Analyzing Recent Morphological Evolution in Central Beijing during Relocating Non-Capital Functions Based on POI Data and Digital Maps

by Yishan Chen & Mingbo Wu
China IPPR International Engineering Co. LTD, Institute of Geographic Sciences and Natural Resources Research, CAS

Keywords: Morphological evolution, Point of interest (POI), Hybridity, Beijing

Abstract: After China's reform and opening up, Chinese cities have undergone large-scale spatial restructuring, which is reflected in urban renewal and urban expansion. Our study focuses on recent morphological evolution in central Beijing during relocating non-capital functions, when Beijing began to adjust the economic structure and spatial structure according to the "Beijing-Tianjin-Hebei Collaborative Planning Outline" since 2014. Based on point of interest (POI) data and digital maps, a series of indicators on urban morphology are specified to classify urban blocks, from the perspectives of the built environment as well as land use. And hybridity, a key morphological characteristic in traditional cities, is measured quantitatively based on these indicators. Through the observation of the trend, extent and intensity of these changing indicators, the evolution of urban form is revealed.
Results suggest that central Beijing can be classified into three different regions: the old city, the newly-built area and the fringe belt. Three regions vary in characteristics of the built environment and land use. The two aspects of urban form are correlated, and certain types of land use tend to locate in blocks with certain built environment characteristics. The changes in land use clearly reveal the effect of relocating non-capital function during this period. However, this compulsory function relocation has led to homogenization in the old city of Beijing, leading to a possible reduction in urban vitality and convenience.

1. Introduction

During 40 years of reform and opening up (1978-2018) in China, the number of residents in Beijing has increased from 4.92 million to 21.7 million, and the urban built-up area has grown from 346 square kilometers to 1,603 square kilometers. A large amount of labor was concentrated in the city when the urban area keeps expanding and merging the nearby villages. Meanwhile, in the old city of Beijing, a proportion of traditional buildings were redeveloped and substituted by buildings of larger scales. The whole city gradually takes on a hybrid and fragmented appearance (Chen, Y.S. & Qi, X. 2018). As China's urbanization is shifting from medium quality to high quality, its driving force also is switching from capital-intensive to innovation-intensive (Fang, C.L. 2019). According to the "Beijing-Tianjin-Hebei Collaborative Planning Outline", the designated function of the city has changed. Beijing began to adjust the economic struc-

DOI 10.36158/978889295357410
www.tabedizioni.it

ture and spatial structure by means of "relocating non-capital functions" after 2014. In this research, we are trying to explore the morphological hybridity of Beijing in this period and the influence this policy shift has caused on the urban form until now.

With the development of information and communications technology (ICT), urban data are accumulating at an unprecedented speed. The explosion of data makes it difficult to understand in traditional ways (Pan, Y. H. 2018), and a number of new analysis methods are put forward in the era of big data, opening up new possibilities for urban research. Through the analysis of multi-source and fast-updated urban data, the evolution in urban form can be quantified more comprehensively and accurately, revealing the relationship between the built environment and urban function in depth.

This study aims to inspect urban form quantitatively based on point of interest (POI) data and digital maps. From two perspectives of the built environment and the land use, several block-level indicators on urban morphology are specified, and hybridity is then calculated. Finally, through comparison of these changing indicators, the evolution of urban form is revealed.

A number of scholars have worked on quantifying urban form since M.P.G.Conzen (1968) and his followers established the connection of urban form with town-plan, building types and land use. From the perspective of the built environment, the Spacematrix method uses multiple parameters of floor space index (FSI), ground space index (GSI), height and open space ratio (OSR) to distinguish between different forms of urban blocks (Berghauser Pont, M. & Haupt, P. 2007). From the perspective of land use, the mixed-use index (MXI) tool quantifies the degree of mixing in land use by calculating the ratio between housing, working and amenities in the area (Van den Hoek, J. W. 2008). With the help of big data, studies with new data source and new methods have emerged. For example, Dutch scholars studied the multi-dimensional character of neighborhoods by clustering algorithms (Gil, J. et al 2009). Long Ying's team used algorithms to construct the relevance between POI and block (Liu, X. & Long, Y. 2016). Yu Ye (2014) combined Space Syntax, Spacematrix and MXI in urban morphological analysis. At the same time, some Chinese researchers began to use big data to analyze the spatial distribution of functions in cities (Wang, F. et al 2015; Chen, W.S. et al 2016; Yang, T. 2018).

2. Methodology

2.1. *Data sources and preprocessing*

The data used in our study include a) data of urban buildings and urban blocks (parcels), which are generated from digital maps in 2015 and kindly made public by Beijing City Lab (*https://www.beijingcitylab.com/*), a virtual research community dedicated to studying China's capital Beijing quantitatively; and b) POI data from 2014 to 2018, which comes from NavInfo, a leading digital map and navigation service provider in China. POI data consist of points representing geographic entities, including latitude and longitude coordinates, address, name, and category, etc. Such data may provide wide coverage and fast update of land use information with relatively high accuracy. After purging invalid records, all the data are transformed into the same type of coordinates and cropped by the 5th ring road, which basically represents the central urban built-up area in Beijing.

Note that we have not obtained data of the built environment through different years. Nevertheless, as the built environment changes rather slowly compared with urban functions

and POI during 2014-2018 (Wegener, M. et al 1986), the influence of this lack of data is limited and reasonable conclusions can still be achieved.

2.2. *Research framework*

Conzen's method combines town plan, patterns of building form, and patterns of land use to inspect urban form (Conzen, M. R. G. 1968). As the data of land lots are hard to obtain in China, we mainly focus on two perspectives of urban morphology, the built environment and the land use. The built environment is analyzed via clustering the four Spacematrix indicators generated from digital maps, while the land use is represented by the POI data, and summarized by the Latent Dirichlet Allocation (LDA) algorithm. Next, the relationship between the built environment and land use is inspected and then hybridity of different blocks is generated. At last, the morphological evolution during relocating non-capital function is revealed through comparison of indicators in different years, and several conclusions are drawn.

In all the above analysis, we take urban blocks surrounded and separated by motorways as the spatial unit of analysis, because an urban block is a relatively coherent entity in urban morphological evolution and data of smaller units (i.e. land lots), are not available.

2.3. *The analysis of built environment based on digital maps*

Spacematrix (Berghauser Pont, M. & Haupt, P. 2007) is an efficient way of quantitative description of built environment, simultaneously representing the building density and various building types. Spacematrix takes into account four correlated indicators of block-level built environment: a) floor space index (FSI) is the sum of area of every floor in buildings in a block divided by the area of this block, which is a common indicator in Chinese urban planning; b) ground space index (GSI) is the total area of buildings footprints in a block divided by the area of this block; c) layers (L), the average story of the buildings in a block, numerically equals FSI divided by GSI; d) open space ratio (OSR) is the total area of open space, i.e. vacant space within blocks, divided by the total built area of buildings in this block. These four variables express the intensity (FSI), the compactness (FSI), the building height (L) and the pressure on non-built space (OSR) in a block respectively. The Spacematrix diagram inspects the four indicators simultaneously. The FSI on the y-axis indicates the intensity in an area and the GSI on the x-axis reflects its compactness. The OSR and L are gradients that fan out across the diagram. Thus, every block is represented by a point in the Spacematrix, and different region in the Spacematrix reflects different patterns in the built environment.

Unlike previous studies, where Spacematrix is classified subjectively into separate regions representing different building types, we utilize classic *k*-means method to cluster points in Spacematrix into different categories of building form. Clustering allows the classification of instances in multi-dimensional space where there are no classes defined beforehand. As a classic and basic clustering method, the *k*-means algorithm iteratively partitions the whole data set into a *k* number of clusters and adjust the locations of *k* cluster centers, seeking to minimize the sum of variance in each cluster (Witten & Frank 2005). In this way, Spacematrix is partitioned into separate regions less subjectively on the basis of the distribution of data, and all the blocks in our study area are classified into a certain cluster of built environment.

2.4. *The analysis of land use based on POI data*

Land use information directly acquired from planning maps usually assigns a single land use in one location. However, in a real urban environment, mixing of different land use is the usual case. Van den Hoek (2009) utilized the mixed-use index (MXI) to quantify the degree of land use mixture, based on the percentages of the gross floor area of dwellings, working places, and commercial amenities among all the floors of buildings in a block.

With the help of POI data, the information on mixing land use can be characterized more accurately with higher spatial resolution and quicker update. The POI data are provided with 47 detailed categorical identifiers, such as personal services, companies, public security agencies, medical institutions, retail, education, etc. Numbers of different types of POI points are summed into corresponding blocks and then normalized according to the area of blocks. The density of different types of POI is representative of diverse land use, revealing the distribution and agglomeration of different urban functions as well as hybridity and correlation between different land use.

However, a drawback in approximating land use with numbers of POI points lies in the varying quality and importance of POI points of different categories. Typically, information about human activities or data on the scale of POI points is needed to deal with this problem. As these data are not available, we utilized the TF-IDF (term frequency–inverse document frequency) transformation, based upon the assumption that rare types of POI are of more importance. TF-IDF is a numerical statistic in information retrieval, which is intended to reflect how important a word is to a document in a collection or corpus (Wu, H. C. et al 2008). The TF-IDF value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contain the word, giving rare words relatively more importance. Here we made an analogy between information retrieval and urban function quantification, with different kinds of POI points corresponding to different words, blocks corresponding to documents, and the entire central Beijing taken as the corpus. The transformed spatial distribution of different types of POI is then used to represent relevant urban functions.

The distribution of different kinds of POI is inter-correlated. Through pairwise correlation between different types of POI, it can be found that some POIs are closely related to each other, while some others are relatively independent. Some POI combinations represent basically the same kind of land use, such as residential area and personal service, car sales and car service. For better comprehension of land use, the original detailed categories in POI data are summarized into fewer latent types. Several methods exist to extract latent factors from observed data, such as principal component analysis (PCA) and factor analysis. In accordance with the aforementioned analogy with information retrieval, Latent Dirichlet Allocation (LDA) is utilized to discover latent topics of land use. LDA is a generative probabilistic model of a corpus (a collection of observed data) (Blei, D. M. et al 2003). The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. In our land use setting, LDA tries to model an urban block with different categories of POI as mixtures over a small number of latent topics of land use, where each topic of land use is characterized by a distribution over categories of POI. The mixture of land use topics is a better representation of the real mixing urban functions, balancing between comprehensibility and detailed precision, and changes in land use topics may reveal the evolution in urban functions.

In practice, the number of topics is an important hyper-parameter affecting the performance of LDA. And we choose the appropriate number of topics where there is a sharp drop in model perplexity or topic similarity.

2.5. *The analysis of hybridity*

Hybridity, the interweaving state of various buildings, people, and urban functions, is a result of conflicts and contradictions in the process of modernization and urban expansion in traditional cities, reflecting urban vitality. In our study, the hybridity of urban blocks is characterized by two aspects, i.e. the built environment and the land use.

From the perspective of built environment, hybridity can be quantified by the degree of mixture between different form types of individual buildings. Types of building forms are determined by the height and shape of buildings. After inspecting the overall distribution of building heights, heights of buildings are classified as "low-rise" (one or two floors), "mid-rise" (three to six floors) or "high-rise" (above six floors). Building shape is measured by area-perimeter ratio, and classified as "strip", "block" or "point" with increasing area-perimeter ratio values. Hence all the buildings are divided into nine form types and an urban block can be characterized by the volume of buildings in each of these nine form types that belong to this block. Then we measure the hybridity of built environment with Shannon entropy, a mathematical measure of the degree of randomness in a set of data in information theory and a popular diversity index in the ecological literature (Jost, L.. 2006).

From the perspective of land use, the hybridity of a certain urban block can be estimated from the degree of mixing in its associated land use topics. Similarly, Shannon entropy of the topic distribution in a block is calculated as an indicator for hybridity.

3. Analysis of Results

3.1. *The analysis of built environment*

The block-level Spacematrix indicators are clustered into 9 types (Figure 1 B), with the cluster center values in Table 1. According to the spatial distribution of clusters (Figure 1 A), three different regions are revealed: old city area, newly-built area, fringe belt. The old city is the historical region surrounded by the ancient walls of the Ming and Qing Dynasties. The newly-built area is the region of urban expansion since the founding of the People's Republic of China. The fringe belt is the interweaving urban and rural areas, where the urban characteristics are gradually reduced away from the city center.

It can be found that the FSI of the building in Beijing is low in the center, high in the newly built area, and the fringe belts are lower again, forming a ring pattern. As shown in Figure 1 A, the scale of the urban blocks gradually increases from the old city to the fringe belt. This indicates that the closer the location is to the city center; the finer division of the land plot is, and vice versa (Table 3).

Due to the hybrid types of building in the block of Beijing, the value of each cluster corresponds to several building types (Figure 1 C). However, Chinese urban land is state-owned, and urban buildings tend to show different characteristics at different times. The form of buildings can be judged roughly by the different built times in the three regions. In the old city, the historical area is of carpet shape, that is, a kind of high-density courtyard building, while some redevelopment areas are high-rise blocks or strips. The newly built area mainly exhibits a high degree of homogeneity with medium density, which is characterized by mid-rise strip buildings. There are still some blocks with low-density forms such as parks and urban villages, and

Table 1. *The index of 9 block clusters.*

Cluster	0	1	2	3	4	5	6	7	8
FSI	0.048700	0.382171	0.918825	1.128469	1.264736	2.037636	2.601148	4.534147	7.295834
GSI	0.032020	0.153036	0.467038	0.299068	0.641320	0.247713	0.415441	0.290448	0.621224
L	1.660942	2.671250	1.987091	3.825410	1.979606	8.369767	6.301446	15.966836	12.137201
OSR	30.898885	2.521109	0.621259	0.667635	0.321749	0.383549	0.231832	0.163752	0.050629

Table 2. *Relationship between region and building type.*

	0	1	2	3	4	5	6	7	8
Old city	Park	Low-rise strip	Less Carpet+ strip	More Carpet-strip	Strip+ high-rise point	Mid-rise block	High-rise strip1	High-rise strip2	High-rise block
Newly built area	Low-rise point	Low-rise strip	Mid-rise point	Carpet+ strip	Strip+ high-point	Mid-rise strip+ high-rise point	High-rise point	High-rise block	X-High-rise block
Fringe belt	Low-Rise point	Low-rise strip	Less Carpet+ strip	More Carpet-strip	Mid-rise block+ mid-rise				

Table 3. *Relationship between region and cluster*

	Main cluster	Minor cluster	Block size
Old city	2\3	0\1\4\5\6\7\8	Small
Newly built area	4\5	0\1\2\3\6\7\8	Medium
Fringe belt	1\2	0\3\4	Large

higher-density forms such as business centers. The form of the fringe belt is the result of the interlaced development of villages, farmland and mid-density residential areas. (Table 2).

3.2. *The analysis of land use*

In Figure 2, urban blocks are colored according to the density of POI of a certain type. The spatial distribution of POI is related to its characteristics. Some public welfare and historical functions are only distributed in specific areas. For example, most of the scenic spots are distributed in the old city, international organizations only appear in the embassy district, and public security organizations are evenly distributed in cities.

The aggregation of POI in some categories has a clear tendency, which reflects the layout of urban planning. For example, culture, social groups, education, and scientific research are concentrated in the northwestern built-up area because the university town is in this area.

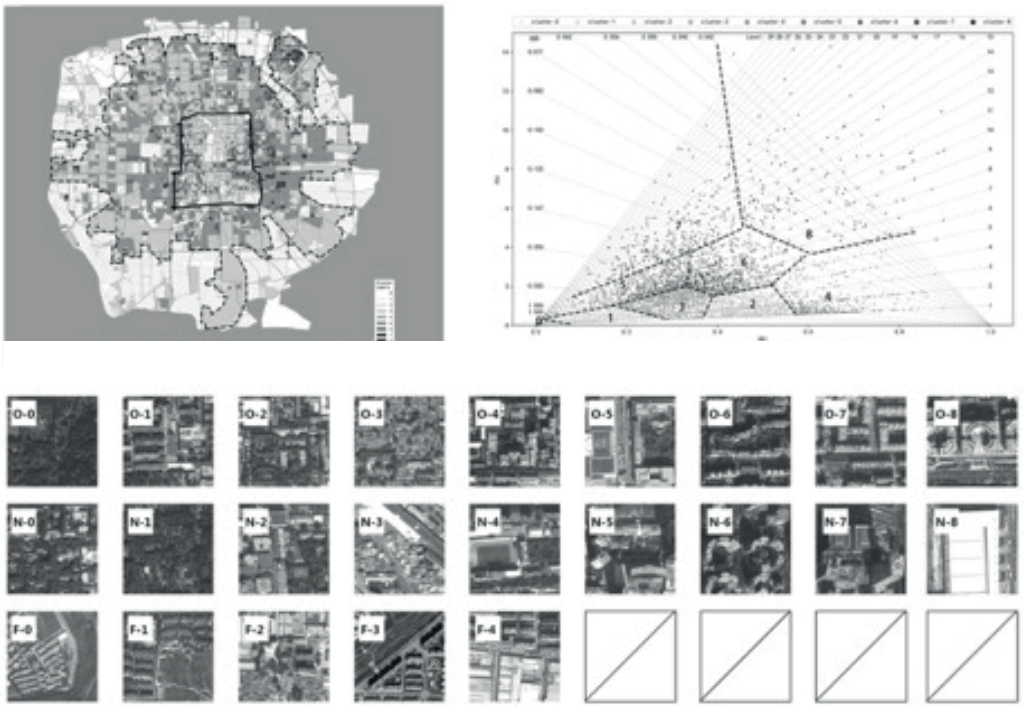

Figure 1. *The spacematrix index, the distribution of the block cluster and remote sense image (a. The distribution of the block clusters; b. The Spacematrix index of the 9 clusters; c. The Remote image of the 9 clusters).*

Commercial facilities are concentrated in Wangfujing-CBD, Xidan-Finance Street, Zhongguancun, and car services are gathered at the edge of the southern city. And profitable POI has a tendency to gather in the center, such as commercial accommodation, children's retail, restaurants, cultural retail.

With the aforementioned method, nine topics of land use are derived from LDA (Table 4). Through the distribution of each topic over different types of POI, the semantic meaning of each topic can be determined. Figure 2 F shows urban blocks with their most prominent land use topic. Different topics exhibit varying spatial preferences. Topic 8, related to "corporate companies" is prone to locate in the east and north directions between the 2nd Ring Road and 3rd Ring Road, and Topic 3 (commercial services) are widely distributed in the west and south. The old city is dominated by Topic 1 (government affairs), Topic 3 (commercial services) and Topic 5 (leisure).

3.3. *The correlation of built environment and land use*

Base on the clusters of built environment and topics of land use, the relationship between the built environment and land use can be explored (Table 5). It can be found that Topic 4 (international organizations and parks) prefers cluster 0, which is characterized by blocks with low density. Topic 7 (automobile logistics) prefers cluster 1, and the large car shops and warehouses make the building level not too high. Topic 3 (commercial and residential) prefers clusters 2 to 6, Topic 6 (residential services) prefers clusters 5 and 7, and Topic 8 (business) prefers clusters 6\7\8, which themselves must maintain a certain density. In particular, "business" is more inclined to a high-density environment.

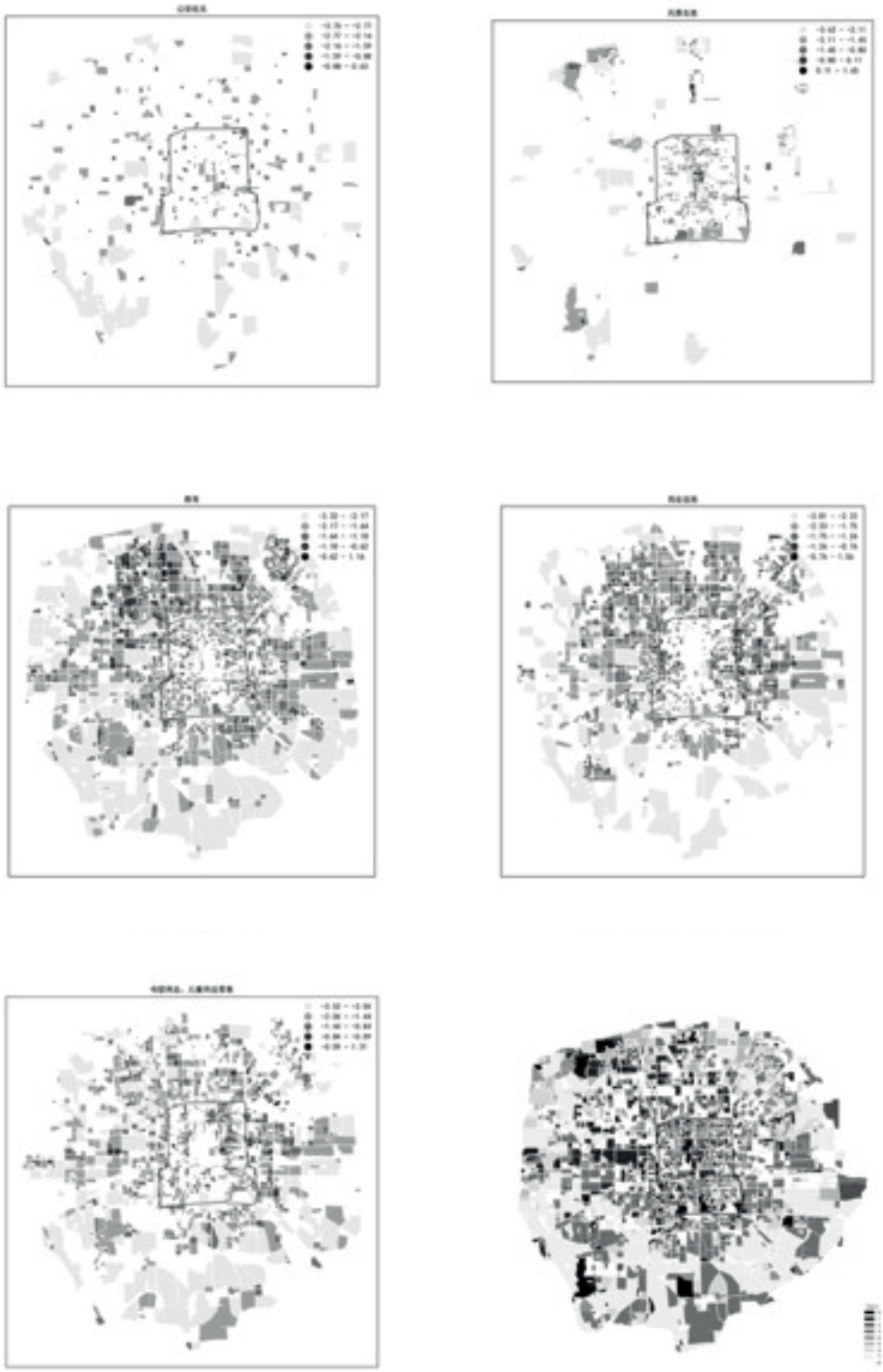

Figure 2. *POI distribution and main function distribution (a. public security organs; b. scenic spots; c. education; d. commercial facility; e. maternal and child products retail; f. the topic of land use).*

Table 4. *Relevance of POI and Topic.*

Topic		Rank1	Rank2	Rank3	Rank4	Rank5
0	POI type	Scenic spots	Medical institutions	Culture, media	Personal service	Public security organs
	Intensity	0.210335	0.110839	0.0909369	0.0526513	0.0439428
1	POI type	Government and regulatory agencies	Other units	Social groups, associations	Research institutions	Personal service
	Intensity	0.112281	0.0775246	0.0635607	0.058314	0.045888
2	POI type	Road auxiliary facilities	Personal service	Integrated retail	Fast food	Retail of beverages and tobacco and alcohol products
	Intensity	0.656289	0.0378517	0.0278673	0.0264926	0.025771
3	POI type	Personal service	Restaurant	Maternal and child products retail	Retail of beverages and tobacco and alcohol products	Fast food
	Intensity	0.147875	0.0908856	0.0898191	0.0764612	0.0724976
4	POI type	Leisure vacation	International organizations	Rental service	Scenic spots	Road auxiliary facilities
	Intensity	0.267939	0.0969138	0.0689593	0.0598811	0.0351575
5	POI type	Retail of culture, sporting goods and equipment	Casual dining	Restaurant	Leisure and entertainment	Sports venue
	Intensity	0.118951	0.110061	0.0755494	0.0753661	0.0726666
6	POI type	Residential accommodation	Education	Personal service	Road auxiliary facilities	Financial Services
	Intensity	0.205713	0.13855	0.061418	0.0563674	0.0497855
7	POI type	Passenger and cargo transportation	Car sales and service	Retail of automobiles and related products	Retail of hardware, furniture and interior decoration materials	the company
	Intensity	0.0867158	0.0850847	0.0810609	0.0690214	0.0569026
8	POI type	The company	Financial Services	Commercial facility	Restaurant	Road auxiliary facilities
	Intensity	0.209603	0.0853332	0.0792545	0.0418569	0.0389604

3.4. *The analysis of hybridity*

From the perspective of built environment and land use, the degree of hybrid can be measured. On the function side, the hybridity represents functional diversity. The higher the degree of hybridity, the more diverse the function representing the street block. From Figure 3 A/B, it can be seen that the old town and the urban fringe are more hybrid than the newly built area. The hybridity of the old city is reflected in the high degree of integration of urban functions after a long time, while the hybridity of the urban fringe belt is reflected in the intertwining agricultural and urban functions. In contrast, the newly built-up area showed homogeneity, which is probably affected by the zoning in modern urban planning.

On the built environment side, hybridity implies the mixing of different building types (Figure 3 C/D). High value of hybridity means that different types are located in the same street block. In the old city, hybridity has relatively low value, indicating the townscape more

Table 5. *Relevance of built environment cluster and land use Topic.*

Cluster \ Topic	0	1	2	3	4	5	6	7	8
0	0.132030	0.068915	0.116309	0.139191	0.179405	0.097928	0.065056	0.146132	0.055034
1	0.107460	0.099910	0.071193	0.115681	0.109350	0.096168	0.092850	0.214517	0.092870
2	0.120778	0.119233	0.044792	0.304577	0.043504	0.125617	0.066901	0.098440	0.076159
3	0.089390	0.153290	0.042706	0.213713	0.043823	0.069446	0.139506	0.122322	0.125803
4	0.095402	0.114566	0.043379	0.345664	0.044404	0.141347	0.058959	0.086864	0.069414
5	0.044617	0.113048	0.034312	0.226000	0.040896	0.067602	0.213924	0.064808	0.194794
6	0.071567	0.160073	0.037311	0.213812	0.036363	0.104025	0.107195	0.058322	0.211331
7	0.047277	0.063575	0.045426	0.186670	0.046475	0.059846	0.217816	0.040032	0.292883
8	0.055979	0.079463	0.041632	0.213389	0.042810	0.126115	0.074951	0.043839	0.321821

Note: the number in grey cells is the most two dominating topic in each cluster of built environment.

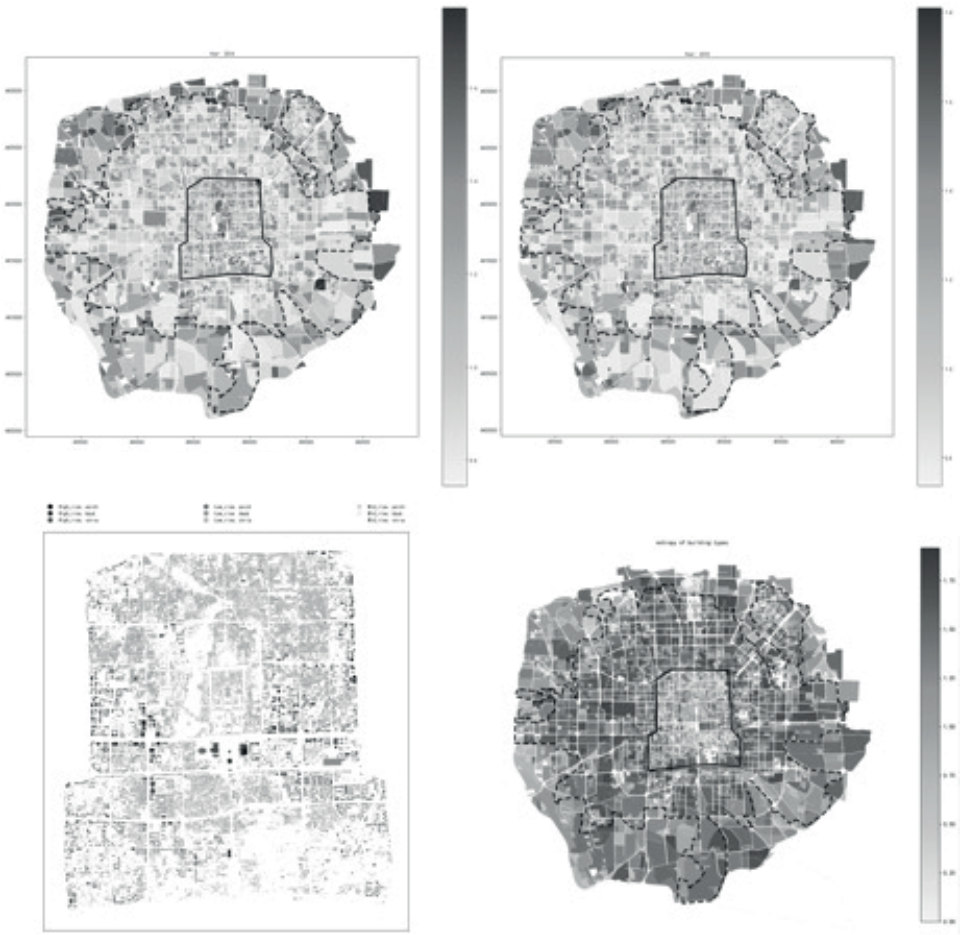

Figure 3. *POI distribution and main function distribution (a. Hybridity of POI in 2014; b. Hybridity of POI in 2018; c. The type of building; d. Hybridity of building types).*

uniform. As the block size becomes larger from the newly built area to the fringe belt, more categories of buildings are included in the same block. By contrast, the blocks in the newly built area are most hybrid, containing high-rise buildings, mid-rise strips, as well as low-rise buildings. These particular hybrid blocks may reflect the DANWEI unit, which is a special Chinese working and living form.

3.5. *The morphological evolution*

Under the guidance of the Beijing-Tianjin-Hebei Collaborative Development Plan, Beijing carried out actions such as "relocating non-capital functions" and "Developing Tongzhou Subcenter". By comparing the topics of land use between 2014 and 2018 (Figure 4), the influence of such policies may be revealed. Topic 1 (government function) has decreased in the old city and increased in the newly built area, which reflects a shift in the government function from municipal to state government. Topic 3 of "commercial and residential services" has been reduced greatly in a large part in the old center in comparison with the newly-built area, reflecting the impact of the relocating action. Topic 6, related to "residential services", have been strengthened in newly built areas, a sign of improvement of living conditions. Topic 7 of "automobile logistics" was gradually substituted, impacted by the relocation policy in Beijing.

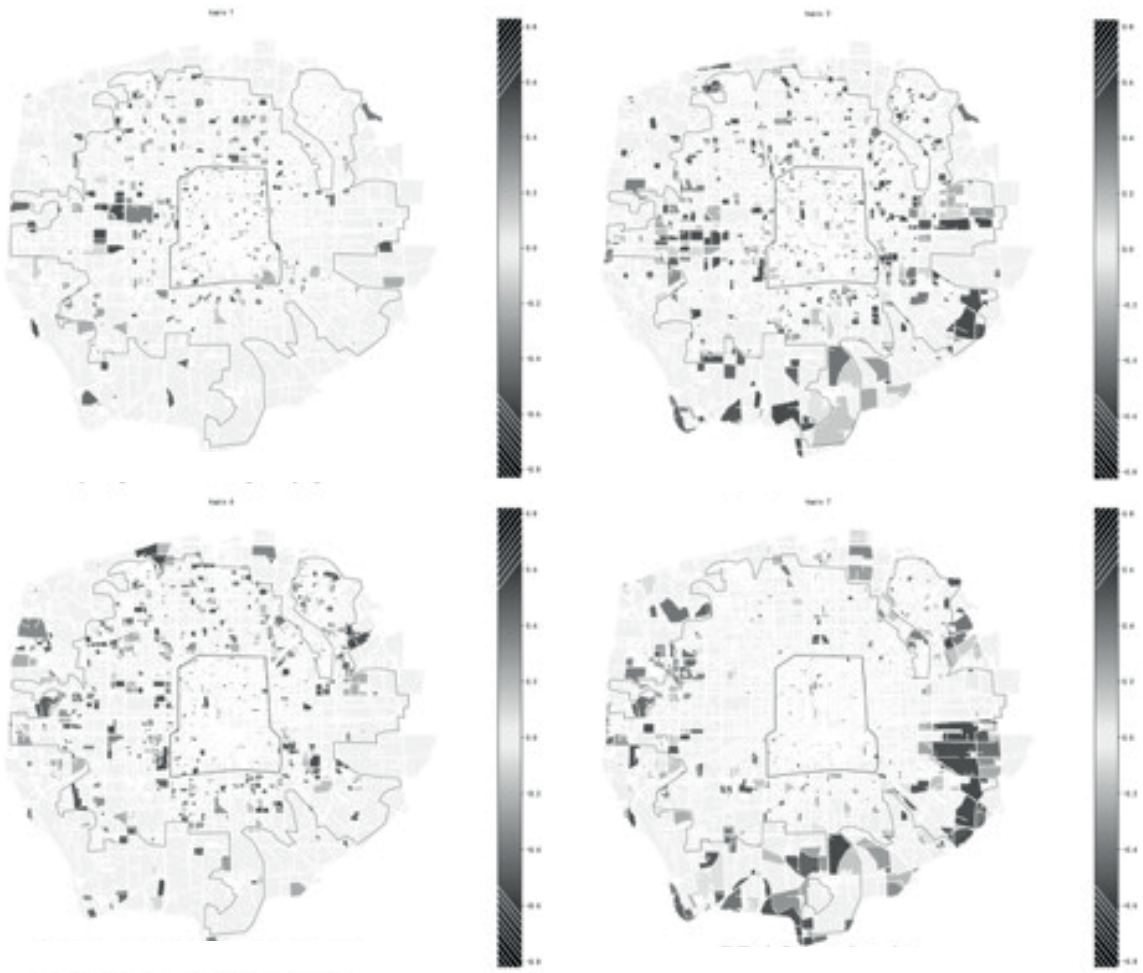

Figure 4. *The change of functions from 2014 to 2018 (a. Topic 1: government and regulatory agencies and etc.; b. Topic 3: personal service and etc.; c. Topic 6: residential accomodation and etc.; d. Topic 7: car service and etc.).*

4. Discussion and Conclusion

This study aims to quantify the morphological evolution in the two aspects of built environment and land use based on POI data and digital maps. Our study provides a possibility of combining the traditional Conzenian methods of morphological analysis with new data sources in the era of big urban data, as well as algorithms from the fields of data science, such as *k*-means, TF-IDF transformation, LDA and Shannon entropy.

However, there are still some aspects to be improved:

1. The data used have some errors. When studying the city as a whole, the increase in the amount of data can make up for the low precision. Analysis base on big data can be more objective in reflecting the characteristics of the city.
2. Spacematrix has shortcomings. Spacematrix is not effective with regard to mixing built types in one same block due to the hybridity of different indicators; the whole distribution of Spacematrix indicators is generally continuous, which requires subjective discrimination to classify into different clusters.
3. LDA algorithm is not stable. The result could vary according to different initial parameters. But through cross-validation and multiple tests, a relatively reasonable result can be achieved.
4. The balance between clarity of meaning and the abstraction. When iteratively processed, the information contained in the data is gradually integrated and abstracted. The key to its intelligibility is whether the comprehensive criteria are classified by human experience or by algorithms.

In conclusion, central Beijing can be classified into three different regions: old city, newly-built area and fringe belt. Three regions vary in characteristics of built environment and hybridity. The distribution of FSI basically forms a ring structure, low in the old city, high in the newly built area, and low in the fringe belt. Profitable land uses tend to agglomerate into centers, while public and welfare urban functions tend to disperse uniformly across the entire city. Some specific types of land use are limited to particular regions out of natural or historical. The built environment and land use are correlated, and certain types of land use tend to locate in blocks with certain built environment characteristics. As for hybridity, the old city region shows a high degree of mixing function and occupies convenient locations, while the newly built area is relatively homogeneous. The degree of hybridity in the fringe belt is higher because of the rural-related POI.

The changes in land use reveal the effect of the government adjusting the distribution of urban function through administrative methods from 2014 to 2018. Some related functions are relocating such as government affairs, residential services, and automobile services. The action optimizes the distribution of industrial and enhances innovation. However, this compulsory function relocation has led to homogenization in the old city of Beijing, leading to a possible reduction in urban vitality and convenience.

References

Berghauser Pont, M., & Haupt, P. (2007). The Spacemate: Density and the typomorphology of the urban fabric. *Urbanism laboratory for cities and regions: progress of research issues in urbanism.*

Blei D.M. , Ng A.Y., Jordan M.I. , Lafferty J. (2003), *Latent Dirichlet Allocation*, in b*Journal of Machine Learning Research, 3*, pp. 993-1022.

Chen W.S., Liu L., Liang Y.T. (2016), [Retail center recognition and spatial aggregating feature analysis of retail formats in Guangzhou based on POI data], in *Geographical Research,35*(4), (in Chinese).

Chen Y.S., Qi X. (2018), [Hybrid Piece of City: The Evolution of Urban Form in Three Historical Patterns in Beijing], in *Community Design, 088*(06), pp. 115-122.,(in Chinese).

Conzen M.R.G. (1968), *The use of town plans in the study of urban history*, in Dyos H.J. (ed.), *The study of urban history,* Edward Arnold, London, pp. 113-30.

Fang C.L. (2019), Basic rules and key paths for high-quality development of the new urbanization in China, in *Geographical Research,38*(01), pp. 15-24, (in Chinese).

Gil J., Montenegro N.C., Beirão J.N., Duarte J.P. (2009), *On the Discovery of Urban Typologies: Data Mining the Multi-Dimensional Character of Neighbourhoods*.

Jost L. (2006), *Entropy and diversity*, *Oikos.*

Liu X., Long Y. (2016), *Automated identification and characterization of parcels with Open Street Map and points of interest*, in *Environment and Planning B: Planning and Design*, *43*(2), pp. 341-360.

Pan Y.H. (2018), [The ternaryization of the world and a new generation of artificial intelligence], in *Modern City*, (in Chinese).

Van den Hoek J.W. (2008), *The MXI (Mixed-use Index) as tool for urban planning and analysis*, in *Corporations and Cities: Envisioning Corporate Real Estate in the Urban Future, Brussels, Belgium.*

Wang F., Gao X.L., Xu Z.N. (2015), [Identification and classification of urban commercial districts at block scale —— Take Beijing as an example], in *Geographical Research,34*(6), pp. 1125-1134, (in Chinese).

Wegener M., Gnad F., Vannahme M. (1986), *The time scale of urban change. Advances in Urban Systems Modelling.*

Witten I.H., Frank E. (2005), *Data mining: practical machine learning tools and techniques,* Morgan Kaufmann, San Francisco, 2nd edn.

Wu H.C. , Luk R.W.P. , Wong K.F. , Kwok K.L. (2008), *Interpreting tf-idf term weights as making relevance decisions*, in *ACM TRANSACTIONS ON INFORMATION SYSTEMS, 26*(3), pp. 1-37.

Yang T. (2018), [A study on Beijing spatial structure and functional location based on big data], in *City Planning Review,42*(09), pp. 30-40, (in Chinese).

Ye Y., Van Nes A. (2014), Quantitative tools in urban morphology: Combining space syntax, spacematrix and mixed-use index in a GIS framework, in *Urban morphology*, *18*(2), pp. 97-118.