URBAN FUNCTIONAL DISTRICT IDENTIFICATION AND ANALYSIS FROM MULTI-SOURCE DATA

Ye Zhang [1,2], Kun Qin [3,*], Wanzeng Liu [1,2], Xiuli Zhu [1,2], Yunlu Peng [1,2], Xinpeng Wang [1,2], Xi Zhai [1,2], Tingting Zhao [1,2], Ran Li [1,2]

[1] National Geomatics Center of China, 28 Lianhuachi West Road, Haidian District, Beijing 100830, China - (zhangye, zhuxiuli, pengyunlu, wangxinpeng, zhaixi, zhaotingting, liran)@ngcc.cn; luwnzg@163.com.
[2] Key Laboratory of Spatio-temporal Information and Intelligent Services (LSIIS), MNR, 28 Lianhuachi West Road, Haidian District, Beijing 100830, China
[3] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Hongshan District, Wuhan 430079, China - qink@whu.edu.cn

Commission IV, WG IV/10

KEY WORDS: Urban Functional District, Social Sensing Data, Taxi Trajectory, OpenStreetMap, Head/tail Breaks Method, Data Mining.

ABSTRACT:

Residents' activities have a significant interaction with urban socioeconomic environment. Taxi trajectory data has been widely used to mine human activity patterns to identify urban functional districts. However, previous studies merely chose several spatiotemporal statistics of taxi pick-up and drop-off points. This paper compares seven time series statistics of taxi pick-up and drop-off points, and selects the best combination to identify urban functional districts. The basic analysis units are not only constructed based on the OpenStreetMap data, but also optimized with the fine-grained clean rasterized pixels, generated from preprocessed taxi trajectory data through the improved head/tail breaks method. The experiment conducted in Wuchang District, Wuhan, shows that the combination of the average statistics of pick-up points, the average statistics of drop-off points, and the ratio statistics of pick-up and drop-off difference achieves the best identification precision of 83.65%, the F1-score of 82.2%, and the recall score of 81.48%. The proposed approach has good scalability and can be transplant to other identification applications.

1. INTRODUCTION

Urban functional districts refer to residential land, commercial land, industrial land, public management and public service facilities as well as other functional zones are gradually formed in the process of urban development (Zhang et al., 2017). The unified, coordinated and reasonable layout of urban functional districts is conducive to improving urban land use efficiency, optimizing resource allocation, realizing balanced urban development, and improving the overall strength of the city. Identifying different types of functional districts and studying their spatial distribution patterns and interaction laws are of great significance for managing the urban spatial structure and establishing formulating scientific and reasonable urban planning policies (Yao et al., 2022).

The emergence of remote sensing data provides a large amount of accurate land cover information, making it possible to effectively characterize the land use status of each urban functional district. It has been widely used in high-precision mapping (Li et al., 2016), typical natural element extraction (Huang et al., 2017), emergency monitoring of natural disasters (Lu et al., 2018), and so on. However, early urban remote sensing research mainly focused on mining natural physical information in urban areas. Therefore, the development of urban remote sensing is constrained by two aspects (Zhu et al., 2019): on the one hand, it is very difficult to directly obtain other socioeconomic information in urban areas; on the other hand, urban land use classification based on remote sensing data always needs the support of prior knowledge. Inspired by 'citizens are sensors', we could transform earth observation into human observation, bringing in more socioeconomic data to address the above issues.

In recent years, social sensing data has been widely used in human activity patterns extraction and urban studies, such as mobile phone data (Jia et al., 2018; Pei et al., 2014; Ratti et al., 2006; Reades et al., 2009; Toole et al., 2012), points of interest (POIs) (Hu et al., 2016; Jiang et al., 2015; Yao et al., 2017; Yuan et al., 2012), check-in data (Gao et al., 2017; Zhang et al., 2017), smart card data (Zhong et al., 2014, 2016), and taxi trajectories (Chen et al., 2017; Liu et al., 2012, 2016, 2020; Pan et al., 2013; Zheng et al., 2011). A few studies focused solely on social sensing data to infer urban land use. For instance, (Pei et al., 2014) used the normalized hourly call volume and the total call volume of mobile phone data, inside each Voronoi polygon, to characterize human communication and applied the semi-supervised fuzzy c-means clustering approach to split Singapore into residential, business, commercial, open space and others. Jiang et al. (2015) used US census data, GIS data and multi-source POIs with the North American Industry Classification System (NAICS) codes to disaggregate land use at the census block level. Their goal is to provide a new way to disaggregate employment data by sizes and industrial category into higher spatial resolution units. Wang et al. (2016) used a regular grid of 400 m x 400 m to divide the study area into cells. Within each cell, they applied K-means clustering algorithm to analyze discussion topics on geotagged social media (Sina-Weibo) data to identify seven types of land use clusters. Then among the seven clusters, they used text mining and word clouds to

* Corresponding author

estimate land use types: residential areas, commercial areas, work areas, transportation hub areas and mixed land use areas. Zhong et al. (2014) used smart card data, the Household Interview Travel Survey, bus stop location points and geo-referenced building footprints to infer urban functions (shopping, studying, working, at home, eating and social visiting) at the building level. Pan et al. (2013) focused exclusively on using taxi trajectories to classify land use at two stages. First, they proposed an improved clustering algorithm (iterative DBSCAN) to extract regions with high taxi pick-up density and manually labeled social function (e.g., stations, campuses, hospitals, scenic spots, commercial districts, entertainment districts, office buildings and residential districts) as training samples. Then they comparatively analyzed the land-use classification performance of four classifiers with designing six taxi pick-up and drop-off features for land-use classification, e.g., daily pick-up feature, daily set-down feature, pick-up and drop-off difference feature, pick-up and drop-off ratio feature, weekly pick-up feature and weekly drop-off feature. The experiment results showed that the first feature has the best discriminative capability and the daily pick-up and drop-off information is very helpful.

Compared with traditional data sources such as questionnaire surveys, the outstanding advantages of taxi trajectory data are large coverage, large sample size and high precision (Kandt and Batty, 2021). Among these social sensing data, taxi trajectory data is more accurate than mobile phone data, demographic coverage is more diverse than POIs and check-in data, and geographical routes are more flexible than smart card data. When it comes to fixed-route public transportation such as buses and subways, although taxi trajectories represent only a small fraction of the total a city's total public transport, it can provide 24 hours a day, 7 days a week, wide coverage, and detailed data on residents' travel routes. Residents' travel behavior can extract urban functional areas by providing information on the following two aspects: on the one hand, when residents arrive or leave a certain area; and on the other hand, where residents arrive or leave (Yuan et al., 2015). Consequently, inferring residents' travel purposes from taxi trajectories can more accurately reflect the current land use in urban areas (Liu et al., 2016).

As demonstrated above, existing researches inferred urban land use at different levels,ranging from single building level (Du et al., 2015; Huang et al., 2017; Zhong et al., 2014) to urban parcel/block level (Voltersen et al., 2014; Zhang et al., 2017). This study constructed the analysis unit based on the OpenStreetMap road network to unify the spatial scale of the aforementioned multi-source data. Urban blocks are defined as the space delineated by road networks. Many researchers generated urban blocks with the help of the local authorities (Voltersen et al., 2014), object-based image analysis maps, or open source data such as OpenStreetMap. However, it is difficult to obtain the official data, and the large number of trees on the roadside hinder the extraction of the complete urban road network from remote sensing images. While previous researches have shown that the OpenStreetMap road network meets the need of designing urban blocks (Zhang et al., 2017) and its availability enhances the portability of the proposed method (Du et al., 2015).

In addition to the OpenStreetMap road network, this study used taxi trajectory data to help refine the road network. Existing studies mostly extract road centerlines from taxi trajectory data in four ways: (1) clustering, (2) the incremental method (Wu et al., 2019), (3) rasterization (Fang et al., 2020), (4) other methods (Yang and Ai, 2017; Zhang et al., 2020). This paper uses mathematical morphological methods to extract road centerlines from the rasterized image of taxi trajectories. After testing and analysis many times, taxi trajectory points were mapped onto an image with a spatial resolution of 2.5 meters. Additionally, an improved head/tail breaks method was used to filter noise pixels.

This study focuses on identifying urban functional zones from multi-source data. It contributes in three ways. First, taxi trajectory data has been used to extract residents' travel behavior for urban functional district identification. Second, the head/tail breaks method has been introduced and improved to reduce the noise pixels when using taxi trajectory data to construct analysis zones. Third, optimum combination of time series statistics of taxi pick-up and drop-off points has been analyzed to identify urban functional districts.

The reminder of this paper is arranged as follows: Section 2 introduces multi-source data and preprocessing and methods in detail. Section 3 analyzes the experimental results of analysis unit construction and urban functional district identification, and discusses the best combination of time series statistics of taxi pick-up and drop-off points for identification. Finally, our conclusions are stated in Section 4.

2. MATERIALS AND METHODS

2.1 Study Area

Wuchang District, is one of the main districts of Wuhan city (Figure 1 (b)), Hubei Province, China. As far as the geographical environment is concerned, Wuchang is bordered by the Yangtze River in the west and East Lake in the east, with complex terrain and intertwined hills and lakes. From the perspective of the socioeconomic environment, Wuchang District has 14 administrative streets, with 141 city communities in 2015. Meanwhile, the permanent population is 1.27 million, and the registered population is 1.06 million. Its gross domestic product achieves 881.56 billion yuan, and the total industrial output value is 201.42 billion yuan (Bureau of Statistics of Wuhan, 2016). In addition, there are government agencies such as Hubei Provincial People's Government and Wuchang District People's Government, places of interests such as Memorial Hall of the Revolution of 1911 and Yellow Crane Tower Scenic Area, residential areas, and other urban functional districts. The complexity of functional types in Wuchang District makes it a good representative study area for identifying functional districts. This paper mainly uses taxi pick-up and drop-off points to extract the travel patterns of residents for describing the socioeconomic environment of Wuchang District. Figure 1 (a) depicts the spatial distribution of taxi pick-up points in Wuchang District, and the distribution of drop-off points is very similar.

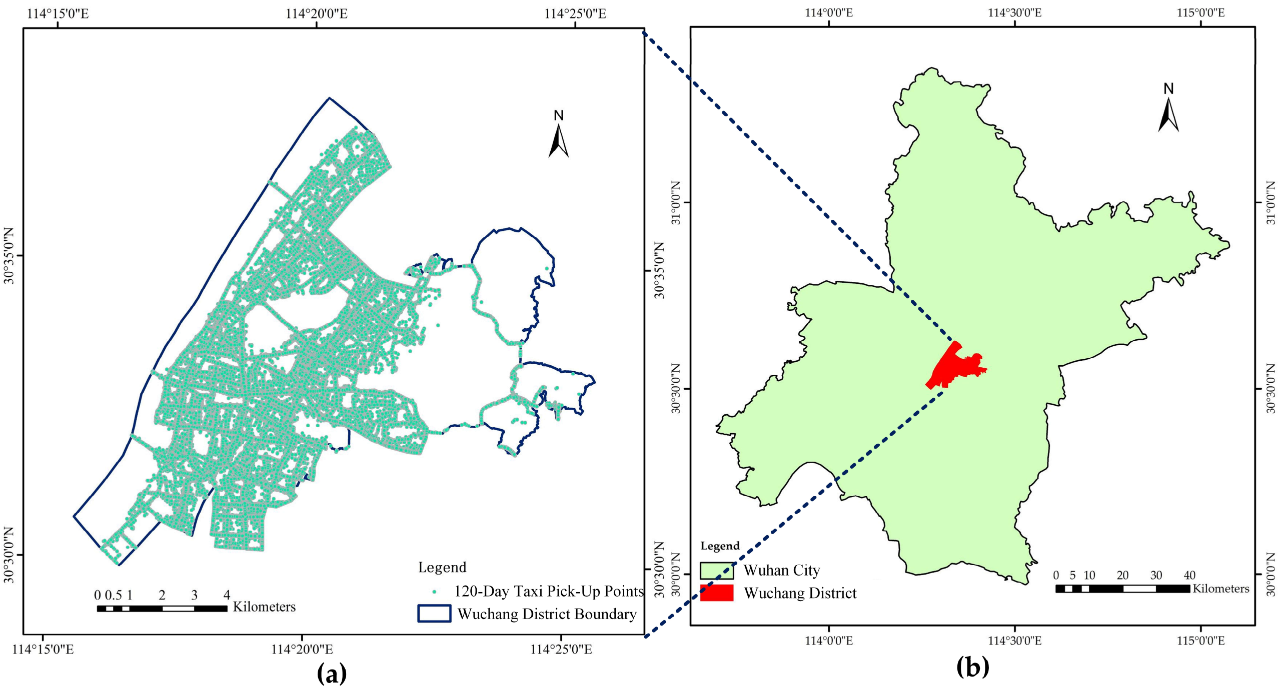

Figure 1. Study area.

2.2 Multi-Source Data and Preprocessing

2.2.1 Taxi Trajectories and Preprocessing

The whole taxi trajectory dataset was collected from February 1st to August 10th, 2015. This paper uses taxi trajectories of Wuchang District to provide sufficient information for road network update, residents' travel pattern analysis, and urban functional district identification. Its sampling interval is less than one minute. The trip records were collected by the GNSS device installed in the taxis. And each record consists of the following information: the ID of the taxi, the timestamp, longitude, latitude, the instantaneous driving direction, the driving speed, the Advanced Cruise Control (ACC), the taxi operation status and the vacant statue. Among them, the vacant status reflects the situation of passengers getting on and off.

The original taxi trajectory data has been preprocessed according to common operations in previous studies (Chen et al., 2017; Zheng et al., 2018). We removed the following records, such as: (1) the location outsides the Wuchang District, (2) the timestamp is beyond the range of February 1st to August 10th, 2015, (3) the attributes are incomplete or invalid in this study. Besides, we modified some records with the correct attributes, but not in the correct order.

Figure 2 shows a histogram of the number of daily taxi pick-up and drop-off point during the study period, where the x tick labels represent each date from February 1st to August 10th. It can be clearly seen that the total numbers fluctuate greatly and sometimes the numbers are wrong. During the research period, the national holidays, represented by oranges cylinders, were first removed, and the research mainly focuses on the daily taxi travel patterns. We sorted the number of daily taxi pick-up and drop-off points, and selected the top 120 days by number. Because Pan et al. (2013) have tested different time lengths of taxi trajectory data for feature extraction and found that the best land-use classification result is achieved using four-month data.

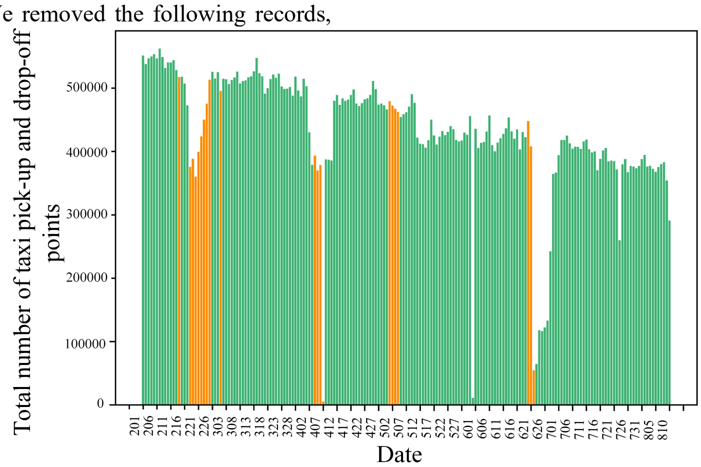

Figure 2. Total number of taxi pick-up and drop-off data per day (Zhang, 2020).

In this study, the key process for matching taxi pick-up and drop-off points into the analysis unit is to determine which side of the road (i.e. the edge of the analysis unit) each taxi pick-up and drop-off point is assigned to. Compared with the traditional nearest-neighbor distance matching method, it is more accurate to project each pick-up and drop-off point into the corresponding analysis unit considering the distance and angle of the taxi pick-up and drop-off point to the road, especially at road intersections (Xu, 2020). Inspired by Xu's work, the analysis unit matching was implemented in four steps in the software ArcMap 10.5.

Firstly, the 'Split Line At Vertices' tool split the final road network (introduced in Section 3) into road segments.

Secondly, the 'Add Geometry Attributes' tool added the 'LINE_BEARING' field to each road segment. The value ranges from 0° to 360°, where 0° means north, increasing in a clockwise direction, which is the same as the instantaneous driving direction of the taxi trajectory data.

Thirdly, the 'Near' tool assigned each taxi pick-up and drop-off point to the nearest road segment by setting the 'Method' parameter to 'Planar' and checking the 'Angle' parameter. The 'NEAR_ANGLE' field has been added to each taxi pick-up and drop-off point, ranging from -180° to 180°, 0° means east, 90° means north, 180° or -180° means west, and -90° means south.

Fourthly, we constructed logical rules to accomplish the analysis unit matching using the relationship between the instantaneous driving direction and the 'NEAR_ANGLE' field: (1) when the direction value is less than or equal to 90°, or greater than or equal to 270°, and the 'NEAR_ANGLE' value is greater than 0°, the taxi pick-up and drop-off point will be assigned to the left side of the road segment, otherwise, the taxi pick-up and drop-off point will be assigned to the right side of the road segment. (2) When the direction value is greater than 90° and less than 270°, and the 'NEAR_ANGLE' value is greater than 0°, the taxi pick-up and drop-off point will be assigned to the right side of the road segment, otherwise, the taxi pick-up and drop-off point will be assigned to the left side of the road segment.

2.2.2 OpenStreetMap Road Network and Preprocessing
OpenStreetMap was used to construct the primary road network of Wuchang District. It is a remarkable project in the field of Volunteered Geographic Information since 2004. It collects data from volunteers with the help of manual survey, Global Positioning System instruments and other free services, and its availability in urban land use classification has been demonstrated in (Grippa et al., 2018). Two preprocessing steps are required before generating the primary road network. The first step is to extract road centerlines, and the second step is to remove the dangling roads (Zhang et al., 2017).

2.3 Methods

The workflow of the proposed method is show in Figure 3. It composes of three parts: analysis unit construction, time series statistic extraction, and urban functional district identification. During the analysis unit construction, the primary road network was extracted from OpenStreetMap. Then, taxi trajectories were used to update the road network, but not all preprocessed trajectory points. We used the improved head/tail breaks method to filter noise pixels from the rasterized image of taxi trajectories and applied morphological operators to extract skeleton pixels as road bases. With reference to Baidu Map and AutoNavi Map, the actual land use of local area was manually tagged. After classifying functional districts according to the national standard and the actual situation of Wuchang District, the analysis units were constructed. Another major part is calculating time series statistics for taxi pick-up and drop-off points and testing different combinations of statistics to identify urban functional districts.

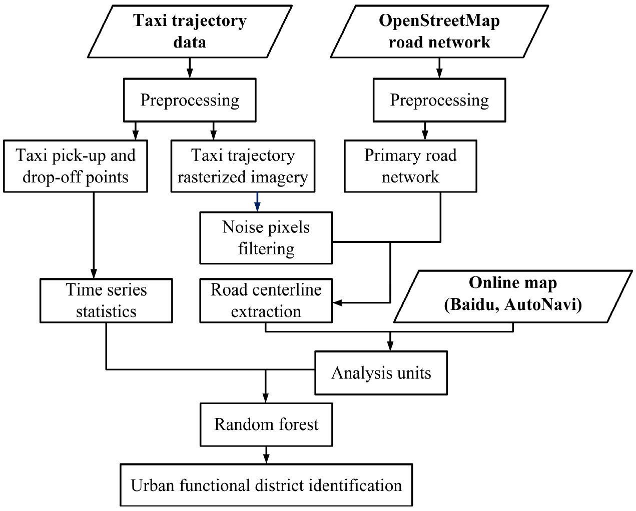

Figure 3. The workflow of the proposed method.

2.3.1 Improved Head/Tail Breaks Method
Inspired by (Ma et al., 2020), this article extended the head/tail breaks method described in (Jiang, 2013) to reduce the noise pixels in the rasterized image of taxi trajectory data and preserve the pixels that contain a certain number of taxi trajectory points. Jiang (2013) used the original head/tail breaks method to equally divide all data values into two parts, and iteratively divided the values (above the mean) until the head part values were no longer heavy-tailed distributed. Generally, researchers rasterize all taxi trajectory data to update the road network, resulting in a large number of abnormal points in the preprocessed taxi trajectory data being mistaken for road pixels. This paper improved the head/tail breaks method by using the weighted average number of taxi trajectory points within a single pixel as the threshold instead of the arithmetic average.

2.3.2 Morphological Operators
After removing a large number of abnormal points, the obtained rasterized image from low-frequency and low-precision taxi trajectory points still has the following problems: 1) The image contains a large number of isolated point groups and holes; 2) The edges of the elements in the image are jagged and uneven. Aiming at these two types of problems, this paper adopted mathematical morphological processing to eliminate outliers, fill holes, and smooth images.

Mathematical morphology, which emerged in 1960s, focuses on the geometric structure of images. The main idea is to scan an image with a structuring element and determine if that element can be effectively filled into the image (Cui, 2000).

The basic morphological operators are dilation and erosion, and the Minkowski form was used here to represent these elements (Haralick and Shapiro, 1993). Dilation used vector addition to filter outside the image and merge the two groups. Erosion used vector subtraction to filter inside the image, merging the two groups. Under the premise of maintaining the main shape features of the image, this paper iteratively used 27 x 27 square structuring elements or 17 x 17 square structuring elements to simplify the rasterized image of taxi trajectories and extracted skeleton pixels.

2.3.3 Time Series Statistics

Time series statistics extraction of taxi pick-up and drop-off points for urban functional area extraction has been extensively studied (Ge et al., 2019; Guo et al., 2012; Liu et al., 2012; Pan et al., 2013). Compared with general time series data and spatial data, taxi pick-up and drop-off points have the characteristics of massive, dynamic, high-dimensional, multi-scale, nonlinear, spatiotemporal correlation and spatiotemporal heterogeneity (Wang et al., 2012). The dynamic nature of taxi pick-up and drop-off points is reflected in their periodicity over time. The spatiotemporal correlation and heterogeneity of data distribution are interrelated in time and space, and at the same time, limited by the urban spatial structure and have spatial heterogeneity.

Referring to the above researches (Ge et al., 2019; Pan et al., 2013), this section calculated seven time series statistics based on 120-day taxi pick-up and drop-off points: the average statistics of pick-up points, the average statistics of drop-off points, the L_2 norm statistics of pick-up points, the L_2 norm statistics of drop-off points, the average statistics of pick-up and drop-off difference, the L_2 norm statistics of pick-up and drop-off difference, and the ratio statistics of pick-up and drop-off difference. We will not summarize the expression of each time series statistic here.

2.3.4 Random Forest

Random forest has been widely applied to combine multiple features. Because the model is insensitive to various scales of features and can measure the importance of features (Zhang et al., 2017). This paper used the random forest model from the Scikit-learn Machine Learning Library.

2.3.5 Analysis Units Construction

Existing studies have generated different types of spatial units for mapping urban land use. For convenience, some researchers split the study area into uniform rectangular grids. This leaves two problems compared to blocks. On the one hand, a grid unit is not as functionally meaningful as the block boundary in analyzing urban land use. For example, it may result in one building being divided into different units or one unit containing buildings with different functions. On the other hand, it retains a certain level of computational complexity. Aiming at these problems, this paper adopted OpenStreetMap road network and taxi trajectories to improve the analysis unit, and was supported by other online map servers. Details are as follows.

(1) The preliminary road network in the study area was constructed from OpenStreetMap data. We used ArcMap software to extract the road centraline from the original OpenStreetMap road network layer.

(2) With the help of taxi trajectory data, we got an accurate road network to improve the spatial division.

(3) In accordance to online map servers, such as Baidu Map and AutoNavi Map, we manually annotated the urban functional classes of all 305 analysis units in ArcMap.

3. RESULTS AND DISCUSSIONS

3.1 Refined Rasterized Image of Taxi Trajectories

According to the number of taxi trajectory points within each pixel, the pixels from the rasterized image of taxi trajectories can be classified into six groups: [1, 11], [12, 22], [23, 58], [59, 100], [101, 256], and [257, 6647] (See Figure 4). The purpose of using the improved head/tail breaks method is to filter noise pixels with sparse points to obtain an accurate and concise road network. We have reserved pixels with more than 12 taxi pick-up and drop-off points. Figure 5 shows the binary classification rasterized image of taxi trajectories. The lines in Figure 5 are thinner and clearer than the ones in Figure 4.

Figure 4. Six-class classification rasterized image of taxi trajectories (Zhang, 2020).

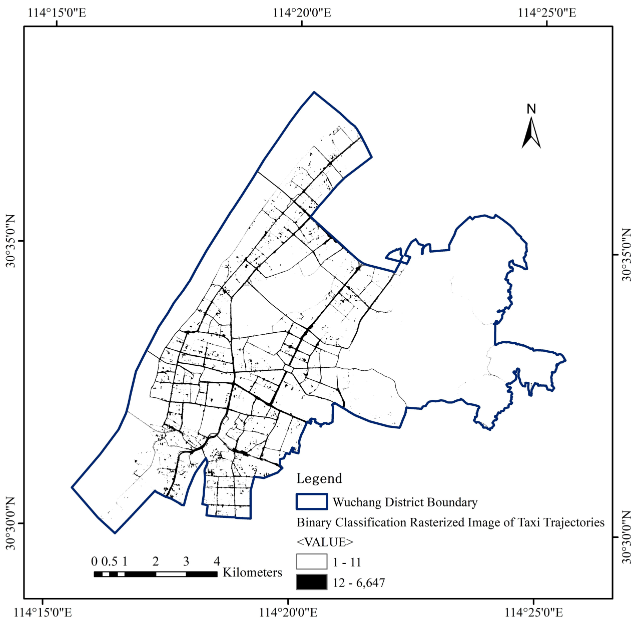

Figure 5. Binary classification rasterized image of taxi trajectories (Zhang, 2020).

3.2 Urban Functional District Identification

This section used the random forest model, combined with time series statistics of taxi trajectories within each analysis unit, to identify urban functional districts and evaluated the performance of different statistics combination. Table 1 shows the experimental accuracy results of different combinations of time series statistics of taxi trajectories for identifying urban functional districts. According to the national standard for basic terminology of urban planning (GB/T 50280-98) and the actual situation, Wuchang functional districts falls into five categories: commercial district, mixed-use district, industrial district, residential district, and central business district. Mixed-use districts can provide administration and public services as well as some other functional services. The functional categories of an urban region are different from the land-use types in that region, and a region may contain multiple land use types but only one functional category. The dominated land use type would be defined as the functional category of the urban region. This relationship links the functional category to the land-use type and links this paper to our previous work (Zhang et al., 2020). Each land use type corresponds to a functional category.

In addition, Zhang et al. (2017) demonstrated that the classification accuracy of pure parcels with a single functional attribute is much higher than that of regions containing mixed functional attributes. Based on our previous work, we retained 269 relatively pure analysis units, except for waters and open spaces, for further functional district identification (see Figure 6). In addition, they can also serve as reference data for subsequent experimental validation. It can be seen from Figure 6 that residential districts cover the main areas of Wuchang District and are distributed in clusters. A few central business districts and commercial districts are clustered, respectively, while the distribution of industrial districts is more scattered.

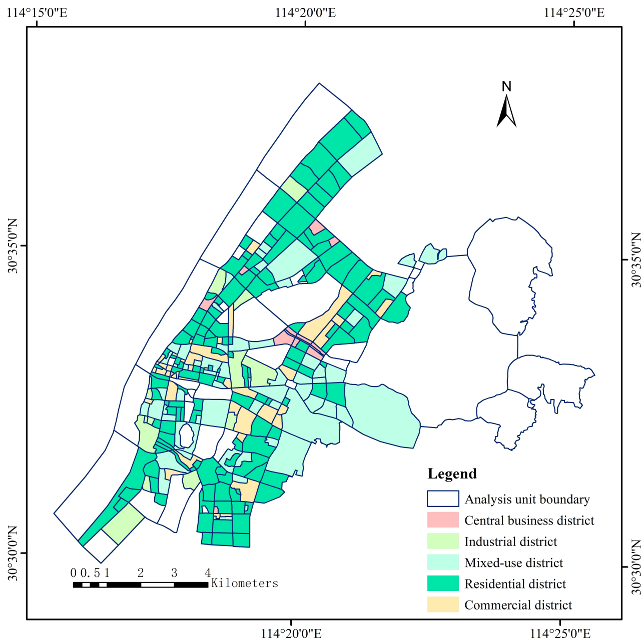

Figure 6. Spatial distribution of 269 functional districts in Wuchang (Zhang et al., 2020).

	Time series statistics of taxi trajectories							Accuracy		
	The L_2 norm statistics		The average statistics		The L_2 norm statistics of pick-up and drop-off difference	The average statistics of pick-up and drop-off difference	The ratio statistics of pick-up and drop-off difference	Precision	F1-score	Recall
	Pick-up points	Drop-off Points	Pick-up points	Drop-off points						
A	√							76.85%	74.69%	77.78%
B		√						62.56%	61.88%	62.96%
C			√					79.07%	76.06%	77.78%
D				√				70.82%	69.23%	70.37%
E	√	√						84.71%	75.39%	77.78%
F			√	√				85.19%	79.63%	77.78%
G					√			64.52%	58.40%	59.26%
H						√		57.78%	59.93%	62.96%
I							√	79.92%	68.69%	70.37%
J			√	√			√	83.65%	82.20%	81.48%

Table 1. Identification accuracy results of urban functional districts with different combinations of time series statistics (Zhang, 2020).

As we can see from Table 1: (1) when only one time series statistic was used to identify urban functional districts, it can be found that using time series statistics of pick-up and drop-off points is more effective than using statistics of pick-up and drop-off difference. But the ratio statistics of pick-up and drop-off difference is a special case. Because the operational form of the ratio enhances the small differences between spatiotemporal taxi trajectories, making the experimental accuracy of the ratio statistics of pick-up and drop-off difference is much higher than the average statistics of pick-up and drop-off difference and the L_2 norm statistics of pick-up and drop-off difference. (2) Compared with the statistics of drop-off points, the accuracy of statistics of pick-up points is higher. The experimental accuracy of the average statistics of pick-up and drop-off points is better that the L_2 norm statistics, which is different from the results of (Pan et al., 2013). In conclusion, the best combination of time series statistics for urban functional district identification is the average statistics of pick-up points, the average statistics of drop-off points, and the ratio statistics of pick-up and drop-off difference. The identification precision reaches 83.65%, the F1-score reaches 82.2%, and the recall score reaches 81.48%.

4. CONCLUSIONS

This paper couples OpenStreetMap, taxi trajectories, and Online maps to identify urban functional districts. OpenStreetMap provides the primary road network. Based on the primary road network, we rasterized taxi trajectories to update the road network to construct analysis units. The improved head/tail breaks method and morphological operators were used to refine the rasterized image of taxi trajectories. Then, seven time series statistics of taxi pick-up and drop-off points were calculated to identify urban functional districts. This paper compared the performance of different combinations of time series statistics to identify urban functional districts. The best combination of time series statistics is the average statistics of pick-up points, the average statistics of drop-off points, and the ratio statistics of pick-up and drop-off difference.

Due to the limitation of taxi trajectories, in the future we plan to study introduce more types of social sensing data that are easier to obtain, such as POIs, subway data, Sina-Weibo data, etc., to extract socioeconomic information in urban areas, analyze the benefits and drawbacks of each data, and pay more attention to feature contribution and feature optimization.

ACKNOWLEDGEMENTS

The comments from three anonymous reviewers are greatly appreciated. The work presented in this paper is supported by the National Key Research and Development Program of China (Grant No. 2018YFC0807005). We are also grateful to Qixin Wang for his help in generating the rasterized image of the original taxi trajectory data, and to Xiaojian Chen for his help in extracting taxi pick-up/drop-off points.

REFERENCES

Bureau of Statistics of Wuhan, 2016. Wuhan Statistical Yearbook - 2016, China Statistics Press. China Statistics Press, Beijing.

Chen, Z., Gong, X., Xie, Z., 2017. An analysis of movement patterns between zones using taxi GPS data. *Trans. GIS*. 21, 1341–1363. doi.org/10.1111/tgis.12281

Cui, Y., 2000. Image Processing and Analysis - Methods and Applications of Mathematical Morphology. Science Press, Beijing.

Du, S., Zhang, F., Zhang, X., 2015. Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS J. Photogramm. Remote Sens*. 105, 107–119. doi.org/10.1016/j.isprsjprs.2015.03.011

Fang, Z., Zhong, H., Zou, X., 2020. Extracting urban road area based on combination of trajectory continuity and image feature similarity. *Cehui Xuebao/Acta Geod. Cartogr. Sin*. 49, 1554–1563. doi.org/10.11947/j.AGCS.2020.20190366

Gao, S., Janowicz, K., Montello, D.R., Hu, Y., Yang, J.A., McKenzie, G., Ju, Y., Gong, L., Adams, B., Yan, B., 2017. A data-synthesis-driven method for detecting and extracting vague cognitive regions. *Int. J. Geogr. Inf. Sci*. 31, 1245–1271. doi.org/10.1080/13658816.2016.1273357

Ge, P., He, J., Zhang, S., Zhang, L., She, J., 2019. An Integrated Framework Combining Multiple Human Activity Features for Land Use Classification. *ISPRS Int. J. Geo-Information*. 8, 90. doi.org/10.3390/ijgi8020090

Grippa, T., Georganos, S., Zarougui, S., Bognounou, P., Diboulo, E., Forget, Y., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., 2018. Mapping urban land use at street block level using OpenStreetMap, remote sensing data, and spatial metrics. *ISPRS Int. J. Geo-Information*. 7, 246. doi.org/10.3390/ijgi7070246

Guo, D., Zhu, X., Jin, H., Gao, P., Andris, C., 2012. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Trans. GIS*. 16, 411–429. doi.org/10.1111/j.1467-9671.2012.01344.x

Haralick, R.M., Shapiro, L.G., 1993. Computer and Robot Vision, Volume II, Computer and Robot Vision. Addison-Wesley, New York.

Hu, T., Yang, J., Li, X., Gong, P., 2016. Mapping urban land use by using landsat images and open social data. *Remote Sens*. 8. doi.org/10.3390/rs8020151

Huang, X., Yuan, W., Li, J., Zhang, L., 2017. A New Building Extraction Postprocessing Framework for High-Spatial-Resolution Remote-Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. 10, 654–668. doi.org/10.1109/JSTARS.2016.2587324

Huang, Y., Zhuo, L., Tao, H., Shi, Q., Liu, K., 2017. A novel building type classification scheme based on integrated LiDAR and high-resolution images. *Remote Sens*. 9. doi.org/10.3390/rs9070679

Jia, Y., Ge, Y., Ling, F., Guo, X., Wang, J., Wang, L., Chen, Y., Li, X., 2018. Urban land use mapping by combining remote sensing imagery and mobile phone positioning data. *Remote Sens*. 10. doi.org/10.3390/rs10030446

Jiang, B., 2013. Head/tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *Prof. Geogr*. 65, 482–494. doi.org/10.1080/00330124.2012.700499

Jiang, S., Alves, A., Rodrigues, F., Ferreira, J., Pereira, F.C., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Comput. Environ. Urban Syst*. 53, 36–46. doi.org/10.1016/j.compenvurbsys.2014.12.001

Kandt, J., Batty, M., 2021. Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*. 109, 102992. doi.org/10.1016/j.cities.2020.102992

Li, M., Stein, A., Bijker, W., Zhan, Q., 2016. Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS J. Photogramm. Remote Sens*. 122, 192–205. doi.org/10.1016/j.isprsjprs.2016.10.007

Liu, H., Xu, Y., Tang, J., Deng, M., Huang, J., Yang, W., Wu, F., 2020. Recognizing urban functional zones by a hierarchical fusion method considering landscape features and human activities. *Trans. GIS*. 24, 1359–1381. doi.org/10.1111/tgis.12642

Liu, X., Kang, C., Gong, L., Liu, Y., 2016. Incorporating spatial interaction patterns in classifying and understanding urban land use. *Int. J. Geogr. Inf. Sci*. 30, 334–350. doi.org/10.1080/13658816.2015.1086923

Liu, Y., Wang, F., Xiao, Y., Gao, S., 2012. Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan*. 106, 73–87. doi.org/10.1016/j.landurbplan.2012.02.012

Lu, C., Yang, X., Wang, Z., Li, Z., 2018. Using multi-level fusion of local features for land-use scene classification with high spatial resolution images in urban coastal zones. *Int. J. Appl. Earth Obs. Geoinf*. 70, 1–12. doi.org/10.1016/j.jag.2018.03.010

Ma, D., Osaragi, T., Oki, T., Jiang, B., 2020. Exploring the heterogeneity of human urban movements using geo-tagged tweets. *Int. J. Geogr. Inf. Sci.* 34, 2475–2496. doi.org/10.1080/13658816.2020.1718153

Pan, G., Qi, G., Wu, Z., Zhang, D., Li, S., 2013. Land-use classification using taxi GPS traces. *IEEE Trans. Intell. Transp. Syst*. 14, 113–123. doi.org/10.1109/TITS.2012.2209201

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.L., Li, T., Zhou, C., 2014. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci*. 28, 1988–2007. doi.org/10.1080/13658816.2014.913794

Ratti, C., Frenchman, D., Pulselli, R.M., Williams, S., 2006. Mobile landscapes: Using location data from cell phones for urban analysis. *Environ. Plan. B Plan. Des*. 33, 727-748 doi.org/10.1068/b32047

Reades, J., Calabrese, F., Ratti, C., 2009. Eigenplaces: Analysing cities using the space - Time structure of the mobile phone network. *Environ. Plan. B Plan. Des*. 36, 824–836. doi.org/10.1068/b34133t

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 Conference on Knowledge Discovery and Data Mining*. pp. 1–8. doi.org/10.1145/2346496.2346498

Voltersen, M., Berger, C., Hese, S., Schmullius, C., 2014. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. *Remote Sens. Environ*. 154, 192–201. doi.org/10.1016/j.rse.2014.08.024

Wang, J., Deng, M., Cheng, T., Huang, J., 2012. Space-time series data analysis and modeling. Science Press, Beijing.

Wang, Y., Wang, T., Tsou, M.H., Li, H., Jiang, W., Guo, F., 2016. Mapping dynamic urban land use patterns with crowdsourced geo-tagged social media (Sina-Weibo) and commercial points of interest collections in Beijing, China. *Sustain*. 8, 1–19. doi.org/10.3390/su8111202

Wu, Q., Wu, Z., Zhang, L., 2019. GPS trajectory agglomeration and refined road network extraction. *Cehui Xuebao/Acta Geod. Cartogr. Sin*. 48, 502–511. doi.org/10.11947/j.AGCS.2019.20180256

Xu, W., 2020. Urban traffic network structure identification and congestion analysis based on complex network. Wuhan University. Wuhan, China. doi.org/10.27379/d.cnki.gwhdu.2020.000076.

Yang, W., Ai, T., 2017. The extraction of road boundary from crowdsourcing trajectory using constrained delaunay triangulation. *Cehui Xuebao/Acta Geod. Cartogr. Sin*. 46, 237–245. doi.org/10.11947/j.AGCS.2017.20160233

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci*. 31, 825–848. doi.org/10.1080/13658816.2016.1244608

Yao, Y., Yan, X., Luo, P., Liang, Y., Ren, S., Hu, Y., Han, J., Guan, Q., 2022. Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf*. 106, 102664. doi.org/10.1016/j.jag.2021.102664

Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 186–194. doi.org/10.1145/2339530.2339561

Yuan, N.J., Zheng, Y., Xie, X., Wang, Y., Zheng, K., Xiong, H., 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Trans. Knowl. Data Eng*. 27, 712–725. doi.org/10.1109/TKDE.2014.2345405

Zhang, X., Du, S., Wang, Q., 2017. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS J. Photogramm. Remote Sens*. 132, 170–184. doi.org/10.1016/j.isprsjprs.2017.09.007

Zhang, Y., 2020. Urban Functional District Extraction and Analysis from Multi-Source Data. Wuhan University. Wuhan, China. doi.org/10.27379/d.cnki.gwhdu.2020.000732.

Zhang, Y., Li, Q., Huang, H., Wu, W., Du, X., Wang, H., 2017. The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China. *Remote Sens*. 9, 865. doi.org/10.3390/rs9090865

Zhang, Y., Qin, K., Bi, Q., Cui, W., Li, G., 2020. Landscape patterns and building functions for urban land-use classification from remote sensing images at the block level: A case study of Wuchang District, Wuhan, China. *Remote Sens*. 12, 1831. doi.org/10.3390/rs12111831

Zhang, Y., She, T., Deng, M., Zhou, F., 2020. A method of multi-temporal incremental extraxtion of road network based on low-frequency trajectory data. *Sci. Surv. Mapp*. 45, 192–198. doi.org/10.16251/j.cnki.1009-2307.2020.04.027

Zheng, L., Xia, D., Zhao, X., Tan, L., Li, H., Chen, L., Liu, W., 2018. Spatial–temporal travel pattern mining using massive taxi trajectory data. *Phys. A Stat. Mech. its Appl*. 501, 24–41. doi.org/10.1016/j.physa.2018.02.064

Zheng, Y., Liu, Y., Yuan, J., Xie, X., 2011. Urban computing with taxicabs, in: *Proceedings of the 2011 ACM Conference on Ubiquitous Computing*. pp. 89–98. doi.org/10.1145/2030112.2030126

Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G., 2016. Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart card data. *PLoS One*. 11, 1–17. doi:10.1371/journal.pone.0149222

Zhong, C., Huang, X., Müller Arisona, S., Schmitt, G., Batty, M., 2014. Inferring building functions from a probabilistic model using public transportation data. *Comput. Environ. Urban Syst*. 48, 124–137. doi.org/10.1016/j.compenvurbsys.2014.07.004