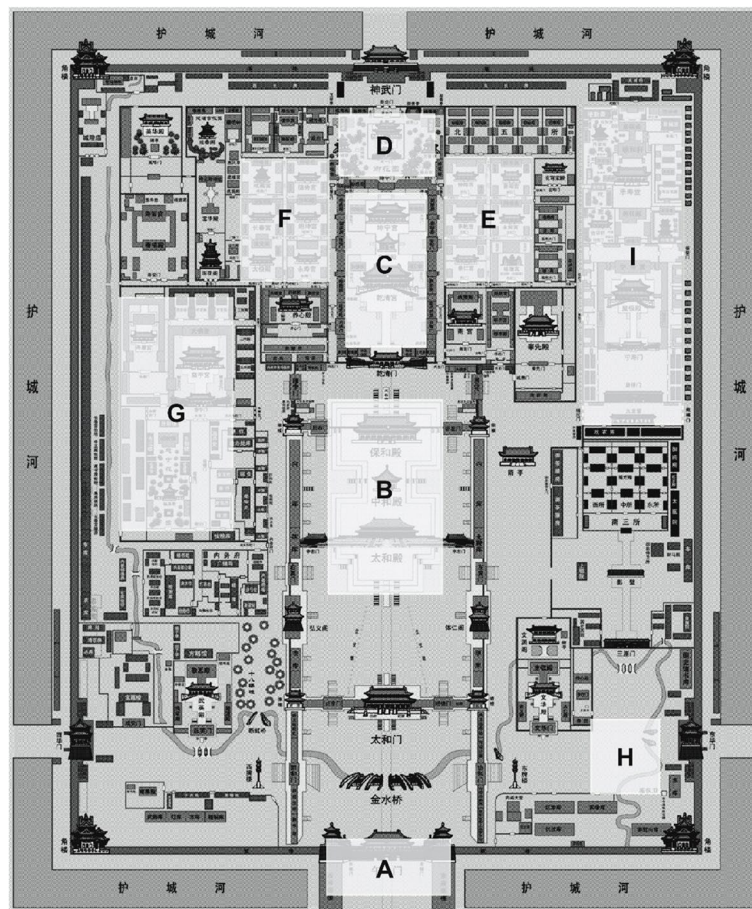


Fig. 7 Coding eight main attraction areas and a secondary exit of the Palace Museum

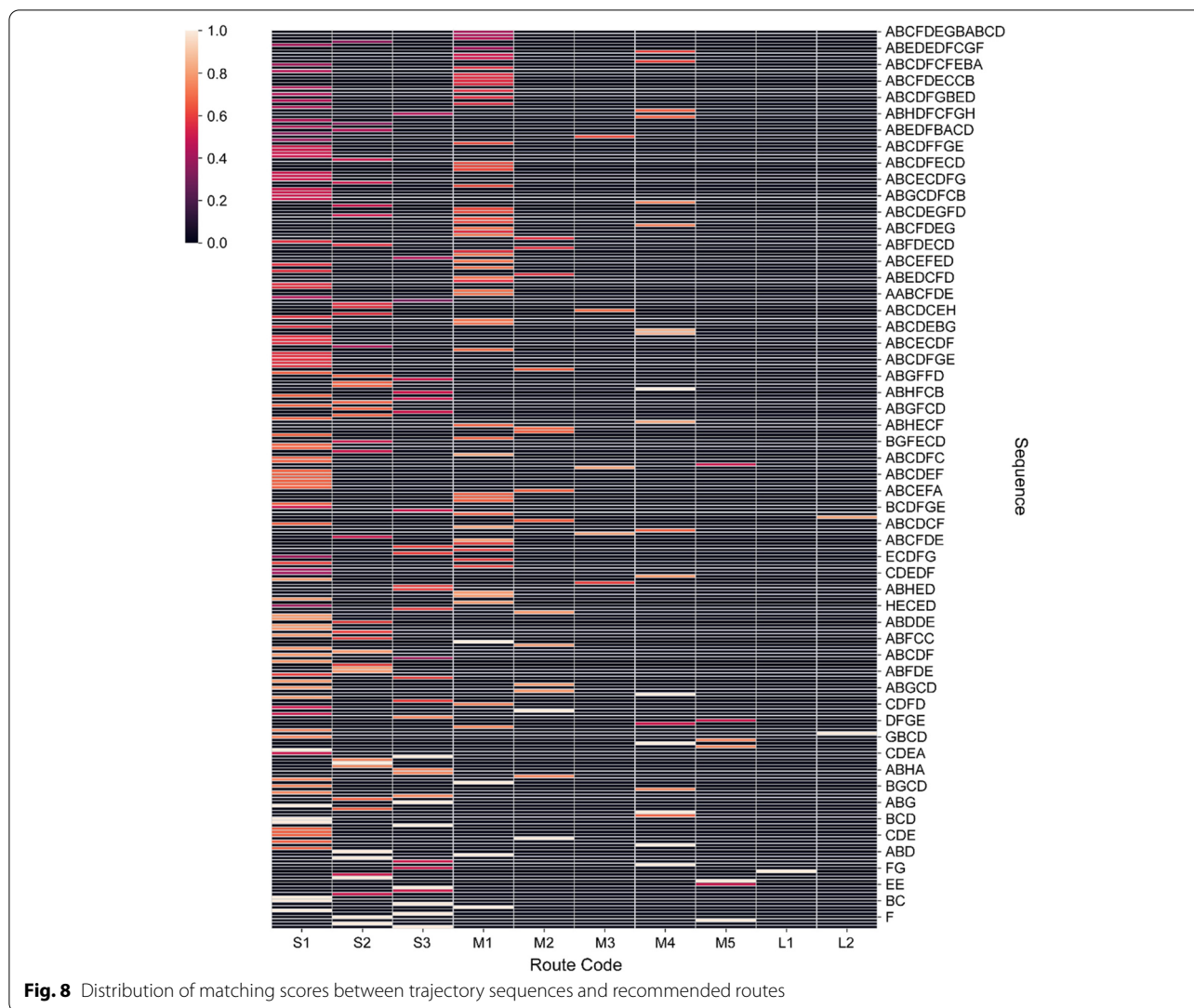
Table 1 Ten recommended tour routes for the Palace Museum

| No. | Sequence | Route Description | Route Code |
|-----|----------|---|------------|
| 1 | ABCD | Fast Tour 1: Central axis | S1 |
| 2 | ABFD | Fast Tour 2: Palaces with original status exhibition | S2 |
| 3 | ABGH | Fast Tour 3: Southern part only | S3 |
| 4 | ABCFD | Half-Day Tour 1: Central axis with western six palaces | M1 |
| 5 | ABCEI | Half-Day Tour 2: Eastern part only | M2 |
| 6 | ABCDH | Half-Day Tour 3: Central axis with southeast exit | M3 |
| 7 | ABGCFD | Half-Day Tour 4: Western part only | M4 |
| 8 | AGDEI | Half-Day Tour 5: Flower viewing | M5 |
| 9 | ABCFDEI | One-Day Tour 1: All attractions without the Cining Garden | L1 |
| 10 | AGFCDEBI | One-Day Tour 2: All attractions | L2 |

4.4 Metrics-based approach: movement behavioral pattern recognition from a quantitative perspective

This experiment selected three basic metrics, including total distance, total duration, and average speed, to describe the movement characteristics of trajectories,

and used them as a feature matrix for trajectory clustering. Since all three metrics were approximately Gaussian distributed, a Gaussian Mixture Model (GMM) using the Expectation-Maximization (EM) framework (Xu & Jordan, 1996) was performed as the clustering algorithm.

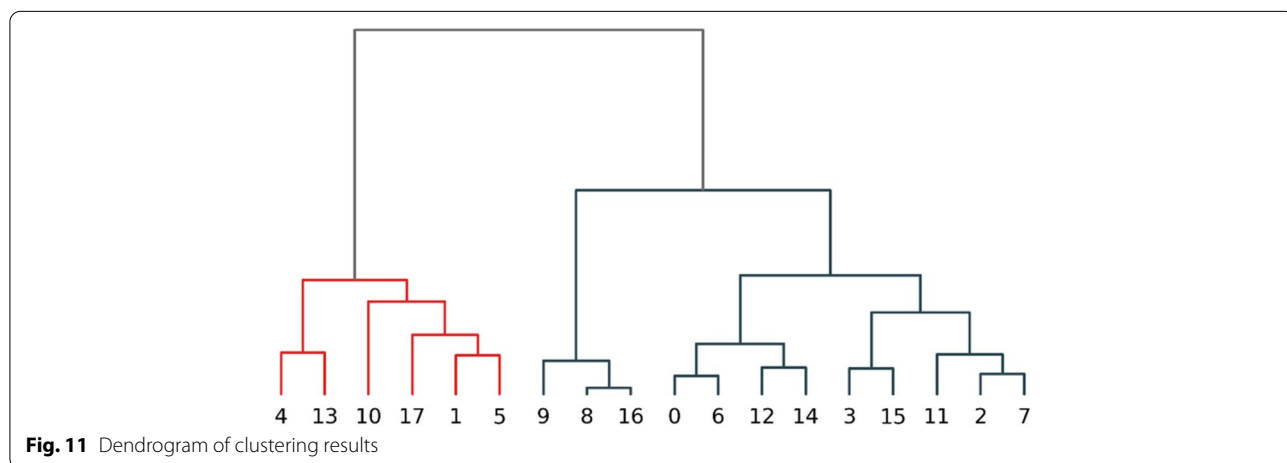
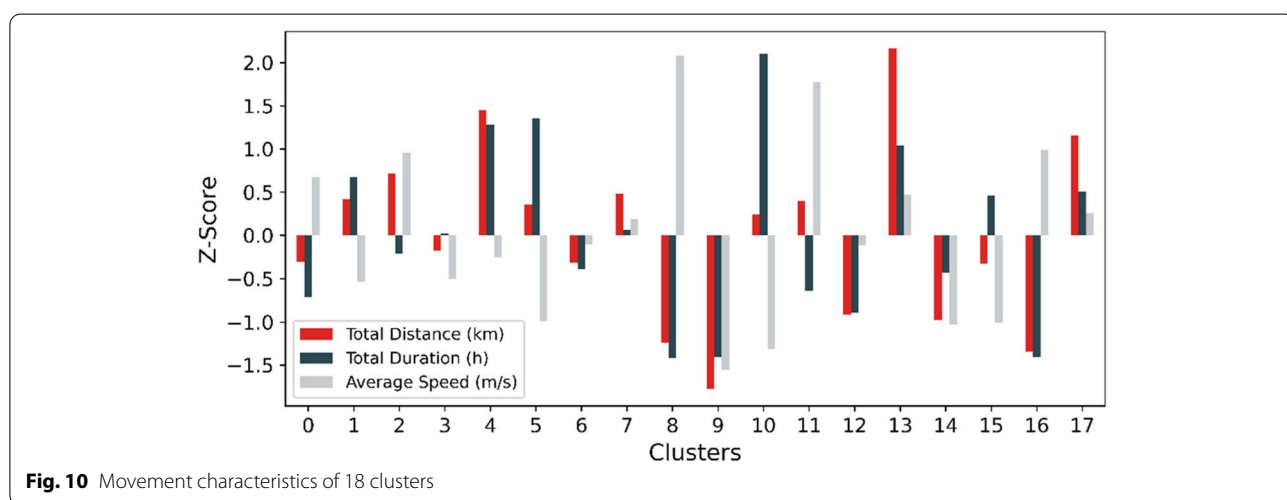
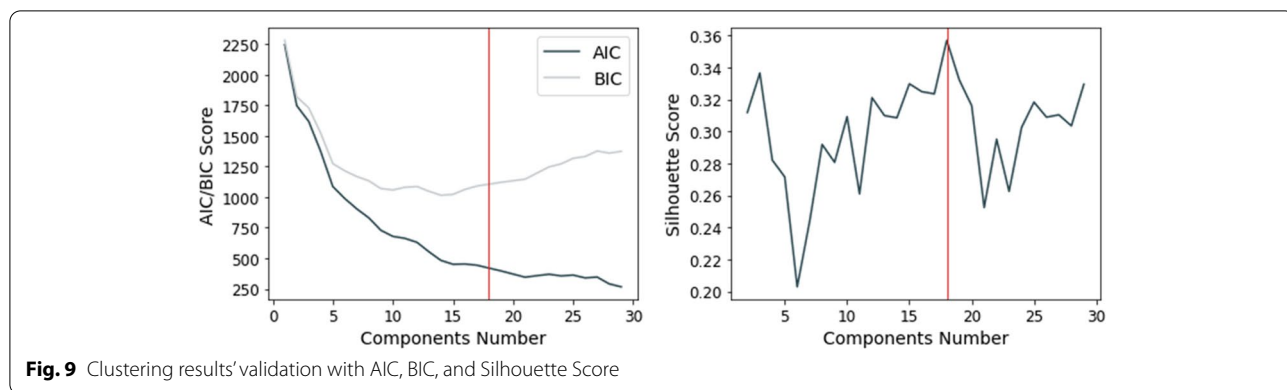


The feature matrix was considered a mixture of multiple Gaussian distributions, and the components of these Gaussian distributions best reproducing the original data are the possible clustering results. GMM is a probabilistic-density-based algorithm, and therefore, two criteria, the Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC) were utilized to test the performance of the fit model. These two criteria helped determine the optimal number of components, and the smaller the AIC/BIC value, the better the model's performance. In addition, this paper used the Silhouette score as a double validation for the quality of the clustering results.

The result of the validation process of clustering model quality by varying the number of components is shown in Fig. 9, and according to three criteria, the optimal number of clusters was 18. The features of 18 clusters were described using the three measures normalized using

the z-score method, as shown in Fig. 10, and a positive or negative value indicates whether the result was above or below the average. Several interesting results were observed, for example, visitors in cluster 13 walked far longer distances relative to the average, but the total duration and average speed were distributed only slightly greater than the average level, not particularly outstanding; while the total duration of cluster 10 exceeded the average by far, but the distance traveled was close to the average length and the average speed was rather slow.

A dendrogram for the 18 clustering results to better observe their categorization and affiliation relationships was calculated and drawn in Fig. 11. The result shows that higher tiers in the hierarchy were divided based on the consistency and the closeness to the average of the distance and duration measures, and lower-tier division was based on the scale of the average speed. Clusters 9, 8, and 16, for example, shared common features that both



distance and duration were much smaller relative to the average level and split over the feature of average speed on the next level, where the speed of clusters 8 and 16 was much faster relative to the average level and cluster

9 was much slower. The consistency of distance and time indicates whether the visitors kept walking or strolled with stopovers, which might be related to the physical strength of the visitors and the purpose of the tour. To

gain relevant conclusions requires further analysis of the clustering results in association with other data regarding human characteristics and spatial features.

As the objective of this experiment was only to verify the effectiveness of the metrics-based approach, it did not dig deeper. Nevertheless, the method has shown great potential due to the good interpretability and high quantification of the clustering results. These uncovered movement behavioral patterns will not lose their meanings as the study site changes, and if supplemented with other semantic information, we could figure out why and how these patterns are formed and discover the associations between human behavioral characteristics, subjective psychology, and other objective influencing factors. These findings could also provide higher dimensional, more quantitative, and more universal strategies for flow design.

5 Discussion and conclusion

This article presented a systematical review of three categories of trajectory mining approaches (i.e. raw-data-based, sequence-based, and metrics-based approaches) and examines the applicability and scope of the three dimensions of the mining approaches, respectively. The results of this study indicate that mass visiting behavior patterns do exist in the Palace Museum, and these patterns were reflected in frequent visiting areas, visiting orders, route planning or selection, and spatio-temporal resource allocation. In other words, it is feasible to employ complex human trajectory data to enhance the understanding of the way people use the built environment and space.

The study aimed to show the applicability and potential contribution of trajectory mining approaches for behavioral analysis. As a result, specific results concerning visitors to architectural sites with similarities to the Palace Museum were limited since extensive explorations of each experiment would be beyond the capability of a single paper. Nevertheless, it is still possible to gather behavioral insights into the great potential of such approaches and provide a wide range of implications. From a methodological point of view, the first and third experiments in this paper used unsupervised clustering methods, while the second experiment exploited an innovative supervised route matching method to classify trajectories by their similarities to the pre-designed routes. This method has not been used in previous studies in the fields of tourism or transportation, although its principle is rather simple. Yet, it is well suited for behavioral studies and flow design evaluation in architectural spaces, since many public buildings possess a sightseeing nature and hence pre-designed flow lines to guide the visitors and improve spatial efficiency. These empirical flow lines

could be used as classification references when analyzing large and complex trajectory datasets. As a result, pre-designed routes with better matching performance could be considered an effective flow design, while trajectories with poor matching results could provide useful insights into individuals' special demands for space usage.

Beyond the methodological insights, the results in this article have a more practical bearing on the reasonable allocation of public resources and could be extended to the intelligent control of building energy services. In this respect, mining trajectories from both a geographic and quantitative perspective are of great value. The geographic perspective could provide spatial and temporal distribution patterns of the users and help avoid operations that would cause waste of resources or energy, while clustering the users based on the movement metrics could facilitate the provision of efficient, energy-saving, and environment-friendly services for various user groups.

The mining approaches used in this study have the potential in uncovering the spatial-temporal behavioral patterns in the field of architecture, but which approach to adopt depends more on the problem to be solved. In general, the higher the dimensionality in utilizing the data, the more effective the approach is in discovering generalizable knowledge of human behavioral patterns. The raw-data-based approach is limited by its overly geographic nature, but it is well suited for addressing problems in a specific space and time, for example how to optimize the setting of environmental facilities in a park or a building complex exactly requires trajectory analysis of this granularity. If we, however, are interested in patterns that are not bound to a specific environment, including mining similar spatial structures, recognizing fixed activity patterns, etc., the sequence-based approach might provide more useful insights due to its semantic advantages. To raise the dimension of the problem, the important role of individuals using a space, including the social-demographic characteristics, psycho-emotional changes, goals, and motivations, cannot be ignored. A more generalized behavioral pattern that can be discussed on the same level as the above issues is then required to be discovered in complex trajectories, where a metrics-based approach from a fully quantitative perspective could be helpful and meaningful.

Another important issue with trajectory pattern mining concerns its relationship with conventional data types. The analyses in this paper revealed a segmentable feature of the visitors' trajectories in the Palace Museum, yet questions such as how multiple trajectory types were generated and which individuals exhibited behavioral consistency are worth exploring in future studies. The current study showed that these questions regarding

individual characteristics and relative semantic information like motivations cannot be achieved using GPS trajectory data alone. Xu et al. (2020) used a K-means clustering algorithm to summarize and classify visitors' spatio-temporal behavior, whose basic idea was similar to the third experiment in this study; however, with a mixed-use of GPS trajectory data and a questionnaire, they further compressed seven clusters into two major spatial patterns and better explained the visitors' behavior. East et al. (2017) also combined GPS and survey data to investigate the effects of the demographic factors exerted on visitors' spatial and temporal behavior and successfully differentiated between the behaviors of different group types. GPS trajectory data and mining techniques could therefore not replace interviews, questionnaires, or field observations which contain important sources of information on human characteristics, psychological motivation, and spatial features (Shoval & Isaacson, 2007). Conversely, trajectory data should be mined along with semantic data to maximize its effectiveness and potential.

Despite the obvious value of GPS trajectory pattern mining, it does have several limitations. The accuracy of GPS, for example, decreases when used indoors, due to satellite signals being blocked by building structures such as reinforced concrete. From this point of view, trajectories collected from more accurate indoor positioning techniques like the mobile phone could be developed deeper in the field of behavioral studies related to the built environment. Other sources or tools for collecting big data positioning data including Wi-Fi IPS and UWB IPS could provide data approximate to trajectory data, and can be applied to the trajectory mining approaches mentioned in the article. In addition to positioning information, human feelings and emotional fluctuations while using the space have great value for research examining environmental psychology. The potential of new emotional monitoring technologies such as portable biometric information measurements should be verified and demonstrated in future research (Shoval et al., 2018).

Another area of importance yet to be explored is data-driven intelligent decision-making concerning architectural programming and smart design. The application of trajectory pattern mining in the context of architectural environments would allow researchers to establish a comprehensive database of human behavior related to the built environments along with spatial information. Typologies of spatial-temporal behavior could be useful in offering suggestions from the design phase and predicting the activity and movement behavior of people in a newly built spatial environment.

Overall, GPS trajectory data and trajectory mining approaches offer an effective means of discovering

behavioral patterns from the spatial-temporal movement data, which, if appropriately applied, could push forward the boundaries of architecture studies. It is worth noting that trajectory mining could also contribute and provide insights to enhance public health and safety. In the current COVID-19 pandemic era, the changes in how people use public buildings and spaces are worthwhile to investigate, which might trigger an even wider enthusiasm for GPS trajectory studies.

Acknowledgments

We would like to thank the Palace Museum for giving consent to collect the data used for this study. We would also like to thank Zhijian Miao and her working group for their assistance with GPS data collection and Jennifer Zhang for her advice in language editing.

Authors' contributions

Weixin Huang: Conceptualization, Funding acquisition, Resources, Supervision, Writing - review & editing. Luying Wang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Investigation, Visualization, Writing - original draft. The author(s) read and approved the final manuscript.

Funding

This work is supported by the National Natural Science Foundation of China [grant No. 52178019].

Availability of data and materials

The datasets generated during and analyzed during the current study are not publicly available due to them containing information that could compromise research participant consent but are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 20 May 2022 Accepted: 17 July 2022

Published online: 26 July 2022

References

- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *Proceedings of the Eleventh International Conference on Data Engineering* (pp. 3–14). IEEE. <https://doi.org/10.1109/ICDE.1995.380415>.
- Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3), 469–486. <https://doi.org/10.1016/j.tourman.2007.05.014>.
- Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic concepts of movement data. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, data mining, and privacy: Geographic knowledge discovery* (pp. 15–38). Springer. https://doi.org/10.1007/978-3-540-75177-9_2.
- Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *KDD Workshop*, 10(16), 359–370.
- Brum-Bastos, V., Long, J., & Demsar, U. (2018). Weather effects on human mobility: A study using multi-channel sequence analysis. *Computers Environment and Urban Systems*, 71, 131–152. <https://doi.org/10.1016/j.compenvurbsys.2018.05.004>.
- Chakka, V. P., Prasad, V., Adam, C., Everspaugh, A. C., & Patel, J. M. (2003). Indexing large trajectory data sets with SETI. In *Proceedings of First Biennial Conference on Innovative Data Systems Research* (pp. 48109–42122). CIDR.
- Chang, D., & Penn, A. (1998). Integrated multilevel circulation in dense urban areas: The effect of multiple interacting constraints on the use

- of complex urban areas. *Environment and Planning B-Planning & Design*, 25(4), 507–538. <https://doi.org/10.1068/b250507>.
- Chen, J., Chen, H., & Luo, X. (2019). Collecting building occupancy data of high resolution based on WiFi and BLE network. *Automation in Construction*, 102, 183–194. <https://doi.org/10.1016/j.autcon.2019.02.016>.
- Cheng, D., Yue, G., Pei, T., & Wu, M. (2021). Clustering indoor positioning data using E-DBSCAN. *ISPRS International Journal of Geo-Information*, 10(10). <https://doi.org/10.3390/ijgi10100669>.
- Cho, N., & Kang, Y. (2019). Identifying staying places with global positioning system movement data using 3D density-based spatial clustering of applications with noise. *Sensors and Materials*, 31(10), 3273–3287. <https://doi.org/10.18494/SAM.2019.2410>.
- Cooper, B. A., Ahrentzen, S., & Hasselkus, B. R. (1991). Post-occupancy evaluation: An environment-behaviour technique for assessing the built environment. *Canadian Journal of Occupational Therapy*, 58(4), 181–188. <https://doi.org/10.1177/000841749105800406>.
- Defays, D. (1977). An efficient algorithm for a complete link method. *The Computer Journal*, 20(4), 364–366. <https://doi.org/10.1093/comjnl/20.4.364>.
- Dodge, S., Laube, P., & Weibel, R. (2012). Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, 26(9), 1563–1588. <https://doi.org/10.1080/13658816.2011.630003>.
- Douglas, D., & Peucker, T. (1973). Algorithms for the reduction of the number of points required to represent a line or its caricature. *The Canadian Cartographer*, 10(2), 112–122. <https://doi.org/10.3138/FM57-6770-U75U-7727>.
- East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, 61, 307–320. <https://doi.org/10.1016/j.tourman.2017.02.021>.
- Ferrante, M., De Cantis, S., & Shoval, N. (2018). A general framework for collecting and analysing the tracking data of cruise passengers at the destination. *Current Issues in Tourism*, 21(12), 1426–1451. <https://doi.org/10.1080/13683500.2016.1194813>.
- Gaffney, S., & Smyth, P. (1999). Trajectory clustering with mixtures of regression models. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 63–72). <https://dl.acm.org/doi/pdf/10.1145/312129.312198>.
- Gelb, A. (Ed.). (1974). *Applied optimal estimation*. MIT Press.
- Grinberger, A. Y., & Shoval, N. (2019). Spatiotemporal contingencies in tourists' intradiurnal mobility patterns. *Journal of Travel Research*, 58(3), 512–530. <https://doi.org/10.1177/0047287518757372>.
- Hadjri, K., & Crozier, C. (2009). Post-occupancy evaluation: Purpose, benefits and barriers. *Facilities*, 27(1/2), 21–33. <https://doi.org/10.1108/02632770910923063>.
- Hallo, J. C., Manning, R. E., Valliere, W., & Budruk, M. (2004). A case study comparison of visitor self-reported travel routes and GPS recorded travel routes. In *Proceedings of the 2004 Northeastern Recreation Research Symposium, GTR-NE-326* (pp. 172–177).
- Han, J., Kamber, M., & Pei, J. (2011). Data mining: Concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 5(4), 83–124.
- Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. (2001). Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *Proceedings of the 17th International Conference on Data Engineering* (pp. 215–224). IEEE. <https://doi.org/10.1109/ICDE.2001.914830>.
- Hershberger, J., & Snoeyink, J. (1992). Speeding up the Douglas-Peucker line simplification algorithm. In *Proceedings of International Symposium on Spatial Data Handling* (pp. 134–143).
- Huang, W., Lin, Y., Lin, B., & Zhao, L. (2019). Modeling and predicting the occupancy in a China hub airport terminal using Wi-Fi data. *Energy and Buildings*, 203, 109439. <https://doi.org/10.1016/j.enbuild.2019.109439>.
- Jin, Y., Yan, D., Kang, X., Chong, A., Sun, H., & Zhan, S. (2021). Forecasting building occupancy: A temporal-sequential analysis and machine learning integrated approach. *Energy and Buildings*, 252, 111362. <https://doi.org/10.1016/j.enbuild.2021.111362>.
- Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises* (pp. 1–227). Dave Kuhlman.
- Kupper, M., & Seyfried, A. (2020). Analysis of space usage on train station platforms based on trajectory data. *Sustainability*, 12(20). <https://doi.org/10.3390/su12208325>.
- Kwan, M.-P., Murray, A. T., O'Kelly, M. E., & Tiefelsdorf, M. (2003). Recent advances in accessibility research: Representation, methodology and applications. *Journal of Geographical Systems*, 5(1), 129–138. <https://doi.org/10.1007/s101090300107>.
- Kwan, M.-P., Richardson, D., Wang, D., & Zhou, C. (Eds.). (2015). *Space-time integration in geography and GIScience*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-9205-9>.
- Lee, J., Han, J., & Whang, K. (2007). Trajectory clustering: a partition-and-group framework. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data (SIGMOD'07)* (pp. 593–604). ACM. <https://doi.org/10.1145/1247480.1247546>.
- Lee, S., Kim, J., & Li, J. (2011). Impacts of store crowding on shopping behavior and store image. *Journal of Asian Architecture and Building Engineering*, 10(1), 133–140. <https://doi.org/10.3130/jaabe.10.1.33>.
- Lee, W.-C., & Krumm, J. (2011). Trajectory preprocessing. In Y. Zheng, & X. Zhou (Eds.), *Computing with Spatial Trajectories* (pp. 3–33). Springer. https://doi.org/10.1007/978-1-4614-1629-6_1.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- Li, Y., Xie, J., Gao, X., & Law, A. (2021). A method of selecting potential development regions based on GPS and social network models—From the perspective of tourist behavior. *Asia Pacific Journal of Tourism Research*, 26(2), 183–199. <https://doi.org/10.1080/10941665.2018.1515092>.
- Lu, Y., Zhao, J., Wu, X., & Lo, S. M. (2021). Escaping to nature during a pandemic: A natural experiment in Asian cities during the COVID-19 pandemic with big social media data. *Science of the Total Environment*, 777, 146092. <https://doi.org/10.1016/j.scitotenv.2021.146092>.
- Meratnia, N., & de By, R. A. (2004). Spatiotemporal compression techniques for moving point objects. In E. Bertino, S. Christodoulakis, D. Plexousakis, V. Christophides, M. Koubarakis, K. Böhm, & E. Ferrari (Eds.), *Advances in Database Technology—EDBT 2004* (pp. 765–782). Springer. https://doi.org/10.1007/978-3-540-24741-8_44.
- Moreno, V., Zamora, M. A., & Skarmeta, A. F. (2016). A low-cost indoor localization system for energy sustainability in smart buildings. *IEEE Sensors Journal*, 16(9), 3246–3262. <https://doi.org/10.1109/JSEN.2016.2524501>.
- Müller, M. (2007). Dynamic time warping. In *Information retrieval for music and motion* (pp. 69–84). Springer. https://doi.org/10.1007/978-3-540-74048-3_4.
- Peng, S., & Maing, M. (2021). Influential factors of age-friendly neighborhood open space under high-density high-rise housing context in hot weather: A case study of public housing in Hong Kong. *Cities*, 115. <https://doi.org/10.1016/j.cities.2021.103231>.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Sankoff, D., & Kruskal, J. B. (1983). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*. Reading: Addison-Wesley Publication.
- Senin, P. (2008). Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1–23), 40.
- Sheng, Q., Wan, D., & Yu, B. (2021). Effect of space configurational attributes on social interactions in urban parks. *Sustainability*, 13(14). <https://doi.org/10.3390/su13147805>.
- Shoval, N., & Isaacson, M. (2007). Tracking tourists in the digital age. *Annals of Tourism Research*, 34(1), 141–159. <https://doi.org/10.1016/j.annals.2006.07.007>.
- Shoval, N., McKercher, B., Birenboim, A., & Ng, E. (2015). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B: Planning and Design*, 42(1), 76–94. <https://doi.org/10.1068/b38065>.
- Shoval, N., Schvimer, Y., & Tamir, M. (2018). Real-time measurement of tourists' objective and subjective emotions in time and space. *Journal of Travel Research*, 57(1), 3–16. <https://doi.org/10.1177/0047287517691155>.
- Tang, L., Zhao, Y., Duan, Z., & Chen, J. (2018). Efficient similarity search for travel behavior. *IEEE Access*, 6, 68760–68772. <https://doi.org/10.1109/ACCESS.2018.2879634>.
- Wang, J., Zhuang, W., Meng, J., et al. (2020). Group discussions: Contemporary city, new human habitat, and architectural design. *Architectural Journal*, 21, 2–27 (in Chinese). <https://doi.org/10.19819/j.cnki.ISSN0529-1399.202003002>.

- Wilson, C. (2008). Activity patterns in space and time: Calculating representative Hagerstrand trajectories. *Transportation*, 35(4), 485–499. <https://doi.org/10.1007/s11116-008-9162-z>.
- Wilson, W. C. (1998). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A*, 30(6), 1017–1038. <https://doi.org/10.1068/a301017>.
- Wu, C.-L., & Carson, D. (2008). Spatial and temporal tourist dispersal analysis in multiple destination travel. *Journal of Travel Research*, 46(3), 311–317. <https://doi.org/10.1177/0047287506304046>.
- Xu, D., Cong, L., & Wall, G. (2020). Visitors' spatio-temporal behavior at a zoo in China. *Asia Pacific Journal of Tourism Research*, 25(9), 931–947. <https://doi.org/10.1080/10941665.2020.1802311>.
- Xu, L., & Jordan, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation*, 8(1), 129–151. <https://doi.org/10.1162/neco.1996.8.1.129>.
- Yang, L., & Huang, W. (2019). Multi-scale analysis of residential behaviour based on UWB indoor positioning system—a case study of retired household in Beijing, China. *Journal of Asian Architecture and Building Engineering*, 18(5), 494–506. <https://doi.org/10.1080/13467581.2019.1682000>.
- Yang, T., Li, M., & Shen, Z. (2015). Between morphology and function: How syntactic centers of the Beijing city are defined. *Journal of Urban Management*, 4(2), 125–134. <https://doi.org/10.1016/j.jum.2015.11.001>.
- Yuan, H., Feng, L., & Qian, Y. (2017). Mining user movement similarity based on massive GPS trajectory data with temporal effects. *Journal of Electronic Commerce Research*, 18(4), 303–316.
- Zhang, X., Cheng, Z., Tang, L., & Xi, J. (2021). Research and application of space-time behavior maps: A review. *Journal of Asian Architecture and Building Engineering*, 20(5), 581–595. <https://doi.org/10.1080/13467581.2020.1800473>.
- Zheng, Y., & Zhou, X. (Eds.). (2011). *Computing with spatial trajectories*. Springer Science & Business Media.
- Zhou, P., Zhang, D., Salzberg, B., Cooperman, G., & Kollios, G. (2005). Close pair queries in moving object databases. In *Proceedings of the 13th Annual ACM International Workshop on Geographic Information Systems (GIS' 05)* (pp. 2–11). ACM. <https://doi.org/10.1145/1097064.1097067>.
- Zhou, Y., Yang, C., & Zhu, R. (2019). Identifying trip ends from raw GPS data with a hybrid spatio-temporal clustering algorithm and random forest model: A case study in Shanghai. *Transportation Planning and Technology*, 42(8), 739–756. <https://doi.org/10.1080/03081060.2019.1675309>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.