

Article

# The Identification and Use Efficiency Evaluation of Urban Industrial Land Based on Multi-Source Data

Lin Qiao<sup>1,2</sup>, Huiping Huang<sup>1,2,\*</sup> and Yichen Tian<sup>1,\*</sup>

<sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; qiaolin@radi.ac.cn

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: huanghp@radi.ac.cn (H.H.); tianyc@radi.ac.cn (Y.T.)

Received: 25 August 2019; Accepted: 28 October 2019; Published: 4 November 2019



**Abstract:** Urban industrial land (UIL), which is classified for industrial purposes, is an indispensable component of urban land. Obtaining up-to-date and timely UIL details from the industrial development perspective has practical significance for UIL planning. Therefore, we propose a practical method for integrating UIL identification and use efficiency evaluation at the parcel scale based on multi-source data. The Open Street Map (OSM) data were utilized to generate parcels, which served as basic analytical units. Point of Interest (POI) data combined with a Continuous Bag-of-Words (CBoW)-based Word2Vec model was utilized to acquire UIL information. The entropy-weight Technique for Order Preference by Similarity to Ideal Solution method, combined with economic and environmental UIL indicators obtained from remote sensing images, ground observation data, and statistical data, was used to evaluate UIL use efficiency, and the spatial distribution and utilization degree of UIL within Beijing's fifth ring road was analyzed. The region within Beijing's fifth ring road was classified into commercial land, industrial land, and other types, with an overall accuracy of 92.24%. With this method, we found that the distribution of UIL presented a ring structure developing outwards along the ring roads and contained concentrated commercial areas. UIL utilization exhibited a south–north differentiation, and industrial land had lower use efficiency. Our work fully utilized the available fine-scale multi-source data.

**Keywords:** urban industrial land (UIL); use efficiency evaluation; multi-source data

## 1. Introduction

China has been experiencing urban expansion at an unprecedented scale and rate, which can be attributed to rapid urbanization [1]. Consequently, several problems such as urban population expansion, unreasonable economic and industrial structures, and the inefficient utilization of land resources are becoming more prominent. Urban land use reflects human activities as well as their use of land [2], and it is closely related to urban planning, economic development, and environmental protection [3–5]. In China, urban industrial land (UIL) can be interpreted as urban construction land that is classified in terms of an industrial purpose, and it usually refers to secondary and tertiary industrial land in the city [6,7]. With rapid economic development and urban expansion, the contradiction between the increasing demand of UIL and limited land resources has been gradually intensifying. Therefore, it is important to acquire up-to-date UIL information for rational planning.

Land Use Land Cover mapping usually relies on manual interpretations of aerial photos or satellite images, combined with field surveys and auxiliary materials (e.g., statistical data) [8]. Urban environments can change rapidly in large city regions when accompanied by rapid urbanization, exceeding the rate by which on-and-off efforts are undertaken to update existing land use database; thus, timely updatable urban land use information is very important [2,9]. Many researchers have

utilized object-oriented classification from satellite-based remote sensing images for obtaining urban land use types, focusing on the utilization of physical properties (e.g., spectrum, texture, and shape information) of ground objects derived from remote sensing images [10,11]; however, the knowledge regarding socioeconomic functions has not been fully utilized therein [12–14]. Some land use types such as commercial land and industrial land cannot be distinguished by simply using remote sensing information and its derived features because urban land use types are strongly correlated with internal socioeconomic activities [13].

Open social data that reflects socioeconomic activities is a particularly promising data resource, as it can capture spatiotemporal patterns of human activities and thus uncover the association with different land use types [12]. Point of Interest (POI) data therein possess rich semantic and location information. A POI record documents a location point and consists of attributes such as name, category (multilevel), and coordinates [15]. Moreover, POIs can denote buildings, stores, universities, or geographical entities of a certain size, and this level of detail strengthens the ability of the data to describe the entity's location. POIs have been utilized to estimate fine-resolution population [16], explore commercial pattern changes [17], and classify urban green space from a social function perspective [18]. Textual information from POIs has been mined using the topic model to infer land cover types [19]. However, this study considered only the frequency and not the spatial distribution of POIs. Yao et al. [14] considered the relationship between the spatial distribution information of POIs and regional functions. Based on these studies, the Word2Vec model can be used to quantify the relationship between POIs and UIL types [14]. The Word2Vec model can map words into high-dimensional vector spaces via their contextual relationships [20].

Researchers have substantially studied urban land use efficiency in China [21], with a focus on urban agglomerations [22] or on specific cities [23] as well as land use types such as construction land [24] or industrial land [25]. Efficiency is a crucial criterion for measuring urban land use level, and urban land use efficiency can be understood as the well-being produced on per unit urban land [23]. Researchers have developed specific indicators to measure land use efficiency in China [26], which covers economic (e.g., the total assets investment) and environmental aspects (e.g., air quality and green rate) [23,27]. Based on this, in our study, UIL use efficiency was used to measure the degree of UIL utilization by accounting for economic and environmental factors: the higher the UIL use efficiency, the better the UIL value realization. We considered utilizing accessible remote sensing data, ground observation data, and necessary statistical data for evaluating UIL use efficiency using the entropy-weight Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method. TOPSIS can generate reasonable evaluations of different problems with multiple criteria [28]. It has been applied in urban planning management studies such as hazard assessment and sustainable and ecological development evaluation [29,30].

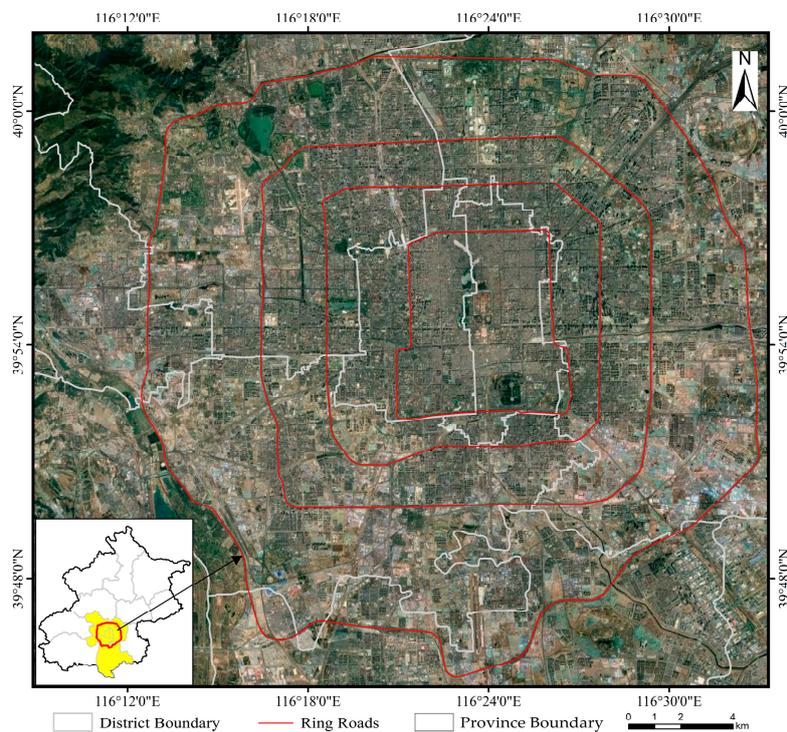
Considering the abovementioned issues, our study focuses on two objectives: first, we utilized timely and updatable POI data and a Word2Vec model to extract UIL types; next, we evaluated UIL use efficiency using the entropy-weight TOPSIS method by generating evaluation indicators from remote sensing data, ground observation data, and statistical data. This paper is structured as follows: the study area, data sources, and implementation methods are described in Sections 2 and 3 and the results, discussion, and conclusions are in Sections 4–6.

## 2. Study Area and Dataset

### 2.1. Study Area

Beijing has a land area of 16,808 km<sup>2</sup> and is the capital of China [31]. It is a typical city experiencing rapid urbanization: the spatial layout of the UIL therein has been undergoing changes due to the change in economic and spatial structures in urban areas. The proportion of tertiary industry has increased, and traditional enterprises such as factories have moved outwards since the implementation of the Beijing Urban Master Plan (2004–2020). Moreover, ring roads gradually formed alongside

rapid urbanization, and they can be seen as the most distinct feature of Beijing's urban structure [18]. The construction of ring roads reflects the characteristics of Beijing's urban development at various stages: for instance, the second ring road was built after demolishing the ring road of the old city wall; the fourth ring road was formed alongside the construction of Zhongguancun high-tech district and Olympic Park Commercial Center. The area within Beijing's fifth ring road is 668.65 km<sup>2</sup> [32], with permanent residents accounting for 48.9% of the population (2014 Beijing population sample survey results). This region has many commercial areas and high-tech industrial bases, and it contains densely distributed industries, including finance, computer services and software, real estate, scientific research and technical services, wholesale and retail services, accommodation and catering, and sports and entertainment. Meanwhile, the development of industry caused a high demand for employees and increased the consumption of resources and energy. Considering the abovementioned reasons, this region was selected as the study area (Figure 1).



**Figure 1.** The study region within Beijing's fifth ring road overlaid on a satellite image and the position of district boundaries and ring roads.

## 2.2. Data Collection and Preprocessing

Open Street Map (OSM) data were used to generate basic analytical units, and POIs were used to identify UIL. GF-1 remote sensing data, ground observation data, and statistical data were used to extract UIL use efficiency indicators.

### 2.2.1. Open Street Map (OSM) Road Network

Beijing's road network data were acquired from OSM (<https://www.openstreetmap.org>), which is a freely editable world map. As an online electronic map, OSM has high positional accuracy and comprehensive content that can be freely downloaded according to individual needs [33]. Data were classified by their relative importance and stored in vector format, which is convenient for the operation.

### 2.2.2. Point of Interest (POI) Data

In our study, POIs were collected using the web services application programming interface (API) of Amap (<http://www.amap.com/>), classified by a pre-defined three-level structure (e.g., for a shopping point, level-1: shopping; level-2: comprehensive market; level-3: seafood market). There were 612,419 POI records of the study region in 2013, covering 19 level-1 categories as follows: 28,187 transportation service facilities points; 9320 accommodation points; 12,901 sports and leisure service points; 79,971 corporation points; 13,974 health care service points; 26,220 business building points; 55,275 address information points; 213 motorcycle service points; 17,759 governmental agencies and public organizations points; 6016 automobile service points; 2425 car maintenance points; 795 car sales points; 73,882 life service points; 30,129 science and education points; 151,761 shopping points; 4828 traffic facilities points; 14,198 financial and insurance service points; 4196 scenic spot points; and 80,369 catering points.

### 2.2.3. GF-1 Remote Sensing Imagery

The multi-spectral sensor with 8 m resolution on GaoFen-1 (GF-1) includes blue, green, red, and near-infrared bands. A cloudless GF-1 multispectral image that covered the study area was obtained on 10 August 2013. The image was preprocessed for radiometric calibration, atmospheric correction, and orthorectification using ENVI 5.3.

The preprocessed GF-1 image was used to extract urban green space with the urban vegetation extraction methodology [34]. The study found that vegetation pixels have good linear separation from non-vegetation pixels with high Normalized Difference Vegetation Index (NDVI) values ( $NDVI > 0.25$ ) in true color composite reflectance space; and they can be separated from low-NDVI-value non-vegetation pixels ( $NDVI \leq 0.25$ ) in false color composite reflectance space. Thus, in this study, hyperplane Equations (1) and (2) were obtained by selecting vegetation and non-vegetation training samples of the study area using the linear kernel function of Support Vector Machines. A pixel can be classified as vegetation if Equations (1) and (2) are positive; if one of them is negative or both negative, the pixel is non-vegetation.

$$Plane_{false} : 3.21 \rho_{green} + 4.24\rho_{red} - 3.47\rho_{nir} + 0.07 = 0 \quad (1)$$

$$Plane_{true} : 7.46 \rho_{blue} + 11.13\rho_{green} - 20.35\rho_{red} + 0.09 = 0 \quad (2)$$

where  $Plane_{false}$  and  $Plane_{true}$  represent false and true color composite space separation planes, respectively;  $\rho_{blue}$ ,  $\rho_{green}$ ,  $\rho_{red}$ , and  $\rho_{nir}$  represent the reflectance of blue, green, red, and near-infrared bands respectively.

### 2.2.4. Ground Observation Data

Ground observation data of 2013 were derived from the Beijing Municipal Environmental Monitoring Center (<http://www.bjmemc.com.cn/>), which provides hourly Air Quality Index (AQI) data collected from 35 ground air monitoring stations in Beijing. The AQI is calculated based on six pollutants following the Chinese Technical Regulation on Ambient Air Quality Index (HJ 633–2012). The higher the value of AQI, the worse the air quality: an AQI value below 50 is generally considered satisfactory air quality.

### 2.2.5. Statistical Data

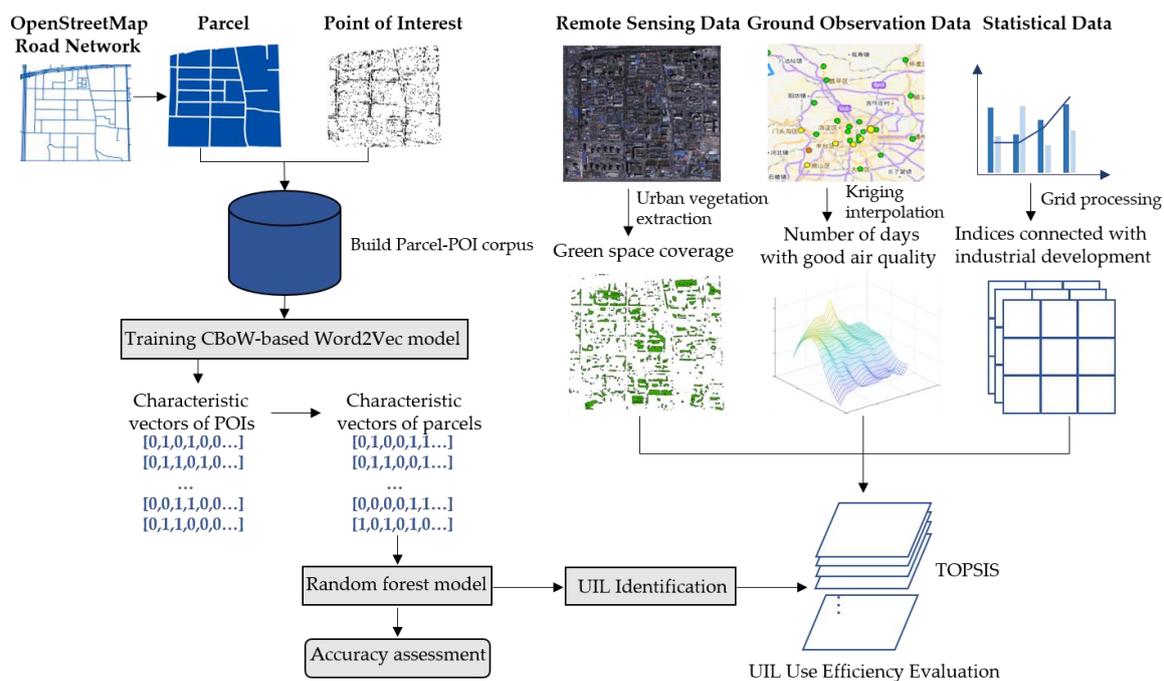
Statistical data were obtained from the Third National Economic Census of Beijing. The data contain detailed information regarding industrial development, such as number of employees, total profits, and total assets, and they utilize street as statistical units.

### 3. Methodology

The identification and use efficiency evaluation of UIL was conducted using the following methodological framework (Figure 2).

We first generated parcels using the OSM road network (including the deletion, trimming, and buffering of roads and removal of road space) based on the method of Long and Liu [35] and utilized them as basic analytical units. Parcels were relatively homogeneous in terms of land use functions [36] and were polygons bounded by road networks [9]. The corpus (i.e., a large and organized collection of well-sampled and processed texts in the field of natural language processing [37]) was constructed based on Parcels and POIs. POI vectors were extracted using a Continuous Bag-of-Words (CBoW)-based Word2Vec model, and parcel vectors were calculated by performing the averaged summation of POI vectors within parcels. The random forest model was adopted to extract UIL types of parcels based on parcel vectors.

Subsequently, at the parcel level, remote sensing data, ground observation data, and statistical data were utilized to extract the following: the green space coverage index; the index of number of days with good air quality; and indices connected with industrial development, including total assets, total profits, number of employees, energy consumption, and net demand of water. The entropy-weight TOPSIS method was employed to evaluate UIL parcel use efficiency based on these indicators.



**Figure 2.** The methodological framework of the identification and use efficiency evaluation of urban industrial land (UIL).

#### 3.1. Classification System of UIL

UIL types in this study were classified according to the Chinese urban land use classification criteria (GB 50137–2011), as shown in Table 1. Considering there are a few warehouses within Beijing's fifth ring road, we classified warehousing into industrial land. UIL contained two primary categories, commercial land and industrial land, and commercial land consisted of commercial services, accommodation, entertainment, and business.

**Table 1.** Classification system of UIL.

	Primary Class	Secondary Class	Description
UIL	Commercial land	Commercial services	Shopping malls, supermarkets, markets, retail stores (e.g., drugstores, convenience stores, food sales, exclusive stores, furniture, and building materials stores), service facilities (e.g., maintenance services, hairdressers, and household services), etc.
		Accommodation	Hotels, inns, resorts, etc.
		Entertainment	Recreational facilities (e.g., amusement parks and golf courses)
		Business	Office buildings, financial services, etc.
	Industrial land		Equipment sales, manufacturing industrial, and warehouses

### 3.2. Construction of the UIL Identification Model

UIL types are difficult to acquire solely using remote sensing images: UIL types are related to socioeconomic activities and human use; however, socioeconomic knowledge extracted from remote sensing images is limited [12,13]. In this study, we mined semantic information and spatial correlations of POI data inside cities, quantified POIs into high-dimensional vectors by utilizing CBoW-based Word2Vec model, and applied POI vectors to determine UIL types at the parcel scale.

#### 3.2.1. Parcel Generation

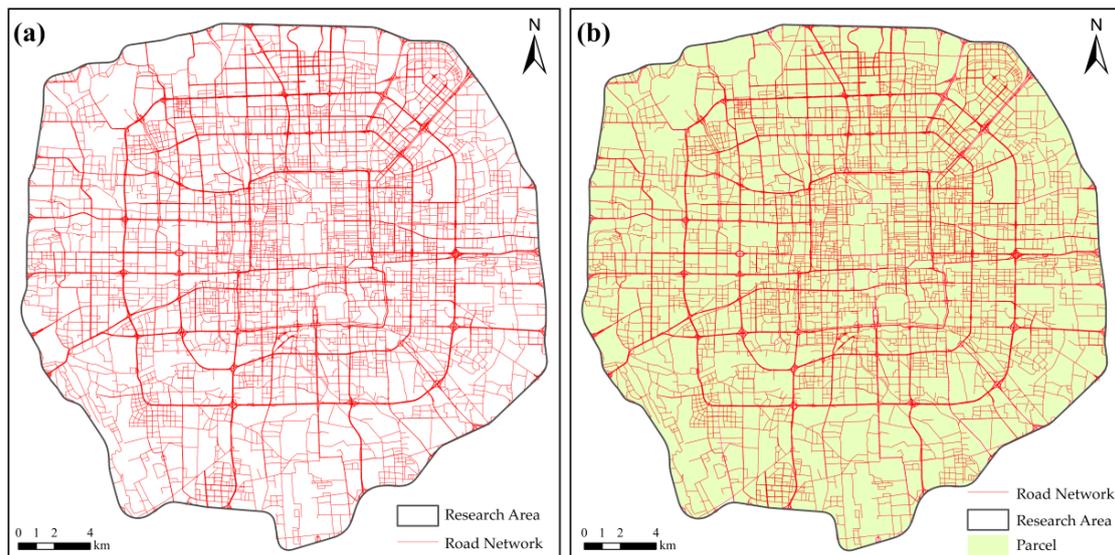
Long and Liu [35] suggested that parcels served as natural segmentation boundaries within urban areas. In this study, the OSM road network was utilized to generate parcels based on this assumption. First, the OSM roads were preprocessed; redundant roads were deleted (e.g., roads within residential districts, university districts); and dangling roads were trimmed. Based on the importance of their original levels (<http://wiki.openstreetmap.org/wiki/Key:highway>) [8], we reclassified the OSM roads into three levels, wherein Level 1 includes trunk, primary, and motorway roads; Level 2 includes secondary roads; and Level 3 includes tertiary, unclassified, residential, and service roads. Next, according to the Chinese road construction standards and the actual road widths within Beijing's fifth ring road, road spaces were generated by creating 40, 20, and 10 m buffer widths [8]. Finally, less meaningful parcels were removed: those with an area below 0.005 km<sup>2</sup> were removed because they lacked actual socioeconomic functions or were located inside overpasses or roundabouts. In this manner, 3005 parcels were obtained (Figure 3).

#### 3.2.2. Building Parcel-POI Corpus and Obtaining Vectors of POI Categories

In the field of natural language processing (NLP), a corpus usually refers to a large and organized collection of well-sampled and processed texts [37]. To be specific, a corpus includes plenty of documents, and each document contains several words. In a corpus, the sequential order of documents and the words therein represents the contextual relationships [14]. Applying these concepts to UIL identification analysis, the study area, each parcel therein, and POIs within each parcel were considered as the corpus, the document therein, and words within each document, respectively. Level-3 categories of POIs were used as descriptions for composing documents in order to guarantee sufficient and effective words.

In our study, the Parcel-POI corpus was constructed using a shortest path method proposed by Yao et al. [14]. Based on this method, POIs in parcels were connected to each other through their spatial relations. Before building the Parcel-POI corpus, we first selected and obtained the POIs within each parcel according to the parcel's location. The process of constructing a Parcel-POI corpus was as

follows. First, for each parcel, the shortest path passing all POIs was calculated and the sequential order of POIs was recorded. Next, parcel-based documents were constructed using words (i.e., level-3 categories of POI) in the same order as that of the POI's acquisition above. Similarly, the sequential order of parcel-based documents in the corpus was recorded by calculating the shortest path passing all parcels. Thus, the Parcel-POI corpus was constructed. Through the organization of the corpus, the spatial distribution attributes and location relationships of the POIs were revealed by the contextual relationships of words (i.e., level-3 categories of POI) [14].



**Figure 3.** (a) The Open Street Map (OSM) data of the study region after preprocessing; (b) parcels inside the study region.

The CBoW model in the Word2Vec tool was selected to train vectors of POI categories in our study. Word2Vec launched by Google is an open source tool for word embedding in the natural language process [20]. By building a neural network language model on a given corpus, the Word2Vec model can map words into high-dimensional vectors via its contextual content [20].

### 3.2.3. Extracting UIL Types for Each Parcel

Parcel vectors were calculated by the averaged summation of all POI vectors inside a parcel based on the study of Zhuang et al. [38]. Assuming  $M$  POIs ( $P_{i,1}, P_{i,2}, P_{i,3}, \dots, P_{i,m}$ ) in the  $i^{th}$  parcel, parcel vectors can be expressed as follows:

$$vector\_Parcel_i = \frac{\sum_{k=1}^M vector\_POI_{category}(P_{i,k})}{M}, \quad (3)$$

where  $vector\_POI_{category}(P_{i,k})$  represents the vector of the  $k^{th}$  POI category in the  $i^{th}$  parcel, and  $vector\_Parcel_i$  represents the vector of the  $i^{th}$  parcel.

The actual UIL types of parcels participated in analysis were identified according to the online digital map, street view data, and field surveys. Considering the existence of multiple types within some parcels, the main UIL type of these mixed parcels was used. Half of the total parcels participated in analysis were randomly selected as training samples, and the remaining half were used as testing samples. The training samples were utilized to train model parameters, in which parcel vectors were considered as the input and their actual UIL types as labels. The trained model was then used for precision assessment of testing samples.

In our study, the random forest model [39,40] was selected to extract UIL types for parcels. The Out-Of-Bag (OOB) [8] accuracy, which is an unbiased estimation of generalization error, was utilized

to evaluate the prediction precision of the random forest model. The classification process was repeated 50 times and the overall accuracy was considered as the evaluation criteria.

### 3.3. UIL Use Efficiency Evaluation

Economic and environmental benefits were accounted for during the selection of urban land use efficiency indicators [27]. Referring to previous studies on indicator selection and considering the availability of data [21,23], we selected total assets, total profits, and number of employees as economic benefits; environmental benefits included green space coverage, number of days with good air quality, energy consumption, and net demand of water. After obtaining the abovementioned indicators of UIL parcels, entropy-weight TOPSIS was utilized to evaluate use efficiency of each UIL parcel.

#### 3.3.1. Use Efficiency Indicators Acquisition

In our study, green space coverage for each UIL parcel refers to the proportion of vegetation within the parcel, as shown in Equation (4).

$$GSC = \frac{S_{veg}}{S_{parcel}} \times 100, \quad (4)$$

where GSC represents green space coverage,  $S_{veg}$  is the vegetation area within the parcel, and  $S_{parcel}$  is the parcel area.

The number of days with good air quality for UIL parcels was calculated as follows. According to the distribution of 35 monitoring stations, 366 10 m gridded surfaces of daily AQI estimations were created using Kriging interpolation, including the gridded surfaces of the standard error to describe the prediction error. The daily AQI of UIL parcels was obtained through spatial averaging of all the gridded daily AQI values inside the parcels. The number of days with good air quality for UIL parcels was obtained by counting days in a year with AQI below 50.

Indices connected with industrial development for UIL parcels, including total profits, total assets, number of employees, energy consumption, and net demand of water, were calculated as follows. Statistical data values of street level were assigned to street units, and 200 m × 200 m gridded data were created within street units' boundaries covering the study area. The value of each grid was determined using the areal weighting method. Each index abovementioned of UIL parcels was calculated by summation of all gridded values of this index inside the parcels.

#### 3.3.2. Use Efficiency Evaluation Method

Entropy-weight TOPSIS, as a multi-criteria decision analysis technique, was adopted to calculate UIL use efficiency scores. This model introduced the entropy-weight method [41] into TOPSIS: first, the entropy-weight method utilized information entropy to determine the weight of indicators, wherein the smaller the information entropy, the greater the weight of indicators [42]; then, TOPSIS was utilized to calculate the degree of proximity between the positive and negative ideal solutions and evaluated objects and accordingly determine comprehensive evaluation scores [28,43]. The process of calculating UIL use efficiency scores was as follows:

The first step was data standardization.

Consider the indicator matrix  $X = (x_{ij})_{m \times n}$ , where  $m$  denotes the number of UIL parcels,  $n$  denotes the number of indexes, and  $x_{ij}$  denotes the value of index  $j$  of the  $i^{\text{th}}$  UIL parcel ( $i = \{1, 2, \dots, m\}$ ,  $j = \{1, 2, \dots, n\}$ ).

Applying vector normalization to convert the original index value  $x_{ij}$  into normalized index value  $r_{ij}$ :

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^m x_{kj}^2}}. \quad (5)$$

The second step was index weight determination by the entropy-weight method. We calculated the proportion of index  $j$  in the  $i^{\text{th}}$  UIL parcel.

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}}. \quad (6)$$

Determining the information entropy value of index  $j$ :

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m p_{ij} \ln p_{ij}. \quad (7)$$

We calculated the weight value of index  $j$ .

$$w_j = \frac{1 - e_j}{\sum_{j=1}^n (1 - e_j)}. \quad (8)$$

The third step was the determination of UIL evaluation scores by TOPSIS.

We determined a set of weights  $\{w_1, w_2, \dots, w_n\}$  for  $n$  indexes and calculated the weighted normalized indicator matrix ( $Z$ ).

$$Z = (z_{ij})_{m \times n}, \quad z_{ij} = w_j \times r_{ij}. \quad (9)$$

We then calculated the positive and negative ideal solutions.

$$\begin{cases} Z^+ = \{ \max_{1 \leq i \leq m} z_{ij} | j = 1, 2, \dots, n \} \\ Z^- = \{ \min_{1 \leq i \leq m} z_{ij} | j = 1, 2, \dots, n \} \end{cases}. \quad (10)$$

Then, we determined  $sep_i^+$  and  $sep_i^-$  by calculating the Euclidean distances between each UIL parcel's evaluated value and the positive and negative ideal solutions.

$$sep_i^+ = \sqrt{\sum_{j=1}^n (z_{ij} - z_j^+)^2}, \quad (11)$$

$$sep_i^- = \sqrt{\sum_{j=1}^n (z_{ij} - z_j^-)^2}. \quad (12)$$

Finally, we calculated the evaluation score ( $C_i$ ) for each UIL parcel; a higher value of  $C_i$  reflected a greater the utilization degree of UIL parcel, ranging from 0 to 1.

$$C_i = \frac{sep_i^-}{sep_i^+ + sep_i^-}. \quad (13)$$

## 4. Analysis and Results

### 4.1. Identification Results and Spatial Distribution of UIL

The study area included a total of 613,973 POIs, which were distributed into 2888 parcels. Due to lack of POIs, 117 parcels were removed and excluded from the analysis. A total of 1548 level-3 categories of POIs were included in building the Parcel-POI corpus. When using the CBoW model to construct word vectors, the sampling window size, output vector dimensions, and number of iterations were set to 5, 200, and 20, respectively. The default settings were maintained for other parameters. The spatial distributions of POIs were represented by the contextual relationships of attribute categories via the conversion of the CBoW model. Thus, 1548 characteristic vectors of level-3 categories of POI were obtained.

Characteristic vectors of 1444 parcels were used as the training samples to construct the random forest model. The characteristic vectors of the same number of parcels were used as the testing samples to evaluate the UIL identification precision of the random forest model. The OOB and testing precisions were 83.19% and 92.24%. The confusion matrix of identifying UIL types using the testing samples was shown in Table 2.

**Table 2.** Confusion matrix of identifying UIL types using the testing samples.

	Commercial	Industrial	Other Types	Total	User's Accuracy
Commercial	235	4	46	285	82.46%
Industrial	3	12	5	20	60.00%
Other types	51	3	1085	1139	95.26%
Total	289	19	1136	1444	
Producer's Accuracy	81.31%	63.16%	95.51%		92.24%

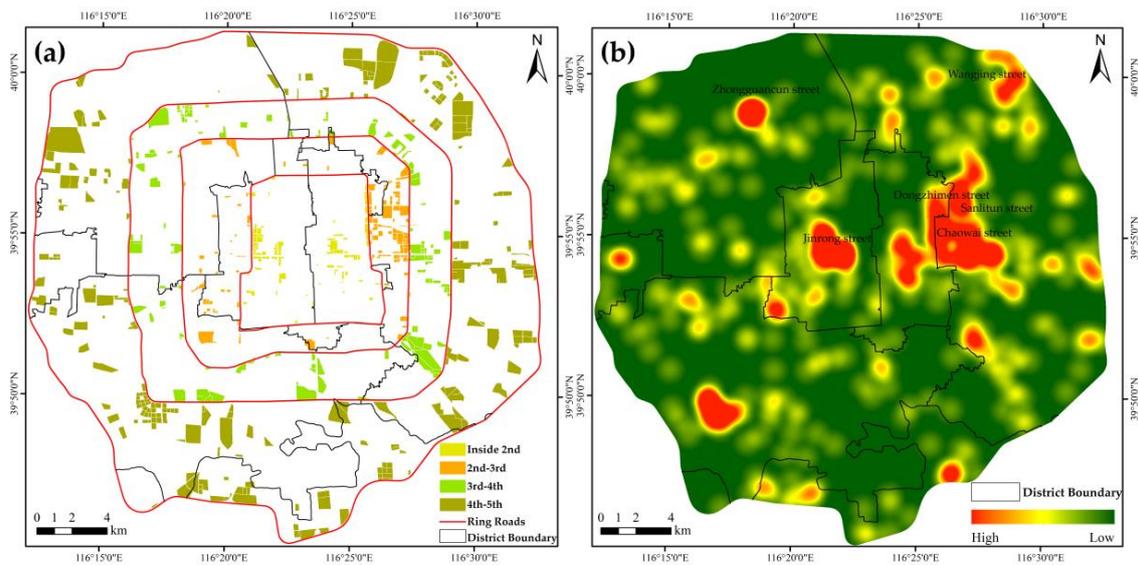
Other types were well differentiated from commercial and industrial parcels because the producer's and user's accuracies were more than 95%. Some of the other types were misclassified as commercial parcels because they contained many POIs that were involved in commercial content; however, the main function of these parcels was not commercial. Commercial parcels were classified slightly better than industrial parcels. Misclassified commercial parcels were principally the large parcels, which were likely to contain other types.

There was a total of 602 UIL parcels within the study region (Table 3). The region from the 4th to 5th ring roads possessed the largest number of commercial parcels, which possibly contributed to reducing congestion in the core area. The distribution of UIL presented a ring structure, developing outwards along the ring roads (Figure 4a), and contained concentrated commercial areas, such as commercial shopping centers (e.g., Xidan, Wangfujing, and Zhongguancun commercial districts) and business office buildings (e.g., Jinrong Street and Guomao business district) (Figure 4b). From the street level, UIL parcels were mainly distributed in Sanlitun Street, Chaowai Street, Dongzhimen Street, Jinrong Street, Zhongguancun Street, and Wangjing Street.

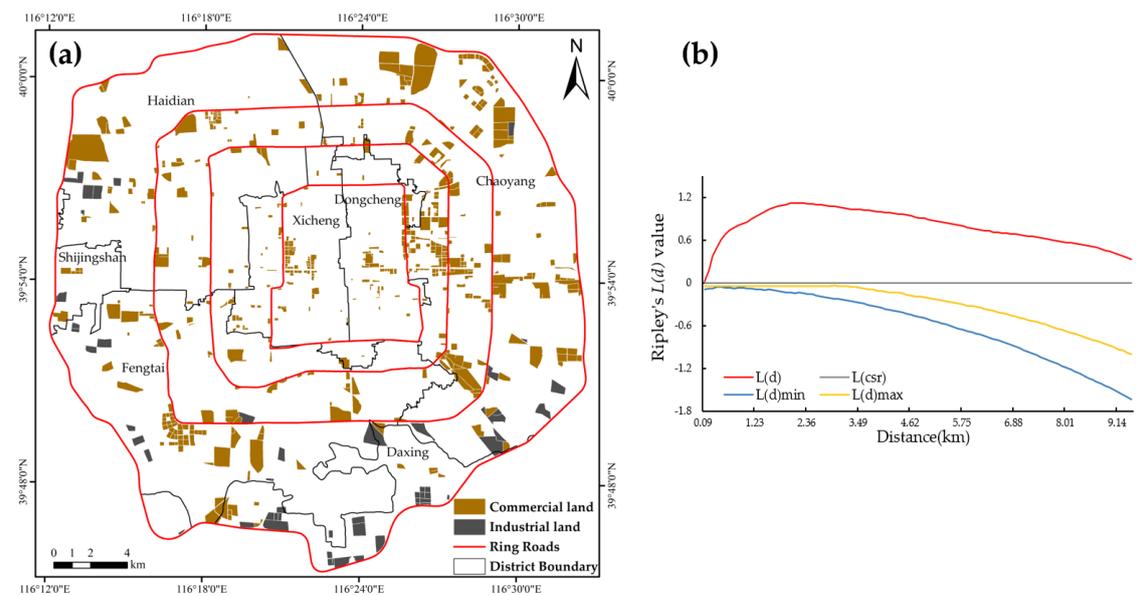
**Table 3.** Count of different UIL types among each pair of ring roads.

Count	Within Second	From Second to Third	From Third to Fourth	From Fourth to Fifth
Commercial land	97	127	133	193
Industrial land	0	0	1	51
total	97	127	134	244

Commercial parcels in Xicheng and Dongcheng districts were mainly used for commercial services (shopping centers) and businesses (finance services). In Haidian, Fengtai, and Chaoyang districts, some were also used for businesses (information services) and entertainment. Industrial parcels were mainly concentrated in the western and southern areas between the 4th and 5th ring roads (Figure 5a), which was likely connected to land prices and the convenience of transportation conditions, as well as urban planning policies.



**Figure 4.** (a) Spatial distribution of UIL among each pair of ring roads inside the study region; (b) kernel density of UIL.



**Figure 5.** (a) Spatial distribution of commercial land and industrial land among each pair of ring roads inside the study region; (b) spatial agglomeration intensity map of UIL by calculating Ripley’s  $L(d)$  value. Positive, negative, and zero values of  $L(d)$  represented clustered, dispersed, and random distribution patterns, respectively.

Ripley’s  $L(d)$  value was used to measure the multi-distance spatial agglomeration pattern of UIL within Beijing’s fifth ring road. The  $L(d)$  value of UIL was greater than 0, indicating that UIL exhibited spatial agglomeration (Figure 5b). Meanwhile, as the distances increased from 0 to 9 km, the agglomeration intensity firstly increased and then slowly decreased.

4.2. Criteria, Evaluation Results, and Spatial Distribution of UIL Use Efficiency

Utilizing the methods in Section 3.3.1, seven criteria were calculated. The weights of seven criteria were determined by the entropy-weight method mentioned in Step 2 of Section 3.3.2, and four criteria (i.e., total assets, total profits, number of employees, and green space coverage) had more impact on UIL use efficiency.

There was no obvious distribution pattern of the green space coverage of UIL. In particular, the green space coverage of commercial parcels that were mainly used for entertainment was up to 90%, and commercial parcels mainly used for business had higher green space coverage than those used for accommodation or commercial services. This was related to UIL types and specific greening planning for construction projects.

On the whole, the number of days with good air quality was marginally higher in commercial parcels than in industrial parcels. Obvious north-south differences existed, and the value of the southern area was relatively lower, which was related to industry distribution and government controls of air quality.

There was a south–north trend in total assets, total profits, number of employees, energy consumption, and net demand of water: high values of parcels comprising these five indicators aggregated inside the 3rd ring road, which was connected to the highly concentrated financial services and business service industries.

The use efficiency evaluation scores ranged from 0.10 to 0.71 and were reclassified into four classes: I (0.1–0.2), II (0.2–0.3), III (0.3–0.4), and IV (above 0.4). Figure 6a presented the proportion of four class scores between each pair of ring roads. Use efficiency scores showed a north-south gradient, and scores inside the 3rd ring road were better than in other regions (Figure 6b). The evaluation scores of commercial parcels were clearly better than those of industrial parcels, and scores above 0.4 mostly appeared in commercial parcels whose main function was accommodation and business (e.g., commercial office buildings, financial services, and hotels).

Global spatial autocorrelation measures the overall correlation degree, spatial distribution pattern, and their significance of spatial objects in the region. The overall correlation of UIL utilization degree was obtained according to UIL use efficiency scores using global Moran’s *I*. The global Moran’s *I* value of 0.647 ( $p = 0.001$ ) passed the significance testing at 1% (Figure 7a), and the figure indicated that UIL utilization degree showed positive spatial correlation. The Local Indicators of Spatial Association (LISA) aggregation graph was drawn to describe local spatial heterogeneity characteristics. High–high regions contained 88 commercial parcels, and were located in areas roughly consistent with the concentrated commercial area (Figure 7b).

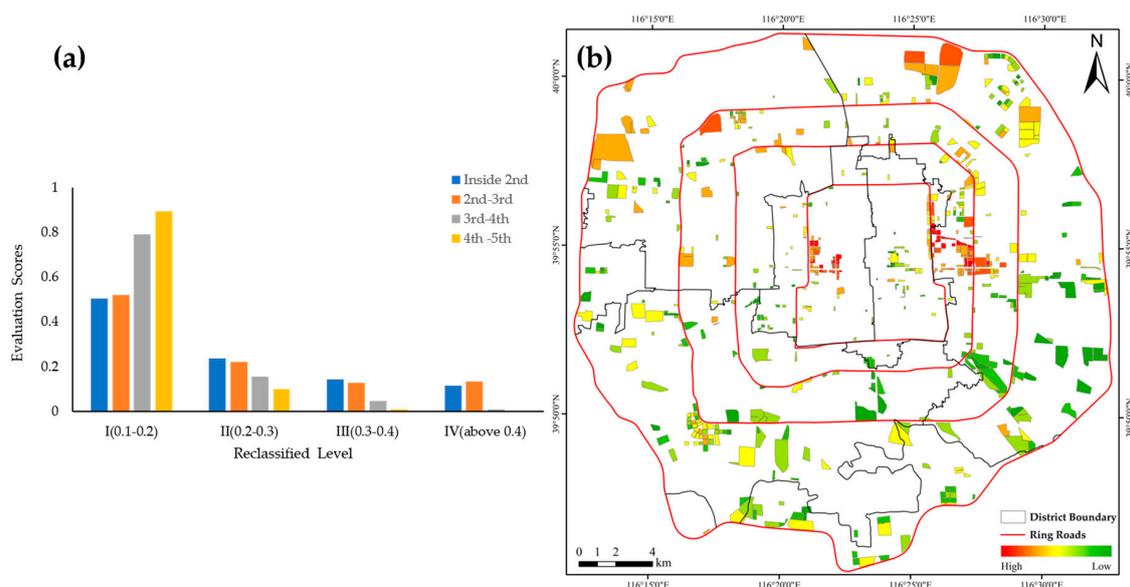
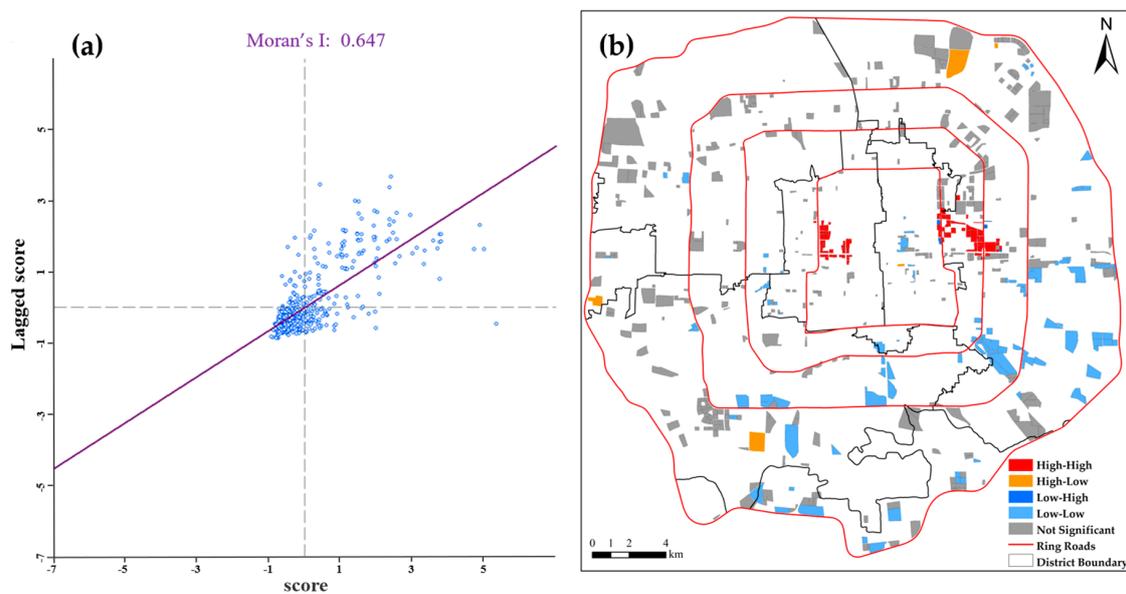


Figure 6. (a) Proportion of four class evaluation scores of UIL use efficiency between each pair of ring roads; (b) UIL use efficiency evaluation scores.



**Figure 7.** (a) Moran scatter plot of UIL use efficiency scores; (b) Local Indicators of Spatial Association (LISA) cluster distribution map of UIL use efficiency scores.

## 5. Discussion

### 5.1. Industrial Development in Beijing

Since 2004, the change in the industrial structure has accelerated in Beijing. The central city area principally consists of finance and business offices, and traditional manufacturing industries have been gradually moved outwards, with an aim to alleviate the current situation of the excessive agglomeration of industrial functions in urban central areas. Industries in key functional areas are more concentrated, such as the financial industry in Jinrong Street, business services in Guomao business district, and information services in Zhongguancun. This is consistent with the results of UIL identification with POIs. Certain problems in industrial land were reported, such as relatively scattered distribution, smaller surface areas, and lower degree of development and utilization. Our study reflects this and shows that industrial land is principally scattered between the 4th to 5th ring roads, and that their degree of utilization is low, as calculated by the entropy-weight TOPSIS method.

### 5.2. Implication of UIL Identification and Use Efficiency Analysis

Combined with easily accessible and timely updated POI data, the ongoing situation of UIL can be obtained and thus provide support for urban land use planning. UIL supports the social and economic activities of cities, and UIL identification can intuitively provide insight into the layout, distribution pattern, and density of UIL. The UIL layout is the comprehensive result of joint action between economy, technology, and policy factors, and the evolution of the UIL layout reflects changes in the industrial structure and economic development. The spatial distribution and combination of urban land can affect urban functional operation, residents' quality of life, and the level of urban development. As an important urban land use type, the rational arrangement of UIL is the key issue of urban development. For instance, the rational arrangement of UIL and residential land can contribute to the easing of urban commuting pressure and thereby reduce environmental pollution [44]. Meanwhile, the causes of the present situation of UIL can be analyzed by considering economic, social and environmental factors.

UIL use efficiency reflects the realization degree of UIL value in the process of urban economic development, and is related to factors of population size, economic level, traffic location, industrial structure, and input of production factors. By analyzing the spatial distribution of UIL use efficiency, measures can be taken to improve UIL utilization degree.

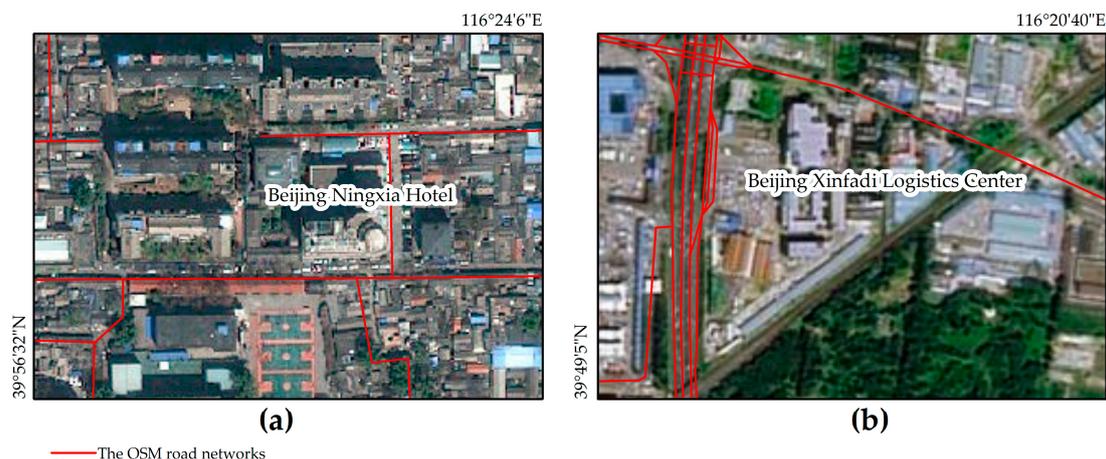
### 5.3. Incorporating Multi-Source Data

POI categories are cognized and conceptualized by individuals, which can help us better understand interactions between the urban environment and UIL types. Compared with traditional urban land use mapping methods or aerial photo interpretation, the utilization of POI data provides a new perspective for UIL identification. POIs used for constructing characteristic vectors are continually updated on the Internet, thus providing a quick method for acquiring the newest UIL types and enhancing the convenience of decision makers for understanding UIL layouts.

We constructed use efficiency evaluation indicators based on economic benefits and environmental impacts. Combined with multi-source data that included remote sensing data, ground observation data and statistical data, a better understanding of UIL use efficiency can be achieved in comparison to solely depending on statistical data. Similar to POI data, remote sensing data and ground observation data can also be updated on a regular basis. However, data related to industrial development are only available through official statistics at present. This suggests that up-to-date UIL use efficiency evaluation results can be maintained through a specific time-scale update.

### 5.4. Issues Associated with OSM-Segmented Parcels

There were some intrinsic problems for parcels generated by road networks because the quality of parcels depended on the accuracy of OSM. Parcels generated for the study region behaved well in the central area. However, they were not satisfactory in margin areas where the road networks were sparse. Thus, there were mixed UIL types distributed particularly between the 4th and 5th ring roads on account of the OSM's quality. Within some parcels, commercial buildings lacked clear boundaries with residential districts, university areas, or government offices, which impeded them from being correctly classified. For example, "Beijing Ningxia Hotel" shown in Figure 8a was located around residences, but no specific separation could be demarcated between the two. In addition, some large parcels were a mix of UIL types and other types, such as green spaces and residences, while few specific roads divided them into distinct units. For example, "Beijing Xinfadi Logistics Centers" were located in a large parcel mixed with other types without specific roads separating them (Figure 8b).



**Figure 8.** Issues with the quality of OSM data: (a) commercial buildings without specific boundaries; (b) a large parcel mixed with other types.

## 6. Conclusions

UIL is closely bound to urban land scales, and urban land sustainable use as well as coordinate development. Understanding the current situation of UIL can aid city planners to allocate urban land resources more efficiently and to alleviate the pressure induced by rapid urbanization. Therefore, the identification and use efficiency evaluation of UIL is necessary, and it has important practical significance for urban land planning. We adopted the strategy of combining regularly updated POI

data with a CBoW-based Word2Vec model used for word embedding to acquire UIL, and it was proven to be feasible. UIL inside Beijing's fifth ring road was distributed along ring roads, exhibiting spatial agglomeration. The entropy-weight TOPSIS method combined with UIL indicators related to the economy and environment obtained from multi-source data facilitated the reliable evaluation of UIL use efficiency at the parcel level. We found that the UIL utilization within Beijing's fifth ring road demonstrates a south-north differentiation and industrial land had a lower utilization degree. Although the use of POI data to identify UIL was proven to be feasible, the uncertainty and incompleteness of POI data collected from crowdsourcing should be accounted for in future applications. In the future, we consider integrating Amap POI data and Baidu POI data to enhance credibility and include more indicators such as spatial accessibility in UIL use efficiency evaluation.

**Author Contributions:** H.H. and Y.T. proposed the framework and edited the draft. L.Q. performed the experiments and wrote the paper. H.H. acquired the funding.

**Funding:** This study was supported by National Key R&D Program of China (2017YFB0503800).

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their systematic review and valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wei, Y.D.; Li, H.; Yue, W. Urban land expansion and regional inequality in transitional china. *Landsc. Urban Plan.* **2017**, *163*, 17–31. [[CrossRef](#)]
2. Hu, S.; Wang, L. Automated urban land-use classification with remote sensing. *Int. J. Remote Sens.* **2013**, *34*, 790–803. [[CrossRef](#)]
3. Mateos, A.C.; Amarillo, A.C.; Carreras, H.A.; González, C.M. Land use and air quality in urban environments: Human health risk assessment due to inhalation of airborne particles. *Environ. Res.* **2018**, *161*, 370–380. [[CrossRef](#)] [[PubMed](#)]
4. Liang, X.; Liu, X.; Li, D.; Zhao, H.; Chen, G. Urban growth simulation by incorporating planning policies into a ca-based future land-use simulation model. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 2294–2316. [[CrossRef](#)]
5. Zhao, Z.-Q.; He, B.-J.; Li, L.-G.; Wang, H.-B.; Darko, A. Profile and concentric zonal analysis of relationships between land use/land cover and land surface temperature: Case study of shenyang, china. *Energy Build.* **2017**, *155*, 282–295. [[CrossRef](#)]
6. Dan, C.; Xong, H. A study on the relationship between the land use structure and industrial development in hainan province. *Resour. Sci.* **2010**, *32*, 718–723. (In Chinese)
7. Lu, Z.; Zhang, J.; Li, X.; Wang, X.; Bai, X.; Zhang, J. Study on the evaluation of urban industrial land intensive use. *Appl. Mech. Mater.* **2012**, 226–228, 2442–2449. [[CrossRef](#)]
8. Zhang, Y.; Li, Q.; Huang, H.; Wu, W.; Du, X.; Wang, H. 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.* **2017**, *9*, 865. [[CrossRef](#)]
9. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping urban land use by using landsat images and open social data. *Remote Sens.* **2016**, *8*, 151. [[CrossRef](#)]
10. Wu, S.-S.; Qiu, X.; Usery, E.L.; Wang, L. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Ann. Assoc. Am. Geogr.* **2009**, *99*, 76–98. [[CrossRef](#)]
11. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [[CrossRef](#)]
12. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]
13. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.-L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [[CrossRef](#)]
14. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [[CrossRef](#)]
15. Liu, X.; Niu, N.; Liu, X.; Jin, H.; Ou, J.; Jiao, L.; Liu, Y. Characterizing mixed-use buildings based on multi-source big data. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 738–756.

16. Bakillah, M.; Liang, S.; Mobasheri, A.; Arsanjani, J.J.; Zipf, A. Fine-resolution population mapping using openstreetmap points-of-interest. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1940–1963. [[CrossRef](#)]
17. Chen, S.; Tao, H.; Li, X.; Zhuo, L. Detecting urban commercial patterns using a latent semantic information model: A case study of spatial-temporal evolution in guangzhou, china. *PLoS ONE* **2018**, *13*, e0202162. [[CrossRef](#)]
18. Chen, W.; Huang, H.; Dong, J.; Zhang, Y.; Tian, Y.; Yang, Z. Social functional mapping of urban green space using remote sensing and social sensing data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *146*, 436–452. [[CrossRef](#)]
19. Xing, H.; Meng, Y.; Hou, D.; Song, J.; Xu, H. Employing crowdsourced geographic information to classify land cover with spatial clustering and topic model. *Remote Sens.* **2017**, *9*, 602. [[CrossRef](#)]
20. Mikolov, T.; Chen, K.; Corrado, G.s.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
21. Cao, X.; Liu, Y.; Li, T.; Liao, W. Analysis of spatial pattern evolution and influencing factors of regional land use efficiency in china based on esda-gwr. *Sci. Rep.* **2019**, *9*, 520. [[CrossRef](#)] [[PubMed](#)]
22. Wu, C.; Wei, Y.D.; Huang, X.; Chen, B. Economic transition, spatial development and urban land use efficiency in the yangtze river delta, china. *Habitat Int.* **2017**, *63*, 67–78. [[CrossRef](#)]
23. Zhu, X.; Zhang, P.; Wei, Y.; Li, Y.; Zhao, H. Measuring the efficiency and driving factors of urban land use based on the dea method and the pls-sem model—A case study of 35 large and medium-sized cities in china. *Sustain. Cities Soc.* **2019**, *50*, 101646. [[CrossRef](#)]
24. Chen, Y.; Chen, Z.; Xu, G.; Tian, Z. Built-up land efficiency in urban china: Insights from the general land use plan (2006–2020). *Habitat Int.* **2016**, *51*, 31–38. [[CrossRef](#)]
25. Zhao, X.; Zhang, L.; Huang, X.; Zhao, Y.; Zhang, Y. Evolution of the spatiotemporal pattern of urban industrial land use efficiency in china. *Sustainability* **2018**, *10*, 2174. [[CrossRef](#)]
26. Yang, X.; Wu, Y.; Dang, H. Urban land use efficiency and coordination in china. *Sustainability* **2017**, *9*, 410. [[CrossRef](#)]
27. Wu, D.; Mao, H.; Zhang, X.; Huang, J. Assessment of urban land use efficiency in china. *Acta Geogr. Sin.* **2011**, *66*, 1111–1121.
28. Rashidi, K.; Cullinane, K. A comparison of fuzzy dea and fuzzy topsis in sustainable supplier selection: Implications for sourcing strategy. *Expert Syst. Appl.* **2019**, *121*, 266–281. [[CrossRef](#)]
29. Irankhahi, M.; Jozi, S.A.; Farshchi, P.; Shariat, S.M.; Liaghati, H. Combination of gisfm and topsis to evaluation of urban environment carrying capacity (case study: Shemiran city, iran). *Int. J. Environ. Sci. Technol.* **2017**, *14*, 1317–1332. [[CrossRef](#)]
30. Estay-Ossandon, C.; Mena-Nieto, A.; Harsch, N. Using a fuzzy topsis-based scenario analysis to improve municipal solid waste planning and forecasting: A case study of canary archipelago (1999–2030). *J. Clean. Prod.* **2018**, *176*, 1198–1212. [[CrossRef](#)]
31. Sun, C.; Sun, C.; Yang, Z.; Zhang, J.; Deng, Y. Urban land development for industrial and commercial use: A case study of beijing. *Sustainability* **2016**, *8*, 1323. [[CrossRef](#)]
32. Song, C.; Pei, T. Exploring polycentric characteristic and residential cluster patterns of urban city from big data. *J. Geo. Inf. Sci.* **2019**, *21*, 384–397.
33. Jokar Arsanjani, J.; Helbich, M.; Bakillah, M.; Hagenauer, J.; Zipf, A. Toward mapping land-use patterns from volunteered geographic information. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2264–2278. [[CrossRef](#)]
34. Yao, F.; Luo, J.; Shen, Z.; Dong, D.; Yang, K. Automatic urban vegetation extraction method using high resolution imagery. *J. Geo. Inf. Sci.* **2016**, *18*, 248–254.
35. Long, Y.; Liu, X. Automated identification and characterization of parcels (aicp) with openstreetmap and points of interest. *Environ. Plann. B* **2015**, *43*, 498–510.
36. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and pois. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194.
37. Ng, H.; Zelle, J. Corpus-based approaches to semantic interpretation in natural language processing. *AI Mag.* **1997**, *8*, 45–64.
38. Zhang, D.; Xu, H.; Su, Z.; Xu, Y. Chinese comments sentiment classification based on word2vec and svmperf. *Expert Syst. Appl.* **2015**, *42*, 1857–1863. [[CrossRef](#)]

39. Zafari, A.; Zurita-Milla, R.; Izquierdo-Verdiguier, E. Evaluating the performance of a random forest kernel for land cover classification. *Remote Sens.* **2019**, *11*, 575. [[CrossRef](#)]
40. Oliphant, A.J.; Thenkabail, P.S.; Teluguntla, P.; Xiong, J.; Gumma, M.K.; Congalton, R.G.; Yadav, K. Mapping cropland extent of southeast and northeast asia using multi-year time-series landsat 30-m data using a random forest classifier on the google earth engine cloud. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *81*, 110–124. [[CrossRef](#)]
41. Tang, H.; Shi, Y.; Dong, P. Public blockchain evaluation using entropy and topsis. *Expert Syst. Appl.* **2019**, *117*, 204–210. [[CrossRef](#)]
42. Chen, P. Effects of normalization on the entropy-based topsis method. *Expert Syst. Appl.* **2019**, *136*, 33–41. [[CrossRef](#)]
43. dos Santos, B.M.; Godoy, L.P.; Campos, L.M.S. Performance evaluation of green suppliers using entropy-topsis-f. *J. Clean. Prod.* **2019**, *207*, 498–509. [[CrossRef](#)]
44. Zhang, M.; Zhao, P. The impact of land-use mix on residents' travel energy consumption: New evidence from beijing. *Transp. Res. Part D Transp. Environ.* **2017**, *57*, 224–236. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).