ORIGINAL PAPER

Open Access

Understanding the urban life pattern of young people from delivery data

Yining Qiu¹, Jiale Ding¹, Mengxiao Wang¹, Linshu Hu¹ and Feng Zhang^{1,2*}

Abstract

Young people are the backbone of urban development and an important pillar of social stability. The growth of young floating population in China has given rise to urban land expansion. Understanding the urban life pattern of urban life for young people benefits rational and effective land expansion. In this article, we introduce food delivery data into the process of exploring behavioral patterns of urban youth in Hangzhou, Zhejiang Province, China. The dynamic time warping (DTW) distance-based k-medoids method is applied to explore the main activity areas and activity patterns of the urban youth population. The results indicate that many young people from Hangzhou work in Internet companies, and most of work hotspot areas are observed in high-tech parks. The existence of overtime work is proved. Combined with the housing price data in Hangzhou, it is found that young people consider both housing prices and education environment when choosing where to live. The analysis combined with road network data reflects the planning characteristics of the city, also looks into differences between the actual city functions and the planning map. The proposed methods can provide useful guidance and suggestions for city planning.

Keywords: Food delivery data, Urban youth, Dynamic time warping, K-medoids

1 Introduction

Many young people have moved from rural areas to urban areas in the past two decades (Luo et al., 2018). The younger generation of rural migrant workers have become the bulk of the so-called floating population in China (Tang et al., 2020). The increase in the floating-population has given rise to urban land expansion, but expansion without guidelines is undesirable. Rational and effective land expansion is an important part of urban management (Luo et al., 2018). Thus, it is an important task for urban planners to understand the main activity areas and activity patterns of urban youth. Jurji et al. (2018) investigated the participation patterns of youth in physical and sporting activity in Kuala Lumpur so that the government or local authorities can take action for the physical and mental health of youths. Louzado et al. (2021) identified the factors related to the

quality of life of young workers in Northeast Brazil. Matt et al. (2020) explored the measures for young talent attractiveness in urban factories.

With the wide-spread use of smartphone, GPS, and cellular network, massive location-based data was recorded by service providers, including taxi trace data (Liu et al., 2016; Pan et al., 2013), public bicycle data (Bao et al., 2017; Kaltenbrunner et al., 2010; Vogel et al., 2011) and phone call data (Pei et al., 2014; Toole et al., 2012). These data have become an important data source for many scholars to conduct research on urban spatial dynamics perception owing to their benefits on coverage, spatial temporal resolution and cost, helping us to recognize the relationship between human activity patterns and urban function. Joewono et al. (2017) explored the action space of young workers in Bandung City, Indonesia by their travel chain data, providing basic information to anticipate future city development. Besides, some scholars use social media data to explore the behavioral patterns of urban youth. Tsou et al. (2013) used social media and web search engines data to

* Correspondence: zfcarnation@zju.edu.cn

¹School of Earth Sciences, Zhejiang University, Hangzhou 310027, China

²Zhejiang Provincial Key Laboratory of Geographic Information Science, Hangzhou 310028, China

explore the information spreading pattern of youth. Zhi et al. (2016) figured out clusters with significant spatio-temporal characteristics in city by analyzing about 15 million social media check-in data. However, many geo-spatial data that reflect the characteristics of human activities have their limitations. Taxi data is largely dependent on the existence of road network, and taxis are usually used for long-distance trips, accounting for only a small portion of overall crowd flow (Liu et al., 2012). Public bicycle data only reflects information near the supply station, and phone cell data is restricted by the locations of base transceiver stations. Besides, these data contain information on the activities of people of all ages, lacking attention to specific age groups. Social media data has been used for urban youth studies, but social media users are not necessarily a representative sample (Martí et al., 2019).

In recent years, food delivery has become popular among the young generation. In China, the food delivery industry in 2020 had reached a turnover of 603.5 billion yuan. Besides, for the 300 million users of food delivery in China, about 75% users age under 35 years old (Iimedia Report, 2018). Statistics show that most of food delivery users are young people. Yan et al. (2020) classified delivery data time series using Euclidean distance based k-means method to study young people's work and residential areas.

In this article, we used dynamic time warping (DTW) distance based k-medoids clustering algorithm (Chen et al., 2017) to discover the information contained in food delivery data that reflects the behavioral patterns of urban youth. Chen et al. (2017) suggested that the DTW distance based method outperforms the Euclidean distance based method in delineating urban functional areas. We explored the main activity areas and activity patterns of the urban youth population in order to understand the behavior patterns of urban youth, which can support urban planning. The locations of residents and firms affect commuting traffic. Restaurants or entertainment venues targeting younger users want to locate in areas where people congregate. It is important to understand the preferences of young people in choosing where to live and work so that urban planning can be adjusted to help attract young talent to the city.

The remainder of this paper is organized as follows. Section 2 introduces the study area and data source. Section 3 discusses the steps of using DTW distance based k-medoids method to extract urban functional areas. Section 4 analyzes the hotspots of working and living places for young people, and further examines the function distribution properties at the traffic analysis zone (TAZ) scale. The article concludes with a summarize as well as directions for further research in Section 5.

2 Study area and data

2.1 Study area

Urban areas are places with high population densities and attract a large number of young workers (Peri, 2002). In order to explore the job-residence characteristics of young people in cities, we selected six districts in the main city area of Hangzhou in Zhejiang Province, a populous city in the east coast of China, as our study area (Fig. 1). Hangzhou is an ancient capital in the south of the Yangtze River, famous for its beautiful scenery and cultural heritage. Since the twenty-first century, the rise of the digital economy has pulled a new round of urban development in Hangzhou. Relying on Alibaba and other head Internet enterprises, Hangzhou has been at the forefront of digital development in China. The administrative boundaries of Hangzhou were downloaded from AutoNavi's open map API (<https://developer.amap.com/api/webservice/guide/api/district>).

2.2 Data

2.2.1 Grids and road network

To generalize the order data, we use grids and road network as two basic research unit.

The 0.001 degrees (approximately 100 m) grids can provide a higher spatial resolution for further analysis, but grids themselves have limited geographic meanings. Thus, road networks (Fig. 2) were introduced into our study. The road networks were constructed by tertiary roads, which can be approached on OpenStreetMap (<https://www.openstreetmap.org/>).

2.2.2 Food delivery data

Food delivery orders can express young people's job-residence patterns, for the destinations of orders changes when they are at home or company. Our study used 6,104,862 food delivery orders during July 28 and September 26, 2017 (Fig. 3), which were provided by Dianwoda (<https://www.dianwoda.com/>), an instant logistics platform.

The order data contains departure and arrival coordinates, which form an OD pair for every order (Table 1). The latitude and longitude provide spatial information, and the arrival datetime made it possible to compose time series of delivery orders.

2.2.3 POIs

POI (point-of-interest), is a specific point location that someone may find useful or interesting. As POIs have specific social function, the spatial distribution of POI categories, in some extent, can explain the function of regions (Gao et al., 2017).

The POI data was provided by NavInfo, a company with core businesses in HD map and high accuracy positioning.

The POIs are collected in the fourth quarter of 2017 and divided into 15 categories, including dining, public facilities, corporate enterprises, education & culture, finance & insurance, housing & accommodation, etc.

2.2.4 House price data
Young generation in China are under pressure from high housing prices (Ma, 2018). In order to explore the inner connection between where the youth live and the house prices there, we use house prices data from the website

Fig. 3 Kernel density map of destination points in the study area

called 58 (<https://hz.58.com/>), a housing information website to get the house price data in Hangzhou in September, 2017, so that we can evaluate how the housing price affects young generation's resident choices.

3 Methodology

3.1 Time series construction

In order to extract the work and live areas of urban youth, we focus on the delivery destination points. The study area is divided into equal grids with spatial size of 0.001 degrees (approximately 100 m). The destination points have longitude, latitude and arrival time so that can be aggregated at grid level. The hourly value of the count of destination points in a grid is used to form a time series.

To reduce the computational burden, the time series is averaged for weekdays, Saturdays and Sundays. This pattern is chosen based on the dynamic time warping (DTW) (Sakoe & Chiba, 1978) distance between normalized timelines of different days in a week (Table 2). A larger distance indicates a greater difference between

two time series. It is shown that the pattern of Monday, Tuesday, Wednesday, Thursday and Friday are much more similar to each other than to Saturday and Sunday. The large distance between Saturday and Sunday indicates that they have separate patterns. Thus weekdays, Saturday and Sunday form a 72-h series.

As the order data is separated extensive points, we first match the destination points into grids, then all the data will be merged by their timestamp in an hourly basis. As result, we get a 24-dimensional vector V_i for each unit:

$$V_i = [d_0, d_1, \dots, d_{22}, d_{23}] \quad (1)$$

where d_t is the hourly density of arrival order. In this way, a 72-dimensional vector V'_i can be constructed by three 24-dimensional vectors:

$$V'_i = \{V_{w_i}, V_{sa_i}, V_{su_i}\} \quad (2)$$

where V_w represents for a 24-dimensional vector of typical weekday, and V_{sa} represents for a 24-dimensional vector of Saturday while V_{su} represents for a 24-dimensional vector of Sunday.

Table 1 Samples of the delivery order data

ID	Departure Datetime	Departure Lon	Departure Lat	Arrival Datetime	Arrival Lon	Arrival Lat	Duration (s)	Distance (m)
222,169,162	2017/9/16 10:03	120.2138	30.19375	2017/9/16 10:19	120.2109	30.20145	979	900.5919
220,577,611	2017/9/16 7:42	120.209	30.27913	2017/9/16 8:06	120.216	30.25956	1438	2277.223
220,577,783	2017/9/16 8:06	120.1232	30.32045	2017/9/16 9:05	120.1366	30.30968	3535	1756.874

Table 2 DTW distance between normalized timelines of different days in a week

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Mon	0	0.224529	0.151254	0.204261	0.201648	0.957914	1.133888
Tue	0.224529	0	0.155798	0.046153	0.198353	1.100853	1.206497
Wed	0.151254	0.155798	0	0.129142	0.165258	1.032749	1.22839
Thu	0.204261	0.046153	0.129142	0	0.174953	1.072367	1.213247
Fri	0.201648	0.198353	0.165258	0.174953	0	1.035618	1.197916
Sat	0.957914	1.100853	1.032749	1.072367	1.035618	0	0.63582
Sun	1.133888	1.206497	1.22839	1.213247	1.197916	0.63582	0

Last, we normalize the vectors by min-max normalization. The normalized vectors V'_i reflects the unit's weekly trend of food delivery orders and are then used in the DTW distance based k-medoids method.

3.2 DTW distance based k-medoids method

Dynamic time warping is used for similarity measure on time series data. Let $A = \langle a_1, a_2, \dots, a_i, \dots, a_I \rangle$ and $B = \langle b_1, b_2, \dots, b_j, \dots, b_J \rangle$ be two time series. An I -by- J matrix is constructed in which the element d_{ij} indicates the distance between the two points a_i and b_j .

$$d_{ij} = \sqrt{(a_i - b_j)^2} \quad (3)$$

A set of neighboring matrix elements are chosen to form a warping path W , which indicates the mapping from pattern A to B . It is defined as $w_1, w_2, \dots, w_k, \dots, w_K$ where $\max(I, J) \leq K < I + J - 1$ and $w_k = (i_k, j_k)$.

The warping path has three following restrictions, namely monotonic condition, continuity condition and boundary condition. Monotonic condition indicates that $i_{k-1} \leq i_k$ and $j_{k-1} \leq j_k$. Continuity conditions restrict that w_k and w_{k-1} should lie in adjacent cells (diagonally adjacent cells are included). Boundary condition requires $w_1 = (1, 1)$ and $w_K = (I, J)$. So that the warping path starts and finishes in diagonally opposite corner cells of the matrix. Among all the paths that satisfy these conditions, the one that minimize the warping cost is chosen:

$$\min \frac{\sum_{k=1}^K w_k}{K} \quad (4)$$

The warping cost is the DTW distance between two time series. This optimal cost can be calculated using dynamic programming to find minimum of the cumulative distances:

$$D(i, j) = d_{ij} + \min\{D(i-1, j-1), D(i, j-1), D(i-1, j)\} \quad (5)$$

The overall similarity is given by the value of $D(I, J)$. In order to find the similarity, DTW "warps" the time axis

of one (or both) series for a better alignment (Keogh & Pazzani, 2001). However, this warping cannot be unconstrained in our study since we need to focus on the behavioral pattern in specific time periods such as lunch and dinner. Thus, we strengthen the boundary restriction by aligning two time series every six hours.

K-medoids (Kaufman & Rousseeuw, 2009) is a well-known clustering method. It is closely related to the k-means (MacQueen, 1967) algorithm. It is more robust than k-means since its center is an actual object which minimize the dissimilarities among data objects instead of the mean position of objects in the cluster (Arora and Varshney, 2016). In the DTW distance based k-medoids method, the similarity of objects is measured by the DTW distance. Objects that are more similar to each other when the DTW distance between them is smaller.

The reliability of different clustering number k is evaluated base on the silhouette score (Rousseeuw, 1987). For an object i , $a(i)$ is the average distance between i and other objects in cluster C_i . For other clusters $C \neq C_i$, $d(i, C)$ is the average distance from i to objects in C . $b(i)$ is the minimum value among $d(i, C)$ where $C \neq C_i$. $s(i)$ is a criterion to evaluate the extent to which i is more suitable for C_i rather than adjacent clusters:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

A large value of $s(i)$ indicates that i is well assigned. If $s(i)$ is close to 0, then it is uncertain whether i belongs to C_i or other adjacent clusters. A negative value of $s(i)$ indicates that the assignment of i is wrong. The average $s(i)$ of all objects is used to measure the overall quality of clustering results.

3.3 Cluster results validation

POIs can be used to feature the function of an urban region. A POIs dataset assists in discovering functional attributes of different areas. This study uses POI data as auxiliary information to help identify the functional attributes of the clustering results. The enrichment factor (Verburg et al., 2004) is used to measure the relative

abundance of a certain type of POIs. The factor is defined by the occurrence of a POI type of a location relative to the occurrence of this POI type in the whole study area:

$$F_{i,k} = \frac{n_{k,i}/n_i}{N_k/N} \quad (7)$$

$F_{i,k}$ characterizes the enrichment of POI type k in the location i . $n_{k,i}$ is the number of POIs in the k^{th} category in the location i . n_i is the total number of POIs in the location i . N_k is the total number of POIs in the k^{th} category. N is the total number of POIs in the study area. The enrichment factor is larger (or smaller) than 1 when the enrichment of POI type k in the location is larger (or smaller) than the regional average enrichment. The factor is 1 when the enrichment of POI type k in the location equals the national average.

The final result is validated by comparing with the urban planning map released by Hangzhou government (Fig. 4). The land use types mainly include residential, industrial, commercial, educational and green. Commercial and industrial sites are generally used to build office buildings and industrial parks, representing places where people work. Educational land is generally used to build schools, which are mixed-function areas that include places where teachers work and dormitories where students live. Residential land, on the other hand, includes neighborhoods and residential buildings, which are places where people live. Our clustering method mainly

distinguishes between workplace and residential places. Work areas include industrial land, commercial land and educational land. Residential areas include residential land and also educational land.

In order to better detect the urban structure, the cluster results are aggregated in the traffic analysis zone (TAZ) scale. Shannon index is used to characterize the function diversity of TAZs:

$$H = - \sum_i (p_i \log_2 p_i) \quad (8)$$

where p_i is the proportion of function i .

3.4 Cluster results analysis

Hotspot and cold-spot analysis is used in this study based on Getis-Ord G_i^* statistic using fixed distance band in ArcGIS software (Getis & Ord, 1992). The Hotspot Analysis tool calculates Getis-Ord G_i^* statistics for each element in the dataset. With the obtained z-score and p -value, we can know where in space clustering of high- or low-value elements occurs. This tool works by looking at each element in the environment of neighboring elements. To be a statistically significant hotspot, an element should have a high value and be surrounded by other elements that also have a high value. The Getis-Ord local statistic is given as:

Fig. 4 Urban Planning Map for Hangzhou (partial)

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1}}} \quad (9)$$

where x_j is the attribute value for feature j , $w_{i,j}$ is the spatial weight between feature i and j , n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (10)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (11)$$

The G_i^* statistic is a z-score so no further calculations are required.

4 Results

4.1 Cluster results

The appropriate cluster number for the DTW distance based distance k-medoids method is determined using the silhouette score. Figure 5 shows that the highest value is reached when the cluster number k equals 2. Therefore, the optimal cluster number is identified as two.

Figure 6 represents the mean order value distribution over time for the two clusters. Different temporal pattern in weekdays, Saturday and Sunday is demonstrated. Figure 7 demonstrates the location distribution of the

clustered grids. By comparing the temporal patterns of delivery orders inside the areas, we can identify their functional attributes, such as residential, work, etc. The work areas usually have more orders at lunch time than dinner time. The number of orders is also higher on weekdays than on weekends in work areas. The behavioral pattern of residential areas tends to act in the opposite way. According to this hypothesis, functional regions can be recognized.

Cluster 1 shows a typical live area temporal pattern. The order values on rest days, Saturday and Sunday, exceed those on work days. Delivery orders during weekday do not have much difference between lunch time and dinner time, while lunch orders on weekends are much higher than dinner orders. Table 3 shows that the enrichment factors of residential community, resident service and restaurant are high in Cluster 1 areas. This corroborates that Cluster 1 belongs to residential area.

Cluster 2 shows a work-related characteristic. Delivery orders reach a peak at lunch time on weekdays. This illustrates that delivery orders in this area are gathered during working hours. The order quantity from largest to smallest is weekdays, Saturday and Sunday, exactly opposite to which of Cluster 1. The enrichment factor of POIs for enterprise, car sales and service, as well as research institution categories are high in Cluster 2. These points of interests are all places where young people work at.

Table 4 illustrates the calculation result of the overlap rate between cluster results and the official urban planning map. According to the temporal pattern of the two clusters, Cluster 1 is defined as live area and Cluster 2 is defined as work area. The classification results were validated according to the urban planning map, and the accuracy rate is 72.67%.

Fig. 5 Silhouette scores in different numbers of clusters

4.2 Analysis

4.2.1 Work areas

After extracting the work area, which is represented by the grids in Cluster 2, using DTW distance based k-medoids method, we use hotspot analysis to detect areas where young generation gather for work. The hot spot is extracted base on the delivery order count during lunch and dinner peaks during weekdays.

Figure 8 indicates 5 hot spots of work area. Area A is located in the heart of the technology industry center of Hangzhou Economic and Technological Development Area. Hangzhou Hi-tech Enterprise Incubate Park, Zhongzi Science Park and many other high-tech industries are all located in this area. It is close to Xiasha district where many colleges and universities locate in. Relying on university science and education resources

for high-tech industry construction is a very typical and orthodox concept.

Area B includes places such as Zhejiang National University Science Park and Hangzhou East Innovation Center. It is located near Jiubao, which is a large transportation hub that brings together the interwoven road network system of Subway Line 1, Passenger Transportation Center, Bus Rapid Transit, Desheng Expressway, Jiubao Bridge and Yanjiang Avenue.

Area C is situated in Hangzhou Hi-Tech Industry Development Zone in Binjiang, which is an area with a high concentration of Internet companies. The completion of the Binjiang Internet Center is, on the one hand, the result of land expansion and, on the other hand, the need to shift the center of gravity of the entire development of Hangzhou southward, i.e., from the West Lake era to

Table 3 Enrichment factors of POI in clusters

	RR	RC	RT	CS	FI	EC	HS
Cluster 1	1.16324	1.20439	1.13993	0.60188	1.02002	0.96108	1.05959
Cluster 2	0.65130	0.56339	0.70108	1.85045	0.95724	1.08315	0.87271
	SL	PF	CF	RS	EP	TW	RI
Cluster 1	1.02803	0.89933	0.83048	1.18354	0.58225	0.87584	0.78261
Cluster 2	0.94013	1.21505	1.36214	0.60792	1.89240	1.26524	1.46438

Note: RR restaurants, RC residential community, RT retail, CS car sales and service, FI financial and insurance, EC education and culture, HS health and social security, SL spots and leisure, PF public facilities, CF commercial facilities, RS resident service, EP enterprise, TW transportation and warehousing, RI research institution

the Qiantang River era, from the development around the lake to the development across the river.

Area D is located in the central city of Hangzhou. Qianjiang New City, which carries Hangzhou's future urban CBD, is the image portal of Hangzhou's future international business and has many high-end office buildings.

Area E contains West Lake Science and Technology Park, Hangzhou Internet Innovation and Entrepreneurship Park and Hangzhou North Software Park.

Hangzhou is better known for its relatively developed Internet industry. The result illustrates that high-tech Internet companies in Hangzhou provide a large number of jobs for young people. The regional gathering effect is obvious for Hangzhou's Internet industry. The formation of Internet industry clusters contributes to the agglomeration effect of talent at the end.

The current state of overtime work for young people is a topic of wide social concern (Liu et al., 2019). The time pattern of Cluster 2 reveals that delivery during weekdays and Saturday are higher than delivery during Sunday. Thus, places where people work overtime are determined using the order data after 8 pm on weekdays and during the lunch-dinner peak on Saturday.

The result shows that the hot spots where people tend to work overtime is highly coincide with those where young people gather to work (Fig. 9). Overtime work has always been a tradition in the tech industry around the world (Xiaotian, 2019). Since a large number of young people work in Internet-related companies in Hangzhou, the health damage caused by demanding management and the rights and interests of workers should be taken seriously (Xiaotian, 2019).

Table 4 The overlap rate between cluster results and urban planning map

	Grid count	Urban planning map			
		Commercial	Educational	Industrial	Residential
Cluster 1	11,342	12.96%	11.24%	7.94%	67.86%
Cluster 2	17,736	18.14%	20.03%	30.40%	31.43%

4.2.2 Live areas

The number of delivery orders during peak dining hours on weekends is used to extract hotspots for accommodation (Fig. 10). The main difference between the residential hotspots and the work hotspots is that while many young people choose to work in the city center, few live in this area.

Figure 11 indicates that a reason why young people choose to not live in the central city might be that the house price is too high to afford. They choose to live in places where housing prices are relatively lower. 70.37% percent of living hotspot belong to areas where house prices are lower than 48,484 yuan every square meter. The only exception is Wenjiao district, which is a residential hotspot but is located in the area with highest prices.

Wenjiao district in the west of the city is home to a number of famous elementary schools and middle schools in Hangzhou. It is the top educational gathering place in Hangzhou. High-quality educational resources have driven up the price of housing in this area. The strong cultural environment, mature facilities and location advantages also make this area popular among young people. This phenomenon shows that in the context of Hangzhou's multi-center development, the city center, where educational resources are concentrated, still has a strong attraction for young people.

4.2.3 Traffic analysis zone (TAZ) level analysis

Function detection based on grids is capable of analyzing urban functional areas at a fine-grained level. As the basic unit of urban function, urban management and urban cognition, traffic analysis zone (TAZ) is one of the important elements in urban planning and constitutes the basic unit of residents' life and urban environment. TAZ-level analysis may contain more socio-economic information.

Shannon index is used to evaluate the function diversity in different TAZs (Fig. 12a). The hot spots and cold spots are shown in Fig. 12b. Area A contains regions with a high degree of mixed functions, basically clustered around the downtown area. The central city undertakes various functions such as living, administrative

Fig. 8 Hot spot and cold spot of work area in the study area

offices, commerce and finance, tourism services, science and technology education, culture and entertainment, high-tech industries, etc. Area B is in Xiasha district, which is a comprehensive new city with Hangzhou Economic and Technological Development Zone and High Education Park as the backbone. The new city has an obvious planned structure, with the education and research area in the north, the industrial area in the south and west, as well as the residential and living area in the central and eastern riverside areas.

Figure 13 shows the difference between incoming orders and outgoing orders in TAZs during different time periods. This helps us understand the dynamic “source-link” structures of information landscape. (Gao, 2015) The visualization result indicates that the number of incoming orders and outgoing orders are in a balanced state. It shows that most of the TAZs in Hangzhou are able to reach a self-sufficient state in the road network planning. The number of areas where incoming order count exceeds outgoing order count is greater than the number of areas where outgoing order count is larger. This phenomenon indicates that the outgoing orders are concentrated in some areas and they provide orders for the surrounding areas.

5 Discussion and conclusion

Our study has the following practical and empirical implications. First, the time pattern of food delivery data can be used to detect work areas and live areas of urban youth. Monitoring and analysis using food delivery data

facilitates the rapid detection of change and can provide urban planners with up-to-date information on the work and residential preferences of young people by targeting specific age groups. Second, urban function detection using the grid has a higher spatial resolution and can better reflect the planning characteristics of the city when combined with the functional analysis of the TAZ as the basic unit. Finally, the methodology used in this study can be more widely applied to explore the functional attributes and hotspots of urban areas using spatial data consisting of time series.

This study proves that the time curve formed by the food delivery data can reflect the functional characteristics of the region’s employment and residence. The DTW distance based k-medoids method is able to cluster the regions with similar function. The temporal pattern in the work area is characterized by higher orders for lunch than for dinner, and higher orders on weekdays than on weekends. The temporal pattern in the residential area is characterized by a smaller difference between lunch orders and dinner orders, and higher orders on weekends than on weekdays.

After the division of work and residence areas, the number of delivery orders during peak meal times can be used for hot-spot analysis to explore the hotspot areas where young people work and live. The results show that Internet companies in Hangzhou provide a large number of employment positions for young people, and most of work hotspot areas are observed in high-tech parks. The aggregation of take-out orders at

(a) Hot spot and cold spot of work area at night

(b) Hot spot and cold spot of work area on Saturday

Fig. 9 Hot spot and cold spot of work overtime areas

different times of the day can also reflect the work overtime situation of young people in the area. Combined with the housing price data in Hangzhou, it is found that young people consider both housing price and location when choosing where to live. Most young people choose to live in suburban areas, where prices are lower. The city center, where prices are high but educational

resources are excellent, is also a hot spot, indicating that young people value the education of the next generation.

The functional analysis combined with the road network data can better reflect the planning characteristics of the city, as well as compare whether there are differences between the actual city functions and the planning map. TAZ-level analysis recognizes the hot

Fig. 10 Hot spot and cold spot of live area in the study area

Fig. 11 The house price map overlay living hotspots

(a) Shannon index of TAZs

(b) Function diversity hot spots and cold spots

Fig. 12 TAZ level function diversity in the study area

spot and cold spot of function diversity. The findings indicate a high degree of functional mix in the old city, while the new city has a more complete plan and a low degree of functional mix. By comparing the data on food deliveries to and from each TAZ, it is found that most of the TAZs are able to achieve self-sufficiency.

The recognition accuracy of the regional function in the study results still needs to be improved. When the function complexity of the study unit is high, the proportion of each function cannot be extracted. The study mainly focuses on the hot spots where young people work and live. It is also possible to dig further into their work-life. In future studies, we plan to incorporate

Fig. 13 The flow of delivery orders at the TAZ scale

multiple sources of human activity information to gain further insight into the working and residential lives of young people, such as commuting routes, etc. Mixed land use remains a major challenge in research, and future research could focus more explicitly on the problem of misclassification in the identification of mixed functional areas.

Code availability

Code can be found here: <https://github.com/qyn0729/Food-delivery-data-analysis>.

Authors' contributions

Conceptualization, Y.Q., M.W., L. H and F.Z.; data curation, Y.Q.M.W. and L.H.; methodology, Y.Q., J.D. M.W. and L.H.; formal analysis, Y.Q. and J.D.; writing—original draft preparation, Y.Q. and J.D.; writing—review and editing, F.Z., and M.W.; project administration, F.Z.; Supervision, F.Z.; funding acquisition, F.Z. All authors have read and agreed to the published version of the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the National Key R&D Program of China (2017YFB0503604).

Availability of data and materials

Food delivery data and points of interest (POI) data presented in this study are available on request from the corresponding author. The data are not publicly available due to they are commercial data available for purchase. Publicly available datasets were analyzed in this study, the road network data can be found here: <https://www.openstreetmap.org/>, and the housing price data can be found here: <https://hz.58.com/>.

Declarations

Conflict of interest

The authors declare no conflict of interest.

Received: 13 May 2021 Accepted: 22 November 2021

Published online: 09 December 2021

References

- Arora, P., & Varshney, S. (2016). Analysis of k-means and K-Medoids algorithm for big data. *Procedia Computer Science*, 78, 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>
- Bao, J., Xu, C., Liu, P., & Wang, W. (2017). Exploring Bikes sharing travel patterns and trip purposes using smart card data and online point of interests. *Networks and Spatial Economics*, 17(4), 1231–1253. <https://doi.org/10.1007/s11067-017-9366-x>
- Chen, Y., Liu, X., Li, X., Liu, X., Yao, Y., Hu, G., Xu, X., & Pei, F. (2017). Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning*, 160, 48–60. <https://doi.org/10.1016/j.landurbplan.2016.12.001>
- Gao, S. (2015). Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spatial Cognition and Computation*, 15(2), 86–114. <https://doi.org/10.1080/13875868.2014.984300>
- Gao, S., Janowicz, K., & Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3), 446–467. <https://doi.org/10.1111/tgis.12289>
- Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Iimedia Report. (2018). 2017–2018 China Online Catering Food Market Research Report. <https://www.iimedia.cn/c400/60449.html>
- Joewono, T. B., Santoso, D. S., Adinegoro, L., & Kharisma, A. H. (2017). Characteristics of travel, activities, and action space of young workers riding motorcycles in developing city. *Transportation Research Procedia*, 25, 5023–5039. <https://doi.org/10.1016/j.trpro.2017.05.202>
- Jurji, N. M., Kasuma, J., Rahman, D., Shahrinaz, I., & Aren, M. (2018). Youth participation patterns in physical activity and sports: The study in urban poverty area in Kuala Lumpur. *Journal of Fundamental and Applied Sciences*, 10(June), 1149–1160. <https://doi.org/10.4314/jfas.v10i5s.96>
- Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4), 455–466. <https://doi.org/10.1016/j.pmcj.2010.07.002>
- Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to cluster analysis (Vol. 344). John Wiley & Sons.
- Keogh, E. J., & Pazzani, M. J. (2001). Derivative dynamic time warping. In *Proceedings of the 2001 SIAM International Conference on Data Mining* (pp. 1–11). <https://doi.org/10.1137/1.9781611972719.1>
- Liu, B., Chen, H., Yang, X., & Hou, C. (2019). Why work overtime? A systematic review on the evolutionary trend and influencing factors of work hours in China. *Frontiers in Public Health*, 7, 1–15. <https://doi.org/10.3389/fpubh.2019.00343>
- Liu, X., Kang, C., Gong, L., & Liu, Y. (2016). Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, 30(2), 334–350. <https://doi.org/10.1080/13658816.2015.1086923>
- Liu, Y., Kang, C., Gao, S., Xiao, Y., & Tian, Y. (2012). Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems*, 14(4), 463–483. <https://doi.org/10.1007/s10109-012-0166-z>
- Louzado, J. A., Cortes, M. L., Oliveira, M. G., Bezerra, V. M., Mistro, S., de Medeiros, D. S., Soares, D. A., Silva, K. O., Kochergin, C. N., de Carvalho, V. C. H. D. S., Amorim, W. W., & Mengue, S. S. (2021). Quality of life and associated factors in young workers. *International Journal of Environmental Research and Public Health*, 18(4), 1–13. <https://doi.org/10.3390/ijerph18042153>
- Luo, J., Zhang, X., Wu, Y., Shen, J., Shen, L., & Xing, X. (2018). Urban land expansion and the floating population in China: For production or for living. *Cities*, 74, 219–228. <https://doi.org/10.1016/j.cities.2017.12.007>
- Ma, X. (2018). Housing challenges in Chinese cities under urbanization. In *Urbanization and Urban Governance in China* (pp. 161–181). Palgrave Macmillan, New York. https://doi.org/10.1057/978-1-137-57824-2_8
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.
- Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, 161–174. <https://doi.org/10.1016/j.compenvurbsys.2018.11.001>
- Matt, D. T., Orzes, G., Rauch, E., & Dallasega, P. (2020). Urban production – A socially sustainable factory concept to overcome shortcomings of qualified workers in smart SMEs. *Computers and Industrial Engineering*, 139, 105384. <https://doi.org/10.1016/j.cie.2018.08.035>
- Pan, G., Qi, G., Wu, Z., Zhang, D., & Li, S. (2013). Land-use classification using taxi GPS traces. *IEEE Transactions on Intelligent Transportation Systems*, 14(1), 113–123. <https://doi.org/10.1109/TITS.2012.2209201>
- Pei, T., Sobolevsky, S., Ratti, C., Shaw, S., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28(9), 1988–2007. <https://doi.org/10.1080/13658816.2014.913794>
- Peri, G. (2002). Young workers, learning, and agglomerations. *Journal of Urban Economics*, 52(3), 582–607. [https://doi.org/10.1016/S0094-1190\(02\)00510-7](https://doi.org/10.1016/S0094-1190(02)00510-7)
- 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)
- Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>
- Tang, S., Hao, P., & Feng, J. (2020). Consumer behavior of rural migrant workers in urban China. *Cities*, 106, 102856. <https://doi.org/10.1016/j.cities.2020.102856>
- Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012). Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, 1–8. <http://dx.doi.org/10.1145/2346496.2346498>
- Tsou, M. H., Yang, J. A., Lusher, D., Han, S., Spitzberg, B., Gawron, J. M., Gupta, D., & An, L. (2013). Mapping social activities and concepts with social media (twitter) and web search engines (yahoo and Bing): A case study in 2012 US presidential election. *Cartography and Geographic Information Science*, 40(4), 337–348. <https://doi.org/10.1080/15230406.2013.799738>

- Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28(6), 667–690. <https://doi.org/10.1016/j.compenvurbsys.2003.07.001>
- Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia - Social and Behavioral Sciences*, 20, 514–523. <https://doi.org/10.1016/j.sbspro.2011.08.058>
- Xiaotian, L. (2019). The 996. icu movement in china: Changing employment relations and labour agency in the tech industry. *Made in China Journal*.
- Yan, Y., Wang, Y., Du, Z., Zhang, F., Liu, R., & Ye, X. (2020). Where urban youth work and live: A data-driven approach to identify urban functional areas at a fine scale. *ISPRS International Journal of Geo-Information*, 9(1). <https://doi.org/10.3390/ijgi9010042>
- Zhi, Y., Li, H., Wang, D., Deng, M., Wang, S., Gao, J., Duan, Z., & Liu, Y. (2016). Latent spatio-temporal activity structures: A new approach to inferring intra-urban functional regions via social media check-in data. *Geo-Spatial Information Science*, 19(2), 94–105. <https://doi.org/10.1080/10095020.2016.1176723>

Publisher's Note

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