IDENTIFYING URBAN DISTRICTS AND SPATIAL COMBINATION OF FUNCTIONS WITH FISHERS' EXACT TEST – A CASE STUDY WITHIN THE SIXTH RING ROAD OF BEIJING, CHINA

Disheng Yi (1), Jing Yang (1), Jing Zhang (1)

[1] Capi Norm Univ., 105 W3th Ring Road North, Haidian district, Beijing, 100048, China
Email: 2180902094@cnu.edu.cn; 2173602025@cnu.edu.cn; zhangjings@mail.cnu.edu.cn

KEY WORDS: Urban district, POI data, Fishers' Exact Test, K-modes clustering

ABSTRACT Urban areas involve different functions which attract individuals and fit personal needs. Understanding the distribution and combination of these functions in a specific district is significant for urban development in cities. At the same time, the information extracted from POI tags is used to express those functions in urban areas in recent years. Many researchers have already studied the methods of identifying the dominant functions in a district. This kind of district is also regarded as urban district. However, most of them only consider the contribution of a particular function for one district, while they ignore the quantity of each tag and combinations of different tags in different districts. Thus, we proposed an improved method to determine the urban districts by types of function in each district and discover the different groups of them. To begin with, we define a functional score based on three statistical features: p-value, odds-ratio and the frequency of each POI tag. P-value and odds-ratio are resulted from a statistical significance test, Fisher's exact test. The higher score we get, the more obvious degree of spatial collection of one tag in a district is. Commonly, this degree reflects the representativeness of one function in the district. Next, we run a K-modes clustering algorithm to classify all urban districts in accordance with the score of each function and their combination in one district and detect four different groups in the study area, namely, *Work and Tourism Mixed-developed district*, *Mixed-developed Residential district*, *Developing Greenland district* and *Mixed Recreation district*. In addition, the number of group we determined passes the evaluation with the highest Silhouette Coefficient score and Calinski-Harabasz Index score. Finally, we used an assessment called Coincidence Degree to evaluate the accuracy of classification that compares the results of some random samples with real situation. In our study, the total accuracy of identifying urban districts is 83.7%. The spatial distributions of different groups show a huge diversity. The first group (*Work and Tourism Mixed-developed district*) and the fourth group (*Mixed Recreation district*) are widespread throughout the whole study area; the second group (*Mixed-developed Residential district*) lies near the subway lines; the third group (*Developing Greenland district*) distributes far from the downtown between the Fourth Ring Road and the Sixth Ring Road. Overall, the proposed identifying method can be used for detecting urban functions and their spatial combination given by types of tags. It is an additional method in various method of identifying functions. Interestingly, the functions of four groups of urban districts we discovered are consistent with actual humans' daily activities. Also, urban spatial structure can be analyzed simply, which has certain theoretical and practical value for urban geospatial planning.

1 INTRODUCTION

Urban functions such as residential, commercial, industrial, transportation and business regions influence human activities in some aspects. It is natural that most of these functions would not appear as a single function in a

[1] This work was supported by the Open Project Program of the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant No.011177220010020 and No. 01119220010011)

particular area. Many previous researches show that these functions are coexistent and several such relationships are often going on within every area in cities. Each function has its own characteristics that we can easily discover through various data. However, identifying their spatial combination is not quite easy. Being able to finish this work gives us a better opportunity to answer some important questions on relationships between human and urban environment. For example, in discovery of different urban districts with different urban functions.

Geo-big data base that match different methods are crucial to development of identifying urban functions and urban districts. For example, remote sensing techniques monitor the physical natures of different kinds of objects to identify specific kind of urban land use (Hu, 2016; Zhang, 2017), while it is not suitable to discover the spatial relationships of functions and fail to understand socioeconomic environment (Gao, 2017). Flow data from GPS and smartcard also could reflect some urban districts caused by human activities at the spatiotemporal level, while they cannot find quantitative relationships of different urban functions in the districts. Increasing amounts of data on points of interest (POIs) is becoming available online. Many researchers have developed heaps of place-based studies that employ POIs which are able to lead to a better understanding of individual-level and social-level utilization of urban space. We could understand land use planning not only at the semantic level but also at the quantitative level using POI data.

Faced with massive data, there is no clear and distinct cognition about the actual meaning of an object and its function because of its vague and non-standard category. However, abundant semantic information could reflect the essence of urban functions and human activities. It becomes an interesting branch in the study of data mining, also is a powerful tool for extracting urban districts in recent years. Most of studies built a LDA model to discover the semantic information from POIs that they really need and then identify different urban districts in accordance with those semantics, such as Xing et al. use semantics and building attributes and identify four kinds of districts (Xing, 2018), Guo et al. combine semantics with the number of check-ins in different time intervals and discover urban districts (Guo, 2018) and the combination of semantics and the origins-destinations flows (Wang, 2018). Another example is employing Bayesian model to connect the spatial objects and zone functions and their semantic information extracted from POIs (Zhang, 2017).

Although semantic information helps to discover urban functions precisely, reclassification on POI categories is also a good solution to improve the accuracy of identification. As a result, large studies focus on finding a general quantitative method to identify urban districts in this field particularly topical. It is more simple-calculated and swift than the methods mentioned above, also with high accuracy. For instance, some of them calculate the density of each POI tag in a districts to represent the importance of tag (Zhao, 2011; Chi, 2016). Furthermore, there is an improved method on the basis of the densities, which consider the influence of the total number of each tag (Kang, 2018; Hu, 2019). Zhao et al. apply entropy weight method and mean square deviation method to measure the functional strength of development land and make a comprehensive assessment of the land type (Zhao, 2018). In addition, some previous studies find urban districts using other kinds of geo-big data and define these districts with these quantitative identification methods in order to improve accuracy. For example, Zhai et al. detected urban functions with spatial context distances and introduced an improved method beyond Word2vec which is a complement on the calculation of POIs' density (Zhai, 2019). However, such rich multiple datasets that complement each other in the same city, especially high-precision data, are usually hard to fully access.

Overall, there is few research considering how a particular function appear in a district, which means they ignore the collection of POI tag representing this function in other districts and how other functions influence the collection of this POI tag. In other word, the functions of district should be judged on the collection of their POI tags and their collection in other districts. To address this issue, here we introduce an exact functional test for directional association by examining the spatial distribution of POIs. Then, we will calculate a functional score for each POI tag in an urban district and classify those districts based on the combination of those scores.

The contributions of this study are as follows:

- We propose a quantitative method to discover urban functions by a kind of statistical significance test, Fisher's exact test. It can combine the relative functions and relative districts efficiently.
- We run a K-modes clustering algorithm to classify all urban districts according to the functional scores and their combination in one district and detect four different groups in the study area.

The remainder of this article is structured as follows. Section 2 discusses the datasets used and the selection of study areas. Section 3 introduces the methods used in our methods. In Section 4, we present the results of clusters, classification accuracy and compare urban districts. Next, we discuss the evaluation of identification using the functional score in Section 5 before concluding and pointing out directions of future work in Section 6.

2 STUDY AREA AND DATASETS

2.1 Study Area

As the center of politics, economy, technology and transportation in China, Beijing attracts attentions from all over the world. Also, it is a suitable area with various and complete urban functions so that we select this city as our study area. Specifically, the area is within the 6th Ring Road of Beijing, including the six main administrative districts. Considering the scale of human activity and spatial distribution of functions, we consider 1000m as the research diameter and divide the whole study area into regular grids (1000*1000 square meters). After removing those grids which do not includes any POI, we regard these 2343 grids as our basic research units (Figure 1).

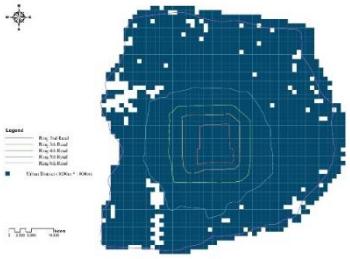


Figure 1 The study area

2.2 Data Prescription and Preprocessing

Human activity always has different topics, these topics can be seen as different functional tags in people's lives, such as workplaces, restaurants, entertainment, schools and so forth. For the sake of location and geographical representation, Point of interest (POI) has become a popular expression in this digital age recently. It can be understood as a position that easy to be found where people are interested in, a kind of basic geographic entity that is used to navigation, smart transportation and location-based services. We employ the POI dataset from BaiduMap API, which includes 383,327 POIs in the study area at 2016. Each piece of record in the dataset has its own attributes like its name, location, latitude, longitude, category and so forth. The raw POI classifications includes sixteen categories, each category also includes several sub-categories, which were confused and redundant. Considering the specific needs for urban functions, we reclassified those POIs into 10 categories after removing some infrastructure (like toilets and ATMs) that barely influence people's daily life. Each new category is clear for expressing urban function that human activities need.

3 METHODS

3.1 Calculation of Functional Score

As for urban functions, we aim at a simple and swift method to express its representativeness for a general way. We already know that P-value we have got from Fisher's exact test can provide us a straight view on how one POI tag aggregates in the district. This degree of collection can be seen as a kind of density of POIs, which is a relatively high density in this district compared with other kinds of POI tags also with same kind in other districts. However, there is a problem that the degree of collection of POI tags in one district might be similar. Only using p-value cannot mark off the density of two different tags. As a result, a functional score has been introduced in this study, and it is made by three statistical features: p-value, odds-ratio and the frequency of each POI tag.

3.1.1 P-Value This value is derived from Fisher's exact test. Section 1 mentioned that the relationships of number and geographical distance of different types of POIs produces an effect on where the POI locates and how many kinds of POIs it surrounds. The association of different POI tags can measure the functional dependency, furthermore, to answer some important questions, for example, the functional combination in urban districts. However, when one has no prior information about the functional form of a directional association, there is not a widely established statistical procedure to detect such an association.

Fisher's exact test (FET) is a statistical significance test for independence as opposed to association in 2 ×2 contingency tables (Table 1), a typical situation where such tables arise is where we have counts of individuals categorized by each of two dichotomous attributes (Sprent, 2011). To perform the test, one calculates these probabilities for all possible n_{11} consistent with the fixed marginal totals and computes the p-value as the sum of all such probabilities that are less than or equal to that associated with the observed configuration.

The most popular application of FET is observing what kinds of variances influence the incidence of disease or the level of healthcare. More importantly, some researches has already applied FET to make a detector, discovered the functional dependency of genes in biological system (Zhong, 2018).

Therefore, p-value calculated from Fisher's exact test is suitable for our study to assess urban functions. This is because 1) It would give us a probability of the number of a POI tag in a district, which means the degree of collection of this tag in a spatial region; 2) It works for small observations and calculates the exact probabilities rather than approximations such as in Chi square test. This low p-value provides very strong evidence of association. In our study, Variance one is one POI tag, and variance two is one specific district. Therefore, A in Table 1 is the number of one POI tag in one district, and B is the number of other kinds of POI tags in the same district, C is the number of same POI tag in the other districts and D is the number of other kinds of POI tags in the other districts, $P_{t,g}$ represents the p-value of the POI tag t in g district we used in this study, the equation (Equation (1)) of it is as below,

$$P_{t,g} = \frac{N_t!N_g!(N-N_t)!(N-N_g)!}{n_t!(N_t-n_t)!(N_g-n_t)!(N-N_t-N_g+n_t)!N!} \quad (t = 1, 2, \ldots, 10;\ g = 1, 2, 2343) \tag{1}$$

where n_t means the number of POI tag t in g district; N_t represents the total number of POI tag t in the whole study area; we use N_g to express the total number of POIs in g district; and finally, N is the total number of POIs.

Table 1 The 2 × 2 contingency table represents two different attributes used in Fisher's exact test

	Variance one	Non variance one
Variance two	A	B
Non variance two	C	D

3.1.2 Odds-Ratio Because it might happen that the two different tags in the same district has similar p-value, we need another index to evaluate which p-value is more reliable. In recent years, odds-ratios have become widely used in medical reports. The odds-ratio is introduced as the ratio of the probability that the event of interest occurs to the probability that it does not. The reasons why we selected it for evaluation are 1) It provide an estimate (with confidence interval) for the relationship between two binary ("yes or no") variables; 2) It enable us to examine the effects of other variables on that relationship, using Fisher's exact test. $O_{t,g}$ (Equation (2)) is seen as the odds-ratio of the POI tag *t* in *g* district.

$$O_{t,g} = \frac{n_t!(N-N_t-N_g+n_t)!}{(N_t-n_t)!(N_g-n_t)!} \quad (t = 1, 2, \ldots, 10;\ g = 1, 2, 2343) \tag{2}$$

where n_t means the number of POI tag *t* in *g* district; N_t represents the total number of POI tag *t* in the whole study area; we use N_g to express the total number of POIs in *g* district; and finally, *N* is the total number of POIs.

3.1.3 The Frequency of Each POI Tag The third statistical feature is the frequency of each POI tag. It calculates the rate of one tag happens or is repeated in a district. This part helps our functional score focusing more on the intra-regional distribution of POIs. Because of its rich application in identifying functional districts in cities, we borrowed their idea and defined F_t (the frequency of each POI tag) as a weight as one part of functional score, the equation of F_t (Equation (3)) is below:

$$F_t = \frac{n_t}{N_t} (t = 1, 2, \ldots, 10) \tag{3}$$

where n_t means the number of POI tag *t* in *g* district; N_t represents the total number of POI tag *t* in the whole study area.

All three features are introduced out in detail, we need to mix them as our functional score, and the final equation (Equation (4)) is below:

$$S_{t,g} = \frac{1}{P_{t,g}} \times O_{t,g} \times F_t \quad (t = 1, 2, \ldots, 10;\ g = 1, 2, 2343) \tag{4}$$

3.2 K-Modes Clustering Algorithm

Now we have already calculated the functional score of each POI tag in a specific district. If we consider the distribution of different POI tags in a particular district, we might discover something new through combining these scores in the district. Previous studies shows that k-means clustering algorithm has a good performance on classifying urban district, which is well known for its efficiency in clustering large data sets (Zhao, 2011; Guo, 2018; Zhai, 2019; Zhang, 2018). However, in our study, k-means cannot run a proper result with clear boundary between clusters. This is because our functional scores in a district are with huge different value between the maximum and the minimum. Also, the category of POIs is large, it caused k-means algorithm in such areas as data mining where large categorical data sets are frequently encountered. To tackle the problem of clustering large categorical data sets in data mining, the k-modes algorithm was proposed by Huang (1998). It is a modified version of the k-means algorithm that uses a simple matching dissimilarity measure for categorical variables, modes instead of means for clusters, and a frequency-based method to update modes in the clustering process. The other benefit of

k-modes clustering is that the time cost is minimizing. As a result, we chose k-modes clustering algorithm to identify urban district in our study.

4 RESULTS

4.1 Classification of Urban Districts

As proposed in Section 3, we first run a Python program in order to calculate the functional score of every tag in a district. However, we found that the scores of different tags with huge different value are in the same district. As a result, a normalization must be involved before we classified urban districts based on the combination of scores. There are lots of normalization method such as max-min normalization and zero-mean normalization. We decided to choose max-min normalization method because it is the most common one with wide applications. More importantly, this method is suitable for those data that do not fit in with Gaussian distribution and it can smooth high-deviation data and improve the accuracy of classification.

K-modes clustering method need a given number as the number of clusters, the most essential thing is choosing the number of clusters properly in this part. Evaluating the performance of a clustering algorithm is not as trivial as counting the number of errors or the precision and recall of a supervised classification algorithm. Some measures require knowledge of the ground truth classes while is almost never available in practice or requires manual assignment by human annotators, others like Silhouette Coefficient and Calinski-Harabasz index do not. Obviously, data we used in this study do not have exact true values. We only evaluated the performance of different k as the total number of clusters for urban districts using two common measures just as we mentioned above. The first one, Silhouette Coefficient, is bounded between -1 for incorrect clustering and +1 for highly dense clustering, where a higher one relates to a model with better defined clusters. Another is faster to compute. Just like Silhouette Coefficient, the higher score we get, the better clusters we have. Figure 2 shows that the results of two measures evaluating how many clusters make the k-modes clustering perform better by setting the range of value k from 2 to 20 (Zhai, 2019; Zhang, 2018). Fortunately, both results tell us that the scores are peaking to the highest value when k = 4. We selected k = 4 as the ideal k value for further analysis and validation, where a different color indicates a different urban district category (Figure 3).

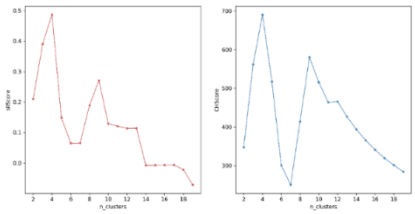


Figure 2 Performance of different k clusters in k-modes algorithm. Red line and blue line represent the evaluation of Silhouette Coefficient and Calinski-Harabasz index, respectively.

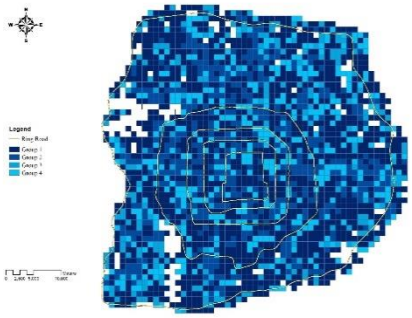


Figure 3 The result of identification and classification of urban districts. Group 1-4 are different type of urban districts based on the combination of POI data.

4.2 Identification and Annotation of Urban Districts

As shown in Figure 3, we can see four different groups of districts with different characteristics. They are Work and Tourism Mixed-developed district, Mixed-developed Residential district, Developing Greenland district and Mixed Recreation district.

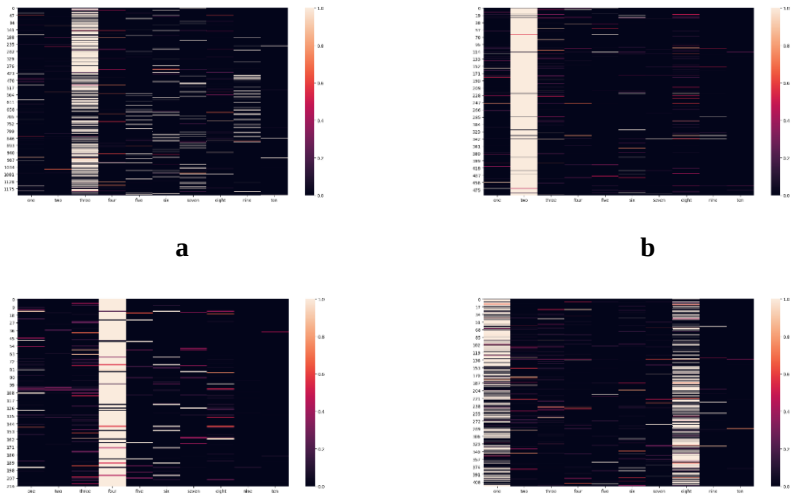
Specifically, Work and Tourism Mixed-developed district (group one), this kind of district is a developed area which relies on work as its main urban function, at the same time, some tourism industry is also located in this group as shown in Figure 4a. For example, the districts where encloses Guomao or Beijing CBD (the most famous business area) and the Palace Museum (the most gorgeous attraction) are both classified as Work Developed district. The number of it is the largest one in four groups as well. However, the urban functions of these districts are completed mixed with many other urban functions related to people's lives such as hospitals. The spatial distribution of them are basically throughout the whole study area (Figure 5a).

Second one is Mixed-developed Residential district, which is with the second largest number. Just like its name, the main function in this group is residence (Figure 4b). There are some commons and differences between group one and group two except their combination of urban functions. At the beginning, they are full of urban functions which are completed for people's lives. While most of functions appeared in group one more likely composed people's working time like public services, higher education and hospital. If someone often checks in at the districts of group one, they probably have a job at this place. Compared to the group one, the functions we discovered in the districts of group two are closer to people's daily lives such as kindergartens and parks. Interestingly, the geographical locations of these districts help us interpret this difference. Because they are distributed around the existing subway lines in Beijing (Figure 5b).

Developing Greenland district is the only kind of district that is not developed in this study, because the most districts with transportation are in this group. According to its combination graph (Figure.4c) of functions, we easily understand the development of these districts where transportation plays a vital role there as well as such buildings like natural parks and university campus with large land area almost cover the entire district. This kind of district do not have various urban functions related to people's needs, they are able to be seen as virgin area with potentials of urbanization. For instance, district No.1662 covers a part of Olympic Forest Park where is rich in forest resources with a green coverage of 95.61%, of course, it cannot exist such other urban functions except public transportation. As for the spatial distribution, they are scattered over the area from the Ring 4th Road to the Ring 6th Road (Figure 5c). Besides, there is one district classified into this group needed our attention because its location lied inside the Ring 3th Road. It is a green land set by government in order to increase city green space and improve air pollution.

The last group is Mixed Recreation district. Figure 4d illustrates that this group's functions mainly includes entertainments and public services which improve cities for leisure and recreation. Other functions such as work

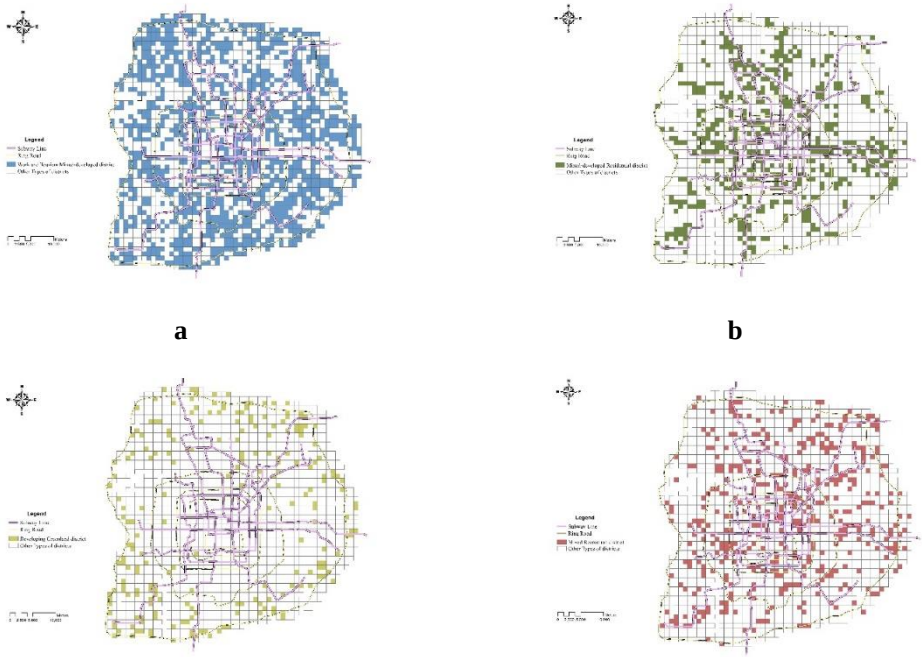
and residence are also discovered in this kind of district, while their proportion are much smaller than the main functions. We can consider them as a mixed district with functions of leisure and recreation. The two typical example, Houhai and Sanlitun, well-known as entertainment area, both appears in the district of this group. The geographical distribution is similar like group two as shown in Figure 5d, however, the number of group four at the suburb between the Ring 4th Road and the Ring 6th Road is smaller than group two.



a **b**

c **d**

Figure.4 The combination of functional scores for four groups of urban district, including a) Work and Tourism Mixed-developed district; b) Mixed-developed Residential district; c) Developing Greenland district; d) Mixed Recreation district, where each line represents one district and the color bar beside each graph shows that the functional scores after normalization.



a **b**

c **d**

Figure 5 The spatial distribution of each group of urban district, a) Work and Tourism Mixed-developed district; b) Mixed-developed Residential district; c) Developing Greenland district; d) Mixed Recreation district.

5 DISCUSSION

In order to test the accuracy of our results of identification and classification by functional scores, this study decided to compare the urban districts we get with the actual condition in accordance with BaiduMap. We used an accuracy assessment named Coincidence Degree to evaluate the accuracy of classification (Kang, 2018). In this assessment, each district will be assigned a 0-3 round number. When a district is assigned to 3, it means that the type of group which it belongs to completely fit with the real land use situation, and 2 means nearly fitting, 1 means slightly fitting as well as 0 means barely fitting. As a result, the C (Equation (6)) is regarded as Coincidence Degree to calculate the total accuracy,

$$C = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} X_i} \times 100\%, \quad (6)$$

where n is the total number of samples, x_i represents the actual mark of district i and X_i represents the full mark of district i.

There is a total of 45 districts selected at random as our samples. Table 2 shows the evaluation of all samples, it can be seen that the overall accuracy rate of identification of urban districts in Beijing has reached 83.7%, which indicates that the functional score we introduced can effectively discover urban functions and districts with a good result.

Table 2 Assessment of Coincidence Degree for urban districts.

The number of districts got 0 mark	The number of districts got 1 mark	The number of districts got 2 mark	The number of districts got 3 mark	Total number of all districts
3	3	7	32	45

6 CONCLUSION

In this study, we first employed Fisher's exact test to improve some common methods which are based on the frequency of POIs' count. This step we get a functional score for the purpose of doing a quantitative analysis of each type of POI tag. Next, we run a k-modes clustering algorithm to discover different groups of urban districts. It is worth to mention that the combination of those scores that belongs to the same district is a prerequisite for our identification and classification of urban districts. We discover that urban functions are mostly mixed in a range of 1000*1000 square meters within the 6th Ring Road of Beijing. There are four groups we found in this study, according to their combination of those POI types, they are able to be regarded as *Work and Tourism Mixed-developed district*, *Mixed-developed Residential district*, *Developing Greenland district* and *Mixed Recreation district*. In the geographical view, only Developing Greenland district is almost scattered at the suburb away from the area inside the 4th Ring Road of Beijing. The distributions of others are all throughout the whole area. More importantly, Mixed-developed Residential district and Mixed Recreation district are closer to the subway lines. This phenomena explains that these two urban functions (residence and recreation) are dependent with public transportation. Finally, we evaluate the result of classification with coincidence degree which is a sample scoring method by human. Over 80% samples are identified by our functional score correctly and the total accuracy of classification is 83.7%.

Although we have successfully proposed an improved quantitative method related to Fisher's exact test to identify urban functions only using POI data, while the result of classification for urban districts is relatively rough. For example, the Work and Tourism Mixed-developed district includes two functions representing two status of human

activities. For registered residents, these two functions can be regarded as work. However, tourists only appear in this kind of district for attractions. Therefore, we consider adding time semantics into the identification of urban districts in the further study.

In conclusion, this study draws more attentions to utilization of land resources, improvement of land use efficiency and urbanization. It plays a significant role in decision supporting in urban planning practice. Another perspective is that urban functions are an excellent expression of human activities. Specific and accurate urban districts could help us to understand the human activities and different types of human.

REFERENCE

Chi, J., Jiao, L., Dong, T., Gu, Y., & Ma, Y., 2016. Quantitative identification and visualization of urban functional area based on poi data. Journal of Geomatics, 41(2), pp. 68-73.

Gao, S., Janowicz, K., & Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location‐based social networks. Transactions in GIS, 21(3), pp. 446-467.

Guo, Z., Zheng, Z., Liu, J., Wang, S., Zhong, P., Zhu, M., ... & Li, J., 2018. Urban Functional Regions Using Social Media Check-Ins. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium pp. 5061-5064.

Hu, T., Yang, J., Li, X., & Gong, P., 2016. Mapping urban land use by using landsat images and open social data. Remote Sensing, 8(2), pp. 151.

Hu, Y., & Han, Y., 2019. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. Sustainability, 11(5), pp. 1385.

Kang, Y., Wang, Y., Xia, Z., Chi, J., Jiao, M., & Wei, Z. W., 2018. Identification and classification of Wuhan urban districts based on poi. J. Geomat, 43, pp. 81-85.

Sprent, P., 2011. Fisher exact test. International encyclopedia of statistical science, pp. 524-525.

Wang, Y., Gu, Y., Dou, M., & Qiao, M., 2018. Using spatial semantics and interactions to identify urban functional regions. ISPRS International Journal of Geo-Information, 7(4), pp. 130.

Xing, H., & Meng, Y., 2018. Integrating landscape metrics and socioeconomic features for urban functional region classification. Computers, Environment and Urban Systems, 72, pp.134-145.

Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z. R., & Gu, C., 2019. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. Computers, Environment and Urban Systems, 74, pp. 1-12.

Zhang, X., Du, S., & Wang, Q., 2017. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. ISPRS Journal of Photogrammetry and Remote Sensing, 132, pp. 170-184.

Zhang, X., Li, W., Zhang, F., Liu, R., & Du, Z., 2018. Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data. ISPRS International Journal of Geo-Information, 7(12), pp. 459.

Zhao, M., Liang, J., & Guo, Z., 2018. Classifying Development-land Type of the Megacity through the Lens of Multisource Data. Shanghai Urban Planning Review, 142(05), 82-87.

Zhao, W., Li, Q., & Li, B., 2011. Extracting hierarchical landmarks from urban POI data. Journal of Remote Sensing, 15(5), pp. 973-988.

Zhong, H., & Song, M., 2018. A fast exact functional test for directional association and cancer biology applications. IEEE/ACM transactions on computational biology and bioinformatics, 16(3), pp. 818-826.