Chin. Geogra. Sci. 2023 Vol. 33 No. 1 pp. 21−35
https://doi.org/10.1007/s11769-022-1320-2

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

Multi-source Data-driven Identification of Urban Functional Areas: A Case of Shenyang, China

XUE Bing[1, 2], XIAO Xiao[1, 3], LI Jingzhong[2, 4], ZHAO Bingyu[1, 3], FU Bo[1, 3]

(1. *Institute of Applied Ecology*, *Chinese Academy of Sciences*, *Shenyang* 110016, *China*; 2. *Key Lab for Environmental Computation and Sustainability of Liaoning Province*, *Shenyang* 110016, *China*; 3. *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*; 4. *College of Urban Planning and Architecture*, *Xuchang University*, *Xuchang* 461000, *China*)

Abstract: Urban functional area (UFA) is a core scientific issue affecting urban sustainability. The current knowledge gap is mainly reflected in the lack of multi-scale quantitative interpretation methods from the perspective of human-land interaction. In this paper, based on multi-source big data include 250 m × 250 m resolution cell phone data, 1.81×10^5 Points of Interest (POI) data and administrative boundary data, we built a UFA identification method and demonstrated empirically in Shenyang City, China. We argue that the method we built can effectively identify multi-scale multi-type UFAs based on human activity and further reveal the spatial correlation between urban facilities and human activity. The empirical study suggests that the employment functional zones in Shenyang City are more concentrated in central cities than other single functional zones. There are more mix functional areas in the central city areas, while the planned industrial new cities need to develop comprehensive functions in Shenyang. UFAs have scale effects and human-land interaction patterns. We suggest that city decision makers should apply multi-sources big data to measure urban functional service in a more refined manner from a supply-demand perspective.

Keywords: human-land relationship; multi-source big data; urban functional area; identification method; Shenyang City

Citation: XUE Bing, XIAO Xiao, LI Jingzhong, ZHAO Bingyu, FU Bo, 2023. Multi-source Data-driven Identification of Urban Functional Areas: A Case of Shenyang, China. *Chinese Geographical Science*, 33(1): 21−35. https://doi.org/10.1007/s11769-022-1320-2

1 Introduction

Urban land functions usually refer to the urban land used for industry, transportation, commerce, education, housing and green space (Zhou et al., 2016), showing spatial hierarchical differentiation under the influence of human activities (Pei et al., 2014). Therefore, urban functional area (UFA) is the typical manifestation of the interaction between human activities and physical space (Shen and Karimi, 2016). For a long time, a series of studies focused on the connotation, identification, and application of UFA (Farmer and Fotheringham, 2011; Kuang et al., 2021). Among them, the identification of UFAs (Yuan et al., 2015; Hu et al., 2020) and the exploration of its influencing factors are technical hotspots, which are essential for evaluating the mixed functions of urban land (Vorontsova et al., 2016), identifying the urban spatial structure (Wang et al., 2020), optimizing the allocation of facility resources (Liu et al., 2003), formulating employment and transportation policies (Novak et al., 2013), solving urban environmental problems (Tian et al., 2010), and thus promoting urban socio-economic development (van de Voorde et al., 2011). It is an essential prerequisite for the city's social and economic development and serves as an essential supporting role in territorial spatial planning and

Received date: 2022-05-14; accepted date: 2022-09-09
Foundation item: Under the auspices of Natural Science Foundation of China (No. 41971166)
Corresponding author: XUE Bing. E-mail: xuebing@iae.ac.cn
© Science Press, Northeast Institute of Geography and Agroecology, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2023

sustainable development decision-making at different scales (United Nations, 2015).

The research paradigm of UFAs was originally formed in the last century, represented by the concentric circle model (Burgess, 2008), the sector model (Hoyt, 1939) and the multi-core model (Harris and Ullman, 1945), which describe the functional structure of land use. Since then, case studies of urban land use function and planning design in practice have emerged, which mainly rely on urban planning data, land use data and remote sensing images (Berry and Kasarda, 1977; Liu and Gu, 2008; Zhang et al., 2016; Kuang et al., 2021). However, the UFA is an important area of modern human activities. It is a complex that includes land material space and human activities, forming a typical regional system of human-land relationship (Liu et al., 2007). In this new era, the research on the human-land system needs to transform from a single-factor pattern to a multi-factor and multi-scale spatial pattern and provide scientific support for the sustainable development regulation (Kuang, 2020). Therefore, the cognition of UFAs should break through a single perspective such as land use and strengthen the multi-scale cognition from the perspective of human activities (Shi et al., 2021). Moreover, it can help urban planners discover urban problems and formulate regulatory strategies by exploring the relationship between land use and human activities.

Actually, there has been an increasing number of studies reflecting land functions from human activities (Goodchild et al., 2010; Martin et al., 2013). Nevertheless, their research's main methods were questionnaire surveys or population censuses (Xue et al., 2016; Sweet et al., 2017). The timeliness and representativeness of the survey samples have been widely discussed (Louail et al., 2014; Crooks et al., 2015; Hu et al., 2020). In recent years, with the continuous development of information technology, the new geographic big data represented by POI, mobile phone signaling, and Location Based Services (LBS) data with basic characteristics of large sample size, rapid data production and diverse types, showed great potential in urban quantitative research (Batty, 2013). Some studies have attempted to use these data to quantitatively identify UFAs (Novak et al., 2013; Shearmur, 2015; Liu et al., 2016). For example, Li et al. (2021) developed a block-scale UFA identification method based on POI data. However, POI data reflects the spatial location of social and economic sectors without the human activities information. Other studies have conducted empirical study using data that reflect human activity. For example, Ahas et al. (2010) used continuous 12-month mobile phone call data to identify about 45% of users' work and residence; Wei et al. (2020) measured urban functional polycentricity from the perspective of human mobility based on taxi GPS data. Cell phone signaling data were considered to have the characteristics of low information loss rate, continuous dynamic recording, involuntary and high holding rate. Therefore, compared with taxi GPS data, bus card data and other fragmented human activity data reflecting minority groups, cell phone data show obvious advantages in the research on the identification of UFAs from the perspective of population distribution and activity trajectory (Pei et al., 2014; Mao et al., 2017; Markonis et al., 2021). However, most of the research units in previous studies are administrative units such as streets, which is difficult to help policy makers implement refined spatial control strategies (Yang et al., 2021). Little study has been devoted to exploring multi-type complex land functions, which is essential for enhancing urban vitality and promoting city smart growth (Xue et al., 2020a). Meanwhile, the identification of UFAs is still inseparable from 'land'. The delineation of functional areas requires to be based on land parcels (Kuang and Yan, 2018). There is a multi-process intrinsic mechanism linking the UFAs reflected by various human activities and the geographic physical elements (Novak et al., 2013). It is also necessary to deeply analyze the formation mechanism of UFAs in the context of the human-land interaction process (Estima and Painho, 2016). Therefore, how to analyze human-land interaction mode of functional areas is crucial to understanding urban functions and accelerate sustainable spatial regulation. 'Land observation' big data such as collaborative mapping platforms (Goodchild, 2007; Jiang et al., 2015; See et al., 2016) or remote sensing data (Mariathasan et al., 2019), has a large sample size, which is time-saving and low-cost to obtain. So, it is usually used to accurately describe the current distribution of urban land use (Novak et al., 2013; Kashian et al., 2019).

Nevertheless, there is still a lack of empirical analysis on the multi-type and multi-scale identification methods of UFAs and the human-land interaction process.

This problem restricts the transformation of UFA research from simple land use description to multi-scale spatial pattern analysis and understanding of complex human-land interaction processes and is not conducive to multi-source big data serving urban spatial decision-making. As typical regional complexes, old industrial cities usually adopt spatial planning adjustment measures to promote their sustainable transformation (Estima and Painho, 2016; Dong et al., 2020), which is directly manifested in the distribution of UFAs (Lee and Newman, 2017). Taking Shenyang City, an old industrial city in Northeast China, as our study area, we built a multi-source big data-based UFA identification methodology. Then we investigated the spatial distribution of multi-scale functional areas as well as its interrelationship with urban land-use facilities. Our results could be beneficial to research attempting to design more comprehensive methodology to understand the urban spatial structure and helpful to policy makers formulating more scientific urban planning strategies.

2 Materials and Methods

2.1 Study area

Shenyang, located in Liaoning Province, China, has experienced different periods in history, such as a frontier military town, a subsidiary of the South Manchurian Railway, a national heavy industrial base, and a city in transition (Xu et al., 2019). Currently, Shenyang is also the only mega-city in Northeast China with an administrative area of 1.29×10^4 km^2 and a resident population of 8.32×10^6 in 2018 (Shenyang Statistics Bureau, 2020). It now has ten municipal districts, two counties and one county-level city under its jurisdiction. We choose the core area of Shenyang (hereinafter referred to as 'Shenyang Proper') as the research area, including five central districts (Heping, Shenhe, Dadong, Huanggu, and Tiexi) and four suburban districts (Hunnan, Shenbei New District, Yuhong, and Sujiatun) (Fig. 1).

The residential population and the employees' information were extracted from the cell phone signalling data for one consecutive month in July 2018 as the original data. The cell phone signalling data consists of both the active signalling data and passive signalling data. Active signalling data are generated when users switch on and off their cell phones, make calls, send and receive SMS, or carry their cell phones to move their

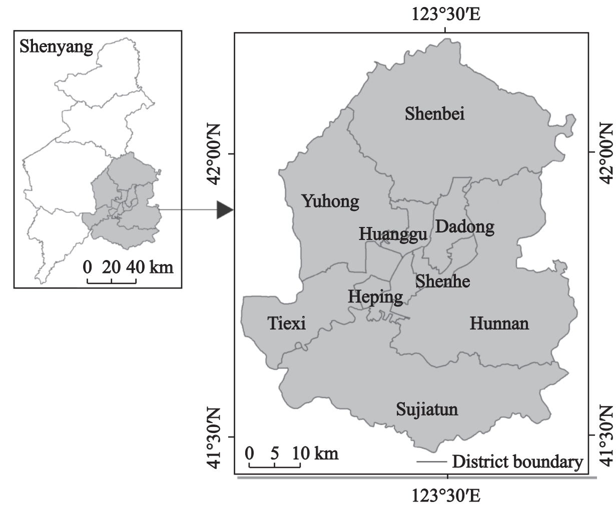

Fig. 1 Location of study area

location so that the base station connected to the cell phone changes. The passive signalling data generate by linking the cell phone to the base station occasionally. The algorithm is deployed in the mobile operator's platform. It first measured the workplace and residence of individual users, and then took the 250 m × 250 m resolution grid as the statistical unit to get the number of people working and living in each grid. The heat map, obtained from operators, 1.51×10^7 items in total, representing the total population appearing in each grid per hour per day, has been chosen to calculate the leisure population in each grid. The operator obtained the heat map from the active and passive cell phone signalling data collected from July 16th to July 22nd, 2018, by the algorithm deployed at the operator. We selected non-holiday heat map to calculate the leisure population. We clipped the employment population, residential population and heat map stored in grides form by the administrative boundary data and unified the projection coordinate system as WGS-84 coordinate system, UTM.

The calculation method of individual user's workplace is to calculate the active and passive observed base station locations of users aged 16–64 years old during the observation period (9:00–17:00) and to measure the actual location of users based on the weighted extrapolation of base station locations and frequencies (referred to as 'base station weighted center of mass algorithm'). The calculation method of individual user's residence is to calculate the active and passive observed base station locations of users during the observation period (21:00–8:00 the next day) and to measure the actual location of users based on the base station weighted center of mass algorithm. The 'heat map' is based on the active and passive cell phone signalling data received by each grid in each hour of the day, and the user's actual location square is determined according to the weighted mass algorithm of the base station. The above algorithm is based on continuous active and passive cell phone signalling for one month. Compared with the previous short-term (1–2 weeks) and only one type of signalling (active or passive signalling) data, its identities the location of individual users more accurately by increasing the number of signalling data samples. At the same time, some constraints have been added in speculating the workplace and residence of individual users to accurately locate the user's workplace and residence. For the first constraint, the algorithm automatically added up the frequency of the user's occurrence at each location within a month and then taking the most frequent location as the user's location. For the second constraint, the occurrence of the most frequent location should be more than 10 d. Thus, the algorithm can effectively filter out the positioning errors of individual users and further improve the accuracy of the number of people working and living in each grid. Using the above method, we finally identified about 2.32 million residents and 1.43 million workers in Shenyang City. There is a strong correlation between the residential population data of each district and the government demographic data (Shenyang Statistics Bureau, 2020), indicating that the method of speculating residential people and working people by cell phone data is feasible (Fig. 2).

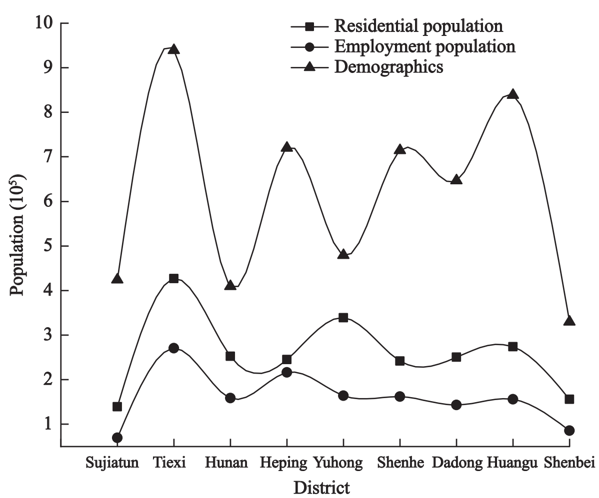

Fig. 2 Comparison of the number of population identified by cell phone data and the demographics in different district of Shenyang City, China

The Points of Interest (POI) data of Shenyang City were mainly collected from Amap, with a total of 4.90×10^5 items and recorded the name, type, latitude, and longitude information of urban socio-economic sector entities item by item. After data cleaning, a total of nine types of urban facility POI data are used to analyse the interrelationship between population activities and urban infrastructure, namely daily services (convenience stores, beauty salons, intermediaries), medical care (healthcare services), public services (public facilities), transportation (bus stations, parking lots), restaurants (food services), shopping (shopping malls, hypermarkets, supermarkets), attractions (scenic spots), accommodation (hotels, hostels, guest houses), and financial offices (excluding ATMs). Finally, we got a total of 1.81×10^5 city facility POIs. The data of road network includes expressways, ring roads, national roads, provincial roads, railroads, and county roads. The socio-economic data includes the demographic data of Shenyang City in 2019 (Shenyang Statistics Bureau, 2020). The 'Land use planning map of the central urban area of Shenyang', 'Spatial structure planning map of central urban area' and other planning documents were collected from the 'Shenyang City's Master Plan (2011–2020)' (Chinese Urban Planning Society, 2016).

2.2 Methods

2.2.1 Concept development

The flowchart of the proposed research framework is illustrated in Fig. 3. Based on the residential and working population numbers in grids obtained by cell phone signalling and other auxiliary data, we measured the residential and employment population densities at different scales. Then, by integrating cell phone data, heat map data, boundary data and demographic data, we calculated the leisure population and mapped the distribution of residential population, employment population and leisure population at different scales. Then we extracted multi-scale and multi-type UFAs of Shenyang City by the multi-level functional area division method we constructed. Finally, with the support of POI data, we calculated the correlation between active population density and the urban facilities density.

2.2.2 Spatial unit division

The district units are the basic administrative units of urban research and can reflect the overall urban functional layout at the macro-level (Tian et al., 2010). The

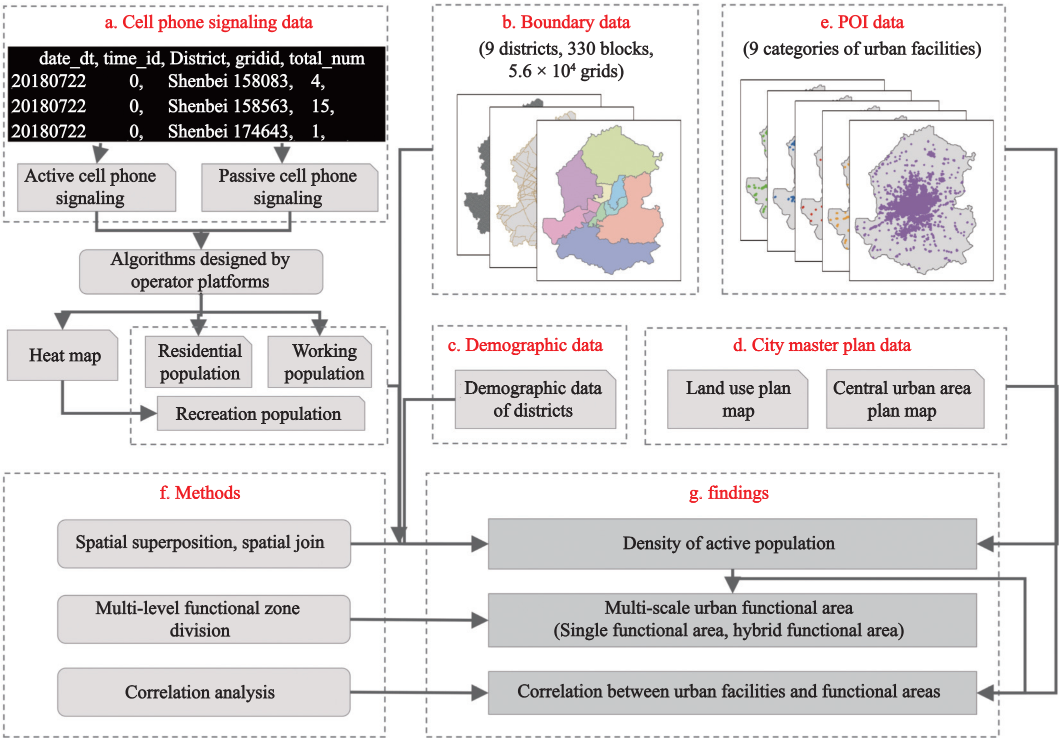

Fig. 3 A framework for identifying and understanding urban functional areas based on multi-source data

block units are the basic support for urban morphological structure cognition (Xue et al., 2020b). The grid unit can characterize the detailed of land functions from a more microscopic perspective (Crooks et al., 2015). Therefore, this study takes the grid, block, and district as three different unit-scale (Fig. 3). The grid and district units were obtained directly from the cell phone signalling data, while the block units were divided based on the study area boundary data and road data. The road space is not within the scope of this study object, so roads areas were excluded from the study area. According to the average width of different roads, the roads area at all levels, i.e., highway, ring road, national highway, provincial highway, railroad, county road, were widened by 40 m, 40 m, 40 m, 40 m, 30 m, and 20 m, respectively. At the same time, the unoccupied green land within the ring road were excluded too. Finally, a total of 330 blocks units were formed for functional area classification.

2.2.3 Functional area identification method

The spatial join tool of ArcGIS 10.3 is applied to obtain the number of residential and employment population maps at different scales. Referring to the study by Niu et al. (2014), we considered 15:00 on weekends as the most intensive period for leisure activities. So, the number of leisure population in each grid was obtained by calculating the average number of people during 15:00–16:00 on Saturdays and Sundays. To distinguish the types of mixed functional areas more accurately and to reduce the identification errors of non-leisure areas where users were located during the rest time, three indicators were used to identify different categories of functional areas. The indicators include residential population density, employment to residential population density ratio, leisure population to residential population density ratio at 15:00 on weekends (Table 1). Niu et al. (2014) classified the functional areas in the central city of Shanghai into high, medium, and low residential population densities, and only selected areas with 'high' and 'low' population densities for the study. However, the research area of this study consists of both the central and the suburban areas, if the population density is only classified as 'high' and 'low', the large suburban areas with 'medium' population density will be ignored. Therefore, the medium-density residential areas are considered in the functional area identification method. The residential population density values were classified into three levels including high (H), medium (M) and low (L) by the natural interruption method (Table 1). The ratio of employment to residential population density less than or equal to 1 is defined as 'l^a' and greater than 1 is defined as 'h^a'. The ratio of leisure population density to residential population density at 15:00 on weekends less than or equal to 1 is defined as 'l^b' and greater than 1 is defined as 'h^b' (Table 1). According to different combinations of residential population density level (RPDL), the ratio of employment to residential population density level (RERPDL), and the ratio of leisure to residential population density level (RLRPDL), we identified different functional area. For example, when RPDL is H, RERPDL is h^a, and RLRPDL is h^b, the area is defined as a 'residential-employment-leisure mixed functional area with high residential population density', represented by 'H: R-E-L'. In Fig. 3, we refer to mixed functional areas with different residential population density levels as 'mixed functional areas'. Using the spatial join tool of ArcGIS 10.3, the ratio of employment to residential population density and the ratio of leisure population density to residential population density at 15:00 on weekends were calculated for different spatial units. Through the calculation of three indicators, the areas were divided into residential functional areas (R), employment functional areas (E), leisure functional areas (L) as well as mixed functional areas (R-E, R-L, E-L, R-E-L) composed of different functions. The mixed functional areas were divided into more detailed types (see the last column of Table 1) according to the residential population density level within each unit.

Table 1 Identification criteria of Shenyang City's urban functional area in July 2018

Residential population density	Ratio of employment to residential population density	Leisure population to residential population density ratio at 15:00 on weekends	Functional area
H	h^a	h^b	H: R-E-L
	h^a	l^b	H: R-E
	l^a	h^b	H: R-L
	l^a	l^b	H: R
M	h^a	h^b	M: R-E-L
	h^a	l^b	M: R-E
	l^a	h^b	M: R-L
	l^a	l^b	M: R
L	h^a	h^b	L: E-L
	h^a	l^b	L: E
	l^a	h^b	L: L
	l^a	l^b	L: R

Notes: High level (H), Medium level (M), Low level (L)

3 Results

3.1 Characteristics of the population activity distribution

In order to analyse the spatial distribution of population activities, we divided Shenyang into three circles: 1) the central urban area within the Third Ring Road, 2) the suburban area outside the Third Ring Road and within the Third Ring Road, and 3) the remaining area outside the Third Ring Road. The population density of residence, employment, and leisure is divided into five levels by the natural interruption method. The population density of residence, employment and leisure at grid scale gradually decreased from center to periphery in the central urban area. There are several high-density centers in the suburbs and many high-density small-scale centers in the suburban area (Fig. 4). This distribution characteristics of the population density is consistent with the urban structure pattern (i.e., 'main city + sub-city + multi-center') proposed in the Shenyang City's Master Plan (2011–2020). The distribution characteristics of residential and leisure population densities are similar, with population densities ranging from 0 to 24 224 persons/km^2 and 0 to 20.368 persons/km^2, respectively. The population density decreased in a stepwise manner along the First, Second and Third Ring Roads. The employment population density values ranged from 0 to 40 496 persons/km^2, showing a more significant concentricity and a steep decline outside the Second Ring Road.

3.2 Multi-scale single-mixed functional area identification results

At a grid scale, the single functional area covered 2.24×10^3 km^2, which was dominated by the residential functional area (H: R, Table 1) with 1.55×10^3 km^2, the leisure functional area (L:L, Table 1) with 515 km^2 and the employment functional area (L: E, Table 1) with 175 km^2. The mixed functional area covered 254 km^2 and was dominated by the employment-leisure functional area (L: E-L) with 222 km^2, followed by the mixed residential-leisure functional area (R-L, including M: R-E and H:R-E) with 17 km^2, the mixed residential-employment-leisure functional area (R-E-L, including H: R-

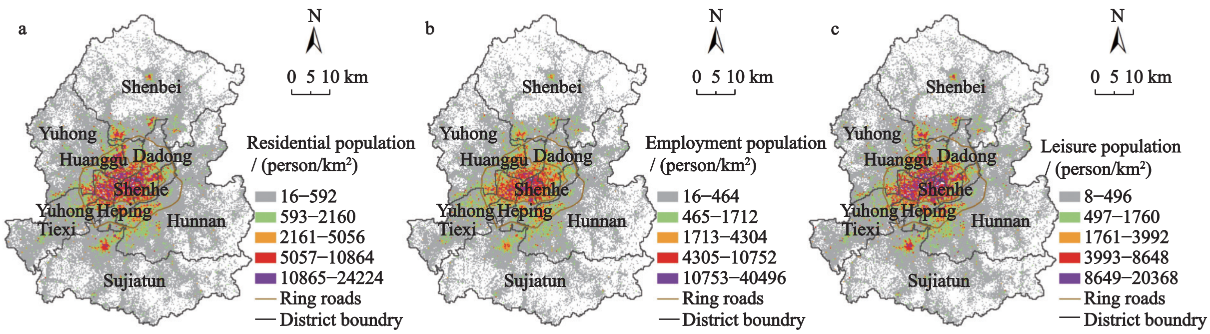

Fig. 4 Residential (a), employment (b) and leisure (c) population density distribution at grid scale in Shenyang City, China

E-L and M: R-E-L) with 13 km^2, and the mixed residential-employment functional area (R-E, including H: R-E and M: R-E) with 2.4 km^2. Single functional areas were widely distributed (Fig. 5a), while mixed functional areas had the characteristic of gathering in the central urban area (Fig. 5b). Especially, the area of R-E-L within the Second Ring Road accounted for 89% of all the R-E-L all over Shenyang City. This is mainly because the central urban area was densely populated with residential and office places and the populations were highly mixed. While the population in the near and far suburban areas was relatively sparse and the different types of functional spaces were dispersed. In addition, single and mixed functional areas appeared to be synergistically

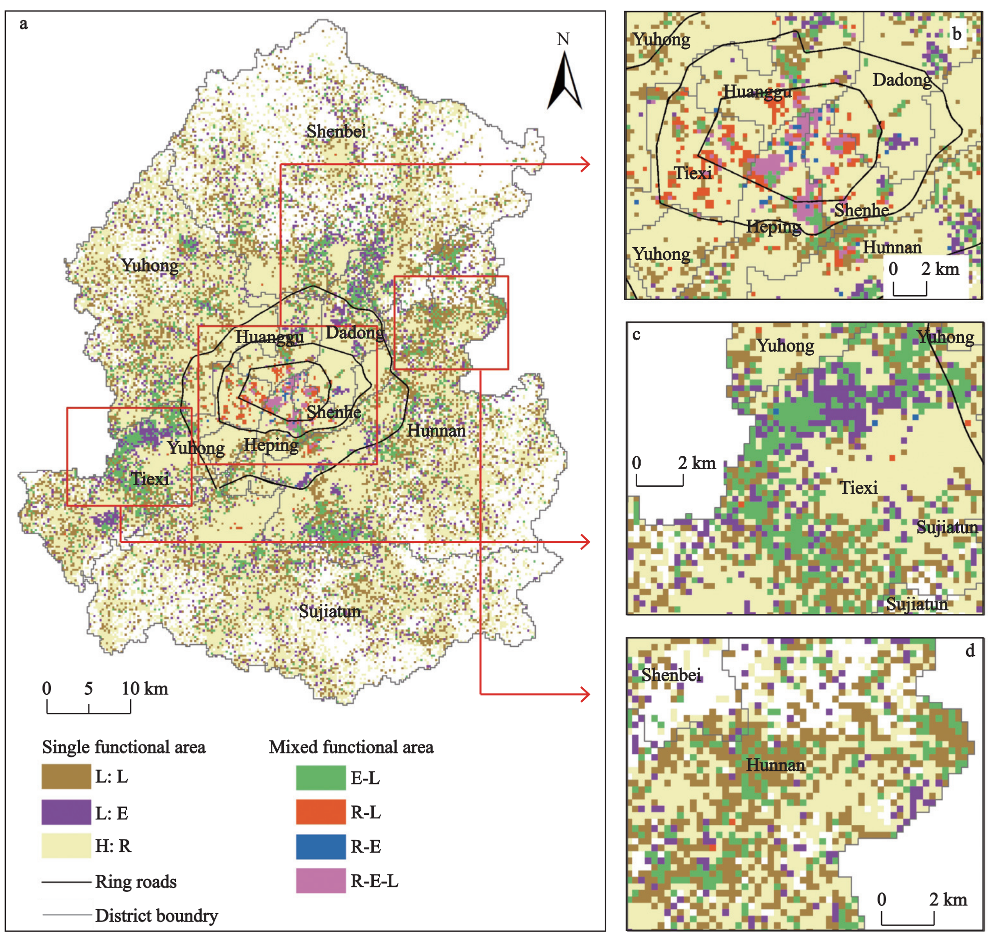

Fig. 5 Spatial distribution of single functional areas include leisure functional area with low residential population density (L: L), employment functional area with low residential population density (L: E) and residential functional area with high residential population density (H: R) and mixed functional areas includes employment-leisure functional area (E-L), residential-leisure functional area (R-L), residential-employment functional area (R-E) and residential-employment-leisure functional area (R-E-L) in Shenyang City, China

clustered in the near and far suburbs (Fig. 5c, Fig. 5d).

At a block scale, the single functional area, only leisure functional area (L: L), covered 3.04 × 10^3 km^2 and widely distributed in the periphery of the city. The mixed functional area covered 389 km^2 (Fig. 6a) and was dominated by H: R-L (283.7 km^2), followed by the E-L that includes M: E-L and H: E-L (95.3 km^2) and the R-E-L that include H: R-E-L and M: R-E-L (9.6 km^2). The H: R-L were mainly distributed in the central urban area, extending southward along Shenyang City's central axis (i.e., Youth Street) to the suburbs. The E-L were strolling in the suburbs in the shape of 'satellites'. The R-E-L mainly located along Taiyuan Street and Zhongjie Street that were proposed to be the future 'municipal centers' in the Shenyang City's Master Plan (2011–2020)'. They were planned to be developed into a comprehensive service center. At a district scale, it contained three kinds of UFAs—L:L, H: R-L and M:R-L (Fig. 6b). The L:L locate in the northern and southern Shenyang, the M:R-L and H: R-L were in the eastern and western Shenyang, reflecting that the residential function was strong in the west, weak in the east, and weaker in the north and south. There were few mixed functional areas at any scale. But there were more residential-leisure mixed functional areas. The residential function and employment function tended to be distributed independently.

Comparing the identification results from the points of scales, we found that the functional area types at block scale or district scale are less than those at grid scale. The functions tend to be diverse and mixed as the study unit becomes larger, showing a scale effect (Fig. 7). The small-scale study unit accommodates a smaller pop-

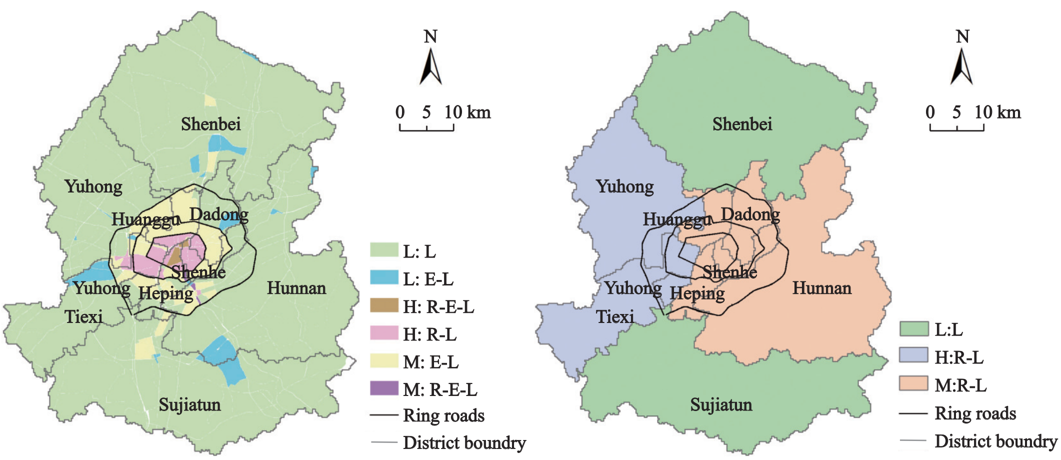

Fig. 6 Spatial distribution of single functional areas (leisure functional area with low residential population density, L: L) and mixed functional areas include employment-leisure functional area with low residential population density (L:E-L), residential-employment-leisure mixed functional area with high residential population density (H: R-E-L), residential-leisure mixed functional area with high residential population density (H: R-L), employment-leisure mixed functional area with medium residential population density (M: E-L), residential-employment-leisure mixed functional area with medium residential population density (M: R-E-L) and residential-leisure mixed functional area with medium residential population density (M: R-L) (left: block scale, right: district scale) in Shenyang City, China

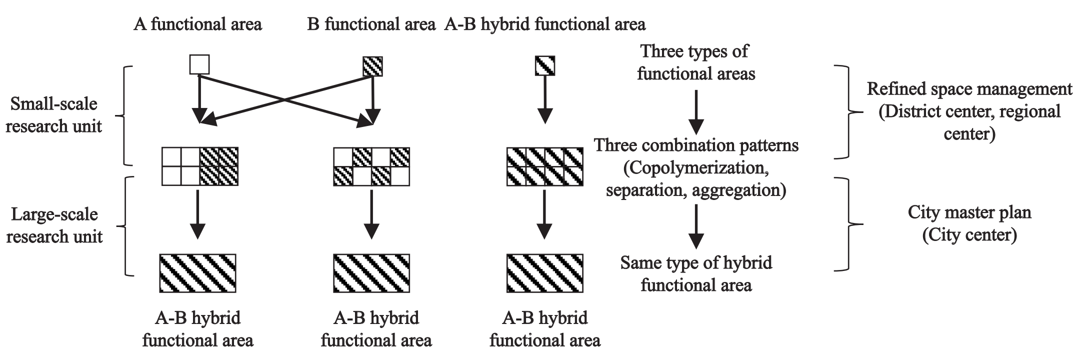

Fig. 7 Schematic diagram of the scale effect of urban functional areas

ulation and reveals diverse local functional details, which is helpful for refined space management. While at the bigger scale, local functional area details are not easy to be discovered, one block or district is more likely to be identified as mixed functional area. However, block-scale UFAs can be compared with traditional urban spatial structure mode, which helps to discover new urban structure theory in newly urbanized area and promotes the urban master planning.

3.3 Spatial pattern of single functional area

At a grid scale, the residential functional areas were becoming increasingly scarce from the center to the periphery. There are few residential functional areas in the areas that were planned to be built into 'new cities'. The H: R were concentrated in the central urban area. The M: R decrease from the First Ring Road outward (Fig. 8a). The L: R were widely distributed outside the Third Ring Road. The residential functional areas mainly correspond to residential land of the land use planning map (Fig. 8b), showing that the identification results of UFA are reliable. There were some multiple clusters of employment functional areas in suburban industrial new city, economic development zones and high-tech industrial zones. These employment functional areas corresponded to the industrial land and storage land of the land use planning map, which proves the high reliability of the identification results. The big employment functional areas were mostly surrounded by L: R, so the industrial new city area should strengthen the comprehensive function development to promote the internal circulation operation of the new city itself. There were few leisure functional areas in the central urban area, but more in the suburbs.

3.4 Spatial pattern of mixed functional areas

The distribution of R-E-L was consistent with planning expectations, which was driven by policy forces. There were four contiguous gathering centers of R-E-L in the central city (Fig. 9). Through comparison with digital maps and field surveys, it shows that the four gathering centers located in Shenyang North Railway Station, Taiyuan Street, Xiaodong Road and Northeastern University area respectively. These areas were going to be built into 'regional comprehensive service industry centers' in the 'Shenyang City's Master Plan (2011–2020)'.

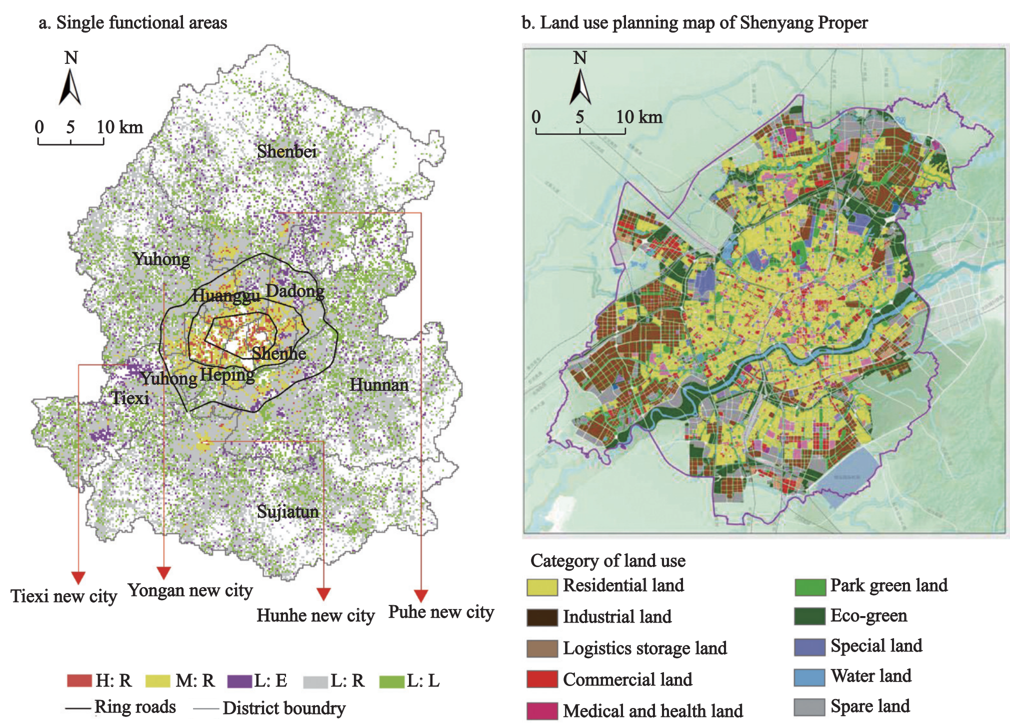

Fig. 8 Comparison between the single functional areas and the land use planning map of central city of Shenyang City, China. The single functional areas include residential functional area with high residential population density (H:R), residential functional area with medium residential population density (M:R), employment functional area with low residential population density (L:E), residential functional area with low residential population density (L:R) and leisure functional area with low residential population density (L:L)

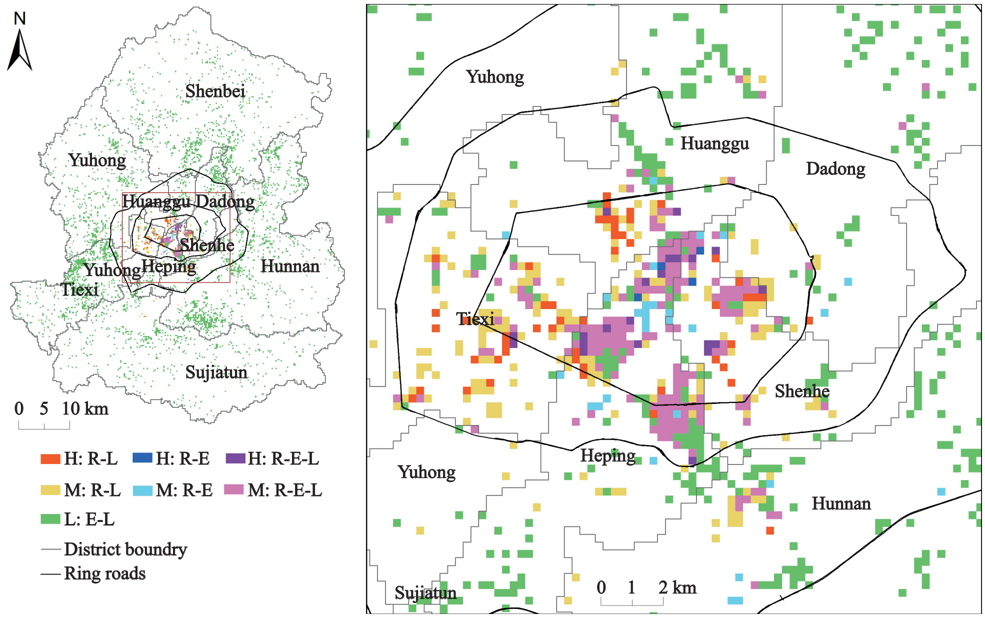

Fig. 9 Spatial distribution of mixed functional areas include residential-leisure mixed functional area with high residential population density (H: R-L), residential-employment mixed functional area with high residential population density (H: R-E), residential-employment-leisure mixed functional area with high residential population density (H: R-E-L), residential-leisure mixed functional area with medium residential population density (M: R-L), residential-employment mixed functional area with medium residential population density (M: R-E), residential-employment-leisure mixed functional area with medium residential population density (M: R-E-L) and employment-leisure mixed functional area with low residential population density (L: E-L)

The R-E were concentrated in the old town of Tiexi District and southern Huanggu District. Since the industrial enterprises were relocated into the new town of Tiexi District when the municipal government's replanning of Tiexi District in 2002, the old industrial Tiexi was transformed into a modern commercial and residential center (Xue et al., 2016). The R-E was the least distributed. The E-L were scattered in the whole city, but less within the Second Ring Road.

3.5 Correlation between urban facilities and active population density

The correlation coefficients between urban facilities and active population density at block scale passed the two-sided hypothesis test. Daily service facilities and catering facilities were highly correlated with the residential population density. Transportation facilities, catering facilities, and daily service facilities were highly correlated with the employment population density. Transportation facilities, daily service facilities and catering facilities were highly correlated with the leisure population density (Table 2). Therefore, four types of facilities, namely, transportation, catering, daily service and shopping, were most closely related to population activities. Three types of facilities, namely, public service, accommodation and financial office, were less closely related to population activities. Medical and attraction facilities were least related to population activities. In addition, there was also a high degree of pairwise correlation among residential population density, employment population density and leisure population density.

4 Discussion

4.1 Multi-source big data application

Previous studies have indicated that multi-source big data has certain advantages in urban land use and refined research on UFAs (Kuang, 2012). Earth observation big data carries rich 'land use' information, while human observation big data reflect human activities, and the combination of earth observation and human observation big data is considered to have great potential in the study of human-land relationship (Li, 2018). Understanding UFAs needs to break through the land use perspective and make a new interpretation from the perspective of human activities (Zhang et al., 2020). This study used 250 m × 250 m grid resolution cell phone data and realized UFA identification from the perspect-

Table 2 Correlation coefficient matrix between active population density / (persons/km²) and urban facility density / (units/km²)

	RP	EP	LP	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
RP	1											
EP	0.932	1										
LP	0.976	0.968	1									
A_1	0.916	0.888	0.94	1								
A_2	0.565	0.527	0.562	0.555	1							
A3	0.732	0.81	0.808	0.824	0.409	1						
A_4	0.906	0.943	0.943	0.941	0.542	0.837	1					
A_5	0.867	0.911	0.929	0.964	0.533	0.894	0.953	1				
A_6	0.907	0.877	0.926	0.957	0.582	0.789	0.906	0.927	1			
A_7	0.416	0.454	0.468	0.517	0.262	0.613	0.474	0.551	0.5	1		
A_8	0.654	0.758	0.776	0.798	0.43	0.802	0.791	0.887	0.758	0.452	1	
A_9	0.681	0.851	0.74	0.706	0.352	0.751	0.795	0.791	0.673	0.407	0.696	1

Notes: RP is residential population density, EP is employment population density, LP is leisure population density; A_1–A_9 are daily service facility density, medical facility density, public service facility density, transportation facility density, restaurant facility density, shopping facility density, attraction facility density, accommodation facility density and financial office facility density, respectively

ive of population activities. Compared with previous studies using a single type of data (Li et al., 2021; Yang et al., 2021), this study integrated multi-source big data consisting of cell phone data, POI data and other auxiliary data to explore the interaction between population activities and land use in UFAs. The feasibility and advantages of multi-source big data in the field of UFA analysis were verified, which enriched the research on human-land relationship.

Although cell phone data can realize high-frequency human activity monitoring, whether it can realize panoramic monitoring and thus improve the reliability of UFA identification is a question worthy of discussion. Early studies indicated that the population identification results based on cell phone data and census data are highly linearly correlated, and the identification results have a high degree of credibility (Goodchild, 2007; Becker et al., 2011). The cell phone heat map used in this study has a strong correlation with the demographic data, but it is limited by the relatively low holding rate of cell phones among children and elderly people. Therefore, we identified the functional areas by the relative differences among different types of active population per unit area instead of the absolute amount of the active people, which reduced errors of identification results. However, it is still necessary to further combine the field research data with the identification results using mobile phone data for mutual verification and accuracy assessment.

4.2 Comparison of functional area identification methods

Unlike previous studies that focused only on the central urban area, this study takes the city as a whole as the study area, which includes the central urban area, suburban areas, and distant suburban areas. This poses a challenge for classifying different UFAs from population density due to the gradient differences in population density in different zones. Based on the study of Niu et al. (2014), we increase the population density gradient and divide different functional areas at each population density level, thereby formulating 12 UFA classification rules. We effectively distinguished the functional structure differences between suburban and distant suburban areas, and generally improved the fineness of functional area identification. In addition, the study of geographic patterns at different spatial and temporal scales is an effective way to understand the mechanisms behind the process (Wang et al., 2010). While previous studies have focused on single-scale UFAs (Li et al., 2021; Yang et al., 2021), this study divided multi-scale spatial units by GIS that breaks through the limitation of the spatial statistical unit that is dominated by a single scale. Our results show that the urban functional characteristics of Shenyang City exhibit certain scale effects.

Although we have divided urban areas into different combinations of residence-employment-recreation function, it is necessary for further research to focus on the detailed classification of these functions. For example, which industries are included in the employment functional area? What are the distribution characteristics of employment in different industries? Which age groups of the population dominate the residential function and in what proportion? Answering these questions is helpful for a more refined urban function division, helping the government to monitor a more comprehensive urban land use status and realize 'people-cantered' smart city construction and management. Therefore, in the future, technical methods such as unsupervised classification and machine learning should be used to perceive more population activity patterns (Zhang et al., 2020). In addition, the user tag information of cell phone data can provide support for the fine division. At present, studies have been conducted to mine more land use features using cell phone user characteristics (e.g., age, gender, *etc*.) and travel characteristics (e.g., commuting distance, *etc*.).

4.3 UFAs spatial distribution and planning management

The traditional theory of urban spatial structure suggests that the UFAs expand from the core to the outer circle through continuous spatial evolution and urban residential areas radiate outward from the city center along the traffic line. Shenyang's R-E-L decrease from the central area to the outer circle and develop in strips along the main traffic corridors, which corroborates the traditional theory of spatial structure. In Shenyang City, the residential population was gathering in the suburbs, but the employment population was not gathering synchronously. Similarly, the employment population in Harbin, another old industrial city in Northeast China, were clustered in the central urban area, but the employment population has a more significant central concentration than the residential population (Yuan et al., 2012). These examples all reflect differences in development between urban and peri-urban and rural areas (Li et al., 2015). Increasing the degree of mixing of urban functions is important to improve urban dynamics and promote sustainable urban development (United Nations, 2015). Finally, this study shows that blocks with higher POI density have higher active population densities, meaning that the formation of urban functions is related to urban infrastructure. Previous studies have indicated that supplementing POI supply may be beneficial to increase urban vitality (Yue et al., 2017). In Shenyang, peri-urban or rural areas may need to improve public services and infrastructure.

The 'Shenyang City's Master Plan (2011–2020)' proposed to build four subsidiary cities including of Hunhe New City, Puhe New City, Yongan New City and Tiexi New City, aiming to accelerate the construction of large residential areas in the subsidiary cities. However, there were mainly low-density residential areas there (Fig. 8a). The uneven distribution of auxiliary facilities in subsidiary city has weakened people's willingness to live, which also led to the separation of employees and residences. We found that monitoring urban functional service areas from the perspective of population distribution based on spatio-temporal big data helps identify discrepancies between urban land planning and actual usage. In an empirical study for Shenzhen, China, and Singapore, a mismatch between the status of UFAs and the official land zoning map was also found (Pei et al., 2014; Tu et al., 2017). Therefore, it can help city managers to scientifically allocate urban infrastructure, enhance urban vitality and improve urban efficiency. On the other hand, demographic activity monitoring through big data can reflect the spatial needs of urban residents in more detail. We suggest that politicians and scientists link up to take measures to optimize the urban spatial structure (Kuang, 2012). It is important to make full use of spatial and temporal big data to monitor the current urban space and control the intensity of land development. This will innovate previous 'ultimate blueprint' mode of land layout and form a good interaction between urban functions and population demand.

5 Conclusions

The identification and spatial analysis of UFAs are important topics in urban geography research and the basis for promoting sustainable urban development planning. Based on cell phone signalling data, POI data, basic geographic data, demographic data and urban master plan document, we constructed a multi-scale UFA identification methodology frame. Taking Shenyang City as a case, we verified the validity of the identification method and carried out UFA identification and spatial distribution analysis. The integrated application of multi-source data provides a new methodology framework for identification of multi-scale multi-type UFAs in the perspective of human activity, complementing the UFA research paradigm. Moreover, with the support of POI big data, this study built an explanation path for the correlation between urban facilities and population activity, enhancing the analysis of UFAs in the perspective of human-land relationship. The distribution of UFAs in Shenyang has the characteristics of a 'concentric circle, fan-shaped, multi-core' pattern. The trend of the resident population spreading to the suburbs is greater than that of the employed population. There are more mix functional areas in the central city areas, the planned industrial new cities need to develop comprehensive functions. In addition, we found that UFAs have scale effects. Residential, employment and leisure functional areas have significant positive correlations with shopping, financial office and transportation facilities respectively. We advocate the use of machine learning methods to sense more population activity patterns and promote the fine delineation of UFAs. We suggest that city managers make full use of multi-source big data to carry out bottom-up urban spatial sensing and establish a human-centered urban spatial control strategy.

References

Ahas R, Silm S, Järv O et al., 2010. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, 17(1): 3–27. doi: 10.1080/10630731003597306

Batty M, 2013. *The New Science of Cities*. Cambridge: MIT Press.

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

Berry B J L, Kasarda J D, 1977. *Contemporary Urban Ecology*. New York: Macmillan.

Burgess E W, 2008. The growth of the city: an introduction to a research project. In: *Urban Ecology*. New York: Springer, 71–78. doi: 10.1007/978-0-387-73412-5_5

Chinese Urban Planning Society, 2016. *Shenyang City Master Plan* (2011–2020). Available at: http://www.planning.org.cn/2016anpc/view?id = 413. (in Chinese)

Crooks A, Pfoser D, Jenkins A et al., 2015. Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, 29(5): 720–741. doi: 10.1080/13658816.2014.977905

Dong L J, Wang Y C, Lin J Y et al., 2020. The community renewal of Shantytown transformation in old industrial cities: evidence from Tiexi worker village in Shenyang, China. *Chinese Geographical Science*, 30(6): 1022–1038. doi: 10.1007/s11769-020-1164-6

Estima J, Painho M, 2016. User generated spatial content-integrator: conceptual model to integrate data from diverse sources of user generated spatial content. *ISPRS International Journal of Geo-Information*, 5(10): 183. doi: 10.3390/ijgi5100183

Farmer C J Q, Fotheringham A S, 2011. Network-based functional regions. *Environment & Planning A:Economy and Space*, 43(11): 2723–2741. doi: 10.1068/a44136

Goodchild M F, 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4): 211–221. doi: 10.1007/s10708-007-9111-y

Goodchild M F, Klinkenberg B, Janelle D G, 2010. A factorial model of aggregate spatio-temporal behavior: application to the diurnal cycle. *Geographical Analysis*, 25(4): 277–294. doi: 10.1111/j.1538-4632.1993.tb00299.x

Harris C D, Ullman E L, 1945. The nature of cities. *The Annals of the American Academy of Political and Social Science*, 242(1): 7–17. doi: 10.1177/000271624524200103

Hoyt H, 1939. *The Structure and Growth of Residential Neighborhoods in American Cities*. Washington, DC: Scholarly Pr.

Hu S, He Z J, Wu L et al., 2020. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data. *Computers, Environment and Urban Systems*, 80: 101442. doi: 10.1016/j.compenvurbsys.2019.101442

Jiang S, Alves A, Rodrigues F et al., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53: 36–46. doi: 10.1016/j.compenvurbsys.2014.12.001

Kashian A, Rajabifard A, Richter K F et al., 2019. Automatic analysis of positional plausibility for points of interest in OpenStreetMap using coexistence patterns. *International Journal of Geographical Information Science*, 33(7): 1420–1443. doi: 10.1080/13658816.2019.1584803

Kuang W H, 2012. Spatio-temporal patterns of intra-urban land use change in Beijing, China between 1984 and 2008. *Chinese Geographical Science*, 22(2): 210–220. doi: 10.1007/s11769-012-0529-x

Kuang W H, 2020. 70 years of urban expansion across China: trajectory, pattern, and national policies. *Science Bulletin*, 65(23): 1970–1974. doi: 10.1016/j.scib.2020.07.005

Kuang W H, Du G M, Lu D S et al., 2021. Global observation of urban expansion and land-cover dynamics using satellite big-data. *Science Bulletin*, 66(4): 297–300. doi: 10.1016/j.scib.2020.10.022

Kuang W H, Yan F Q, 2018. Urban structural evolution over a century in Changchun city, Northeast China. *Journal of Geographical Sciences*, 28(12): 1877–1895. doi: 10.1007/s11442-018-1569-7

Lee J, Newman G, 2017. Forecasting urban vacancy dynamics in a shrinking city: a land transformation model. *ISPRS International Journal of Geo-Information*, 6(4): 124. doi: 10.3390/ijgi6040124

Li D, 2018. Brain cognition and spatial cognition: on integration of geo-spatial big data and artificial intelligence. *Geomatics and Information Science of Wuhan University*, 43(12): 1761–1767. (in Chinese)

Li J Z, Xie X, Zhao B Y et al., 2021. Identification of urban functional area by using multisource geographic data: a case study of Zhengzhou, China. *Complexity*, 2021: 8875276. doi: 10.1155/2021/8875276

Li M Y, Shen Z J, Yang T et al., 2015. Application of spatial and temporal entropy based on multivariate data for measuring the degree of urban function mix. *China City Planning Review*, 24(1): 8–16. (in Chinese)

Liu J G, Dietz T, Carpenter S R et al., 2007. Complexity of coupled human and natural systems. *Science*, 317(5844): 1513–1516. doi: 10.1126/science.1144004

Liu J Y, Liu M L, Zhuang D F et al., 2003. Study on spatial pattern of land-use change in China during 1995–2000. *Science in China Series D:Earth Sciences*, 46(4): 373–384. doi: 10.1360/03yd9033

Liu X, Kang C G, Gong L et al., 2016. Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, 30(2): 334–350. doi: 10.1080/13658816.2015.1086923

Liu X T, Gu C L, 2008. Decoding urban land-use spatial structure: a case study on the city of Nanjing. *Urban Planning Forum*, (5): 78–84. (in Chinese)

Louail T, Lenormand M, Cantu Ros O G et al., 2014. From mobile phone data to the spatial structure of cities. *Scientific Reports*, 4: 5276. doi: 10.1038/srep05276

Mao H N, Ahn Y Y, Bhaduri B et al., 2017. Improving land use inference by factorizing mobile phone call activity matrix. *Journal of Land Use Science*, 12(2–3): 138–153. doi: 10.1080/1747423x.2017.1303546

Mariathasan V, Bezuidenhoudt E, Olympio K R, 2019. Evaluation of earth observation solutions for Namibia's SDG monitoring system. *Remote Sensing*, 11(13): 1612. doi: 10.3390/rs11131612

Markonis Y, Pappas C, Hanel M et al., 2021. A cross-scale framework for integrating multi-source data in Earth system sciences. *Environmental Modelling & Software*, 139: 104997. doi: 10.1016/j.envsoft.2021.104997

Martin D, Cockings S, Harfoot A, 2013. Development of a geographical framework for census workplace data. *Journal of the Royal Statistical Society:Series A (Statistics in Society)*, 176(2): 585–602. doi: 10.1111/j.1467-985X.2012.01054.x

Niu X Y, Ding L, Song X D, 2014. Understanding urban spatial structure of Shanghai central city based on mobile phone data. *Urban Planning Forum*, (6): 61–67. (in Chinese)

Novak J, Ahas R, Aasa A et al., 2013. Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of Estonia. *Journal of Maps*, 9(1): 10–15. doi: 10.1080/17445647.2012.762331

Pei T, Sobolevsky S, Ratti C et al., 2014. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28(9): 1988–2007. doi: 10.1080/13658816.2014.913794

See L, Mooney P, Foody G et al., 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information*, 5(5): 55. doi: 10.3390/ijgi5050055

Shearmur R, 2015. Dazzled by data: big data, the census and urban geography. *Urban Geography*, 36(7): 965–968. doi: 10.1080/02723638.2015.1050922

Shen Y, Karimi K, 2016. Urban function connectivity: characterisation of functional urban streets with social media check-in data. *Cities*, 55: 9–21. doi: 10.1016/j.cities.2016.03.013

Shenyang Statistics Bureau, 2020. *Statistical Yearbook 2019* (data year 2018). Available at: http://tjj.shenyang.gov.cn/sjfb/ndsj/202207/t20220711_3410131.html.

Shi W Z, Goodchild M F, Batty M et al. , 2021. *Urban Informatics*. Singapore: Springer. doi: 10.1007/978-981-15-8983-6

Sweet M N, Bullivant B, Kanaroglou P S, 2017. Are major Canadian city-regions monocentric, polycentric, or dispersed. *Urban Geography*, 38(3): 445–471. doi: 10.1080/02723638.2016.1200279

Tian G J, Wu J G, Yang Z F, 2010. Spatial pattern of urban functions in the Beijing metropolitan region. *Habitat International*, 34(2): 249–255. doi: 10.1016/j.habitatint.2009.09.010

Tu W, Cao J Z, Yue Y et al., 2017. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, 31(12): 2331–2358. doi: 10.1080/13658816.2017.1356464

United Nations, 2015. Transforming our world: the 2030 agenda for sustainable development. Available at: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf.

Van de Voorde T, Jacquet W, Canters F, 2011. Mapping form and function in urban areas: an approach based on urban metrics and continuous impervious surface data. *Landscape & Urban Planning*, 102(3): 143–155. doi: 10.1016/j.landurbplan.2011.03.017

Vorontsova A V, Vorontsova V L, Salimgareev D V, 2016. The development of urban areas and spaces with the mixed functional use. *Procedia Engineering*, 150: 1996–2000. doi: 10.1016/j.proeng.2016.07.277

Wang T Y, Yue W Z, Ye X Y et al., 2020. Re-evaluating polycentric urban structure: a functional linkage perspective. *Cities*, 101: 102672. doi: 10.1016/j.cities.2020.102672

Wang Y F, Fu B J, Lü Y H et al., 2010. Local-scale spatial variability of soil organic carbon and its stock in the hilly area of the Loess Plateau, China. *Quaternary Research*, 73(1): 70–76. doi: 10.1016/j.yqres.2008.11.006

Wei L, Luo Y, Wang M et al., 2020. Multiscale identification of urban functional polycentricity for planning implications: an integrated approach using geo-big transport data and complex network modeling. *Habitat International*, 97: 102134. doi: 10.1016/j.habitatint.2020.102134

Xu Y Y, Liu M, Hu Y M et al., 2019. Analysis of three-dimensional space expansion characteristics in old industrial area renewal using GIS and Barista: a case study of Tiexi District, Shenyang, China. *Sustainability*, 11(7): 1860. doi: 10.3390/su11071860

Xue B, Zhang L M, Geng Y et al., 2016. Extended land-use coding system and its application in urban brownfield redevelopment: case study of Tiexi District in Shenyang, China. *Journal of Urban Planning & Development*, 142(3): 05015014,1–11. doi: 10.1061/(ASCE)UP.1943-5444.0000309

Xue B, Xiao X, Li J Z, 2020a. Identification method and empirical study of urban industrial spatial relationship based on POI big data: a case of Shenyang City, China. *Geography and Sustainability*, 1(2): 152–162. doi: 10.1016/j.geosus.2020.06.003

Xue B, Xiao X, Li J Z et al., 2020b. Analysis of spatial economic structure of Northeast China cities based on Points of Interest big data. *Scientia Geographica Sinica*, 40(5): 691–700. (in Chinese)

Yang Z S, Su J H, Yang H et al., 2021. Exploring urban functional areas based on multi-source data: a case study of Beijing. *Geographical Research*, 40(2): 477–494. (in Chinese)

Yuan N J, Zheng Y, Xie X et al., 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 27(3): 712–725. doi: 10.1109/TKDE.2014.2345405

Yuan Y H, Raubal M, Liu Y, 2012. Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Computers, Environment and Urban Systems*, 36(2): 118–130. doi: 10.1016/j.compenvurbsys.2011.07.003

Yue Y, Zhuang Y, Yeh A G O et al., 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, 31(4): 658–675. doi: 10.1080/13658816.2016.1220561

Zhang S Q, Yang Y, Zhen F et al., 2020. Exploring temporal activity patterns of urban areas using aggregated network-driven mobile phone data: a case study of Wuhu, China. *Chinese Geographical Science*, 30(4): 695–709. doi: 10.1007/s11769-020-1130-3

Zhang Z X, Wang Y H, Liu Q J et al. , 2016. A CNN based functional zone classification method for aerial images. In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Beijing, China: IEEE, 5449–5452. doi: 10.1109/IGARSS.2016.7730419

Zhou G L, Li C G, Li M Y et al., 2016. Agglomeration and diffusion of urban functions: an approach based on urban land use conversion. *Habitat International*, 56: 20–30. doi: 10.1016/j.habitatint.2016.04.002