Article

Classifying Urban Functional Zones Based on Modeling POIs by Deepwalk

Xin Yang *, Shuaishuai Bo and Zhaojie Zhang

School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China
* Correspondence: yangxin@qut.edu.cn

Abstract: Developing urban functional zone classification method to study urban spatial structure is a hotspot in current research. Using the word embedding model to excavate spatial relationship of the geographic elements in urban functional zones is an important way to develop urban functional zone classification method. However, in these studies, the spatial relationship of geographic elements was regarded as their homogeneity, while the structural similarity of geographical elements was ignored, which inevitably reduces the classification accuracy of urban functional zone classification method. This paper proposes to develop an urban functional zone classification method based on Deepwalk model, which could extract homogeneity and structural similarity of nodes in graph. The proposed method uses POI data to represent geographical elements, organizes POIs into graphs, and uses Deepwalk to embedding POIs for urban functional zone classification. It was applied to classify the urban functional zones of Chaoyang district in Beijing; and the classification results were compared with those of two baseline method based on Word2vec model and Place2vec model. The experimental results show that considering both the homogeneity and structural similarity of geographical elements, the proposed model has higher accuracy than the models only considering the homogeneity of geographical elements.

Keywords: urban functional zone classification; homogeneity of geographic elements; structural similarity of geographic elements; POI; deepwalk

1. Introduction

With the development of urbanization and modern civilization, various urban functional zones (UFZs) have emerged in cities, which have become the spatial carriers of different social and economic activities to meet people's diverse needs for life, work, education and public services [1–3]. Classifying UFZs to get the spatial distribution of UFZs and generate UFZ map is of great significance to urban planners and managers, and can provide many useful applications for travel recommendation, commercial location selection, etc. [4–7]. However, UFZ is not only artificially designed by urban planners, but also formed naturally according to people's actual lifestyles; the rapid urbanization process in recent years has further increased the complexity of the spatial structure of UFZs [8,9]. All of these pose serious challenges to UFZ classification. Therefore, scholars have been studying the methodology of UFZ classification for decades.

Early studies on the classification of urban functional zone were mainly subjective and qualitative. These studies benefit from, but are limited to, human subjective knowledge. The spatial scales of these studies are generally large, mostly towns, districts or streets [10]. With the emergence of geographic big data and social big data, the classification method of urban functional zones has gradually changed to objective analysis methods based on big data, and the spatial scales of urban functional zone researches have also been refined into blocks, regular grids and so on. In these studies, UFZ classification methods can be divided into four categories: (1) methods based on statistical characteristics of geographical elements of UFZs [11–14]; (2) methods based on local spatial relationship characteristics

of geographical elements in UFZs [15,16]; (3) methods based on the semantic features of UFZs [17–21]; (4) methods based on abstract features of UFZs extracted by algorithms such as deep learning [22–24].

Constructing or advancing a method based on the local spatial relationship characteristics of geographical elements is an important research direction for UFZ classification method. This particular method utilizes POI data to represent geographical elements, embeds POI categories into vector space through the use of a word embedding model, and calculates UFZ vectors for UFZ classification based on the distribution of POIs [11–14].

Method based on local spatial relationship characteristics of geographical elements was first introduced by Yao et al. [15]. They employed Word2vec (a word embedding model) to explore the spatial relationship of POIs in terms of land use classification. In their work, a traffic analysis zone (TAZ)-based corpus was constructed by connecting POIs sequentially based their spatial distances. The spatial distribution relationships of POIs are represented in POI sequence, like the language sequence. Hu et al. [25] pointed out that Word2Vec model ignores the ubiquitous homonymy and polysemy issues of words and embeds each word using only a single vector and make improvements accordingly. Niu et al. [26] considered the Word2Vec model fails to capture the spatial heterogeneity among urban areas because some areas may share the same ratios of different POI classes, but differ in terms of the spatial arrangement of the POIs. They proposed to integrate the Doc2Vec model, an extension of Word2Vec, with POI data to train vectors of urban areas directly, as well as training the vectors of POI classes through the neural network. Yan et al. [27] believe that the structure of geographic space differs substantially from the sequence in language. They introduce the spatial distance among POIs into Word2Vec to augment the spatial contexts of POI types, and presented a novel embedding-based model called Place2Vec. Based on the Place2Vec model and POIs data, Zhai et al. [28] constructed a neighborhood area–based corpus and trained high-dimensional characteristic vectors of POIs categories to identify urban functional regions. Sun et al. [29] proposed Block2vec based on Word2vec to extract the spatial semantic information among regions. The Block2vec model can map the spatial correlation between the POIs, as well as the regions, to a high-dimensional vector, in which classification of urban functional regions can be better performed. Through the continuous efforts of scholars, the method of UFZ classification based on the local spatial relationship characteristics of geographical elements has been continuously improved.

However, these studies still have not fully explore the spatial relationships of geographical elements. These studies specify the spatial context information of a geographical element as its set of nearby geographical elements. For example, Zhai et al. [28,29] construct corpus for each POI based on its K-nearest POIs to mine the spatial correlation between the POIs; Niu et al. [26] construct corpus for each POI based on its neighboring POIs within a certain search radius to mine the spatial correlation between the POIs. In these cases, the closer the spatial distance, the higher the similarity of geographical elements, which is consistent with the first law of geography [27]. This means that in these methods, the spatial relationship of geographical elements is considered as the homogeneity of geographical elements. In fact, the spatial relationship of geographical elements is very complexity. As the carrier of urban social and economic functions, geographical elements form an invisible social network based on spatial relationships. Geographical elements, as nodes in the network, not only have homogeneity in spatial relationship, but also have similarity in structure (geographical elements that are far apart but adjacent to similar structures have higher similarity) [30]. Obviously, the existing researches ignore the structural similarity of geographic elements, which inevitably reduces the feature expression accuracy of geographic elements, thus reduces the classification accuracy of urban functional zone classification method.

To fill the research gap, this paper proposes a method for UFZ classification based on the Deepwalk model. Deepwalk is a representative network structure analysis model that generates node sequences through random walks, and then inputs the sequences into the

Word2vec to learn node embedding [31,32]. The model is proven effective in extracting node homogeneity and structural similarity, and is widely used in POI recommendation [30], community detection [33,34], and other areas [35,36]. In the proposed method, POI data is used to represent geographical elements; POIs within a zone are organized on a graph based on the spatial distances among POIs; then, the Deepwalk model is used to study the spatial relationship of POIs for urban functional classification.

The remainder of this paper is organized as follows. Section 2 presents the methodology. Section 3 presents the case studies. In Section 4, conclusions are presented, along with some suggestions regarding the direction for future research.

2. Methodology

POIs are discrete points in geographic space. When using the Deepwalk model to model POIs, it is necessary to organize the POIs of each UFZ into a graph. The process of the proposed method includes three parts: (1) Construct the POI graphs of UFZs; (2) Generate the vectors of POI categories based on the Deepwalk algorithm; (3) Calculate vector of each UFZ by aggregating vectors of POIs in it; (4) Classify UFZs based on SVM (Support Vector Machine). The flowchart of the proposed method is illustrated in Figure 1.
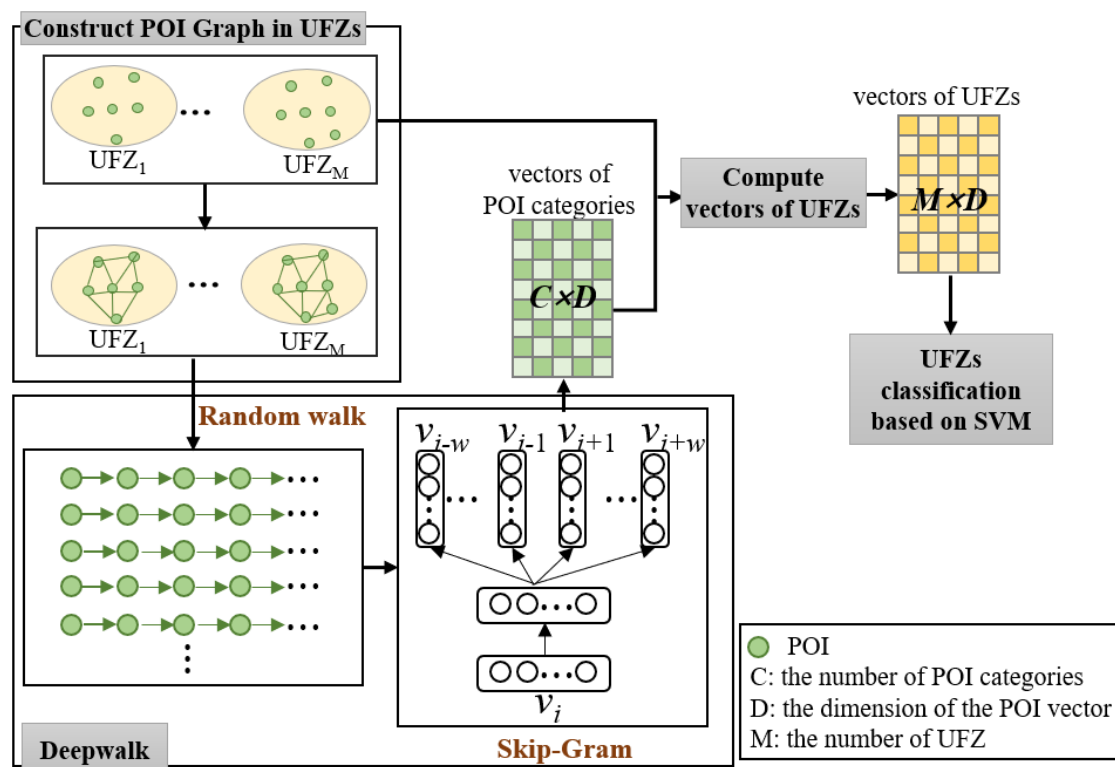

Figure 1. Flowchart of the proposed method.

2.1. Constructing POI Graphs in UFZs

The key to constructing a POI graph is to determine the edges between POIs according to the spatial relationship of POIs. In the POI recommendation literatures, POIs are organized into graphs according to human behavior, emphasizing the relationship between POIs and people [37]. In this method, the spatial distribution pattern of POIs is emphasized, and POIs are organized into a graph according to the spatial distance between POIs.

There are two ways to connect POIs to generate edges and construct graph, one is to connect each POI to neighboring POIs within a certain search radius [26], and the other is to connect each POI to its K-nearest POIs [28,29]. In this paper, it was decided to connect each POI to its K-nearest POIs for the following reasons. POIs are points and cannot represent the geometric properties of geographic entities. However, as we known, different geographic

entities have different geometric characteristics, and their spatial distributions are also different. Some geographic entities are small, and these POIs are spatially concentrated, such as restaurant and convenience store. Some geographic entities are relatively large, and these POIs are relatively sparse in geographic space, such as residential quarter, park, and factory. Therefore, searching the neighbors of all POIs with the same radius will weaken the relationship between large geographic entities and other geographic entities.

2.2. Embedding POI Type in Vector Space Based on Deepwalk

DeepWalk was proposed by Perozzi et al. in 2014 based on Word2Vec, it can learn latent representation of vertices in a network graph using the co-occurrence relationship among the whole nodes in the graph [30]. There are two stages in the process of the Deepwalk method: (1) The random walk algorithm is used to construct sequences of nodes. For each vertex v_i, γ random walks with length t are conducted, with v_i as the starting vertex. (2) The Skip-Gram language model is used to process the special language composed of the set of randomly-generated walks, then the discrete nodes can be represented as vectors in the network. Skip-Gram maximizes the co-occurrence likelihood of the vertices that come into view within a window w using an independent assumption as follows [36]:

$$P\left(\frac{v_{i-w},\cdots,v_{i-1},v_{i+1},\cdots,v_{i+w}}{v_i|\Phi(v_i)}\right) = \prod_{k=i-w,k\neq i}^{i+w} P(v_k|\Phi(v_i)) \qquad (1)$$

In Equation (1), $v_{i\pm k}$ are the context of the word v_i, w is the size of the window. In addition, we map each vertex v_k to its current representation vector $\Phi(v_k) \in R^d$, and d is the dimension of the vertex vector. To speed up the training time, $P(v_k|\Phi(v_i))$ is factorized with Hierarchical SoftMax by allocating the vertices to the leaves of a binary tree, and $P(v_k|\Phi(v_i))$ is then computed as follows [36]:

$$P(v_k|\Phi(v_i)) = \prod_{l=1}^{log|V|} 1/\left(1 + e^{-\Phi(v_i)\bullet\psi(b_l)}\right) \qquad (2)$$

In Equation (2), $\psi(b_l)$ represents the parent of tree node b_l. $(b_0, b_1, \cdots, b_{log|v|})$ is the sequence of tree nodes to identify v_j, where $b_l = root$.

In order to reduce the number of model parameters and the complexity of program implementation, in this paper, the walk length t in the random walk is set to the same as the window w in the Skip-Gram, which does not change the principle of the model.

It can be seen from the above process that the node sequences generated by random walk in network includes not only the proximity of the nodes, but also the network structure of the nodes. Therefore, the similarity of node embeddings includes both homogeneity and structural similarity.

2.3. Embedding Urban Functional Zone in Vector Space

After completing the training by the Deepwalk, the vectors of all POI categories are obtained. Then, the UFZ vectors are calculated by averaging the vectors of the inner POIs. The UFZ vectors can be specified mathematically by Equation (3).

$$UFZ_i = \frac{\sum_{k=1}^{N} typeVec(POI_{ik})}{N} \qquad (3)$$

In Equation (3), N is the number of POIs in the i-th UFZ, POI_{ik} denotes the k-th POI in the i-th UFZ, $typeVec(POI_{ik})$ denotes the vector of the k-th POI in the i-th UFZ.

2.4. Classifying Urban Functional Zone Using SVM

In this section, based on the UFZ vectors calculated by the above method, the SVM is used to classify UFZs. The reason for choosing the SVM algorithm is that the UFZ vector is high-dimensional, and the SVM has shown high classification efficiency for high-dimensional features in previous studies.

In this study, 70% of UFZs in each urban functional type are selected as training samples, and the other 30% are used as testing samples. The SVM model requires two finely tuned parameters: the penalty C factor and the kernel parameter. To find the best parameter configuration, 75% of the training samples are used for model training and the other 25% are used for validation. The parameter combination leading to the highest validation accuracy is deemed the best parameters. Then, the optimized SVM is trained using the training samples and applied to the testing samples for accuracy assessment.

3. Case Study

To demonstrate the feasibility of the proposed model, it was realized by our research team using Python and several python libraries (such as Scipy, Genism and sklearn). Then a case study was conducted to classify UFZs in the Chaoyang District, which is one of the six main districts of Beijing, China. Chaoyang has a large population proportion within the jurisdiction of Beijing. It has developed industry and frequent diplomatic activities, and is highly representative in economy, housing, education and other aspects. Urban functions in Chaoyang are complex and diversity.

3.1. Data Source

In this paper, we analyze and classify the functional zones of Chaoyang district at TAZ (traffic analysis zone) level, which is most commonly used in the existing research. Road net data of Chaoyang district were prepared, and were input into the Feature To Polygon function of ArcMap to generate UFZs. The Chaoyang district were divided into 2411 UFZs, shown in Figure 2a.
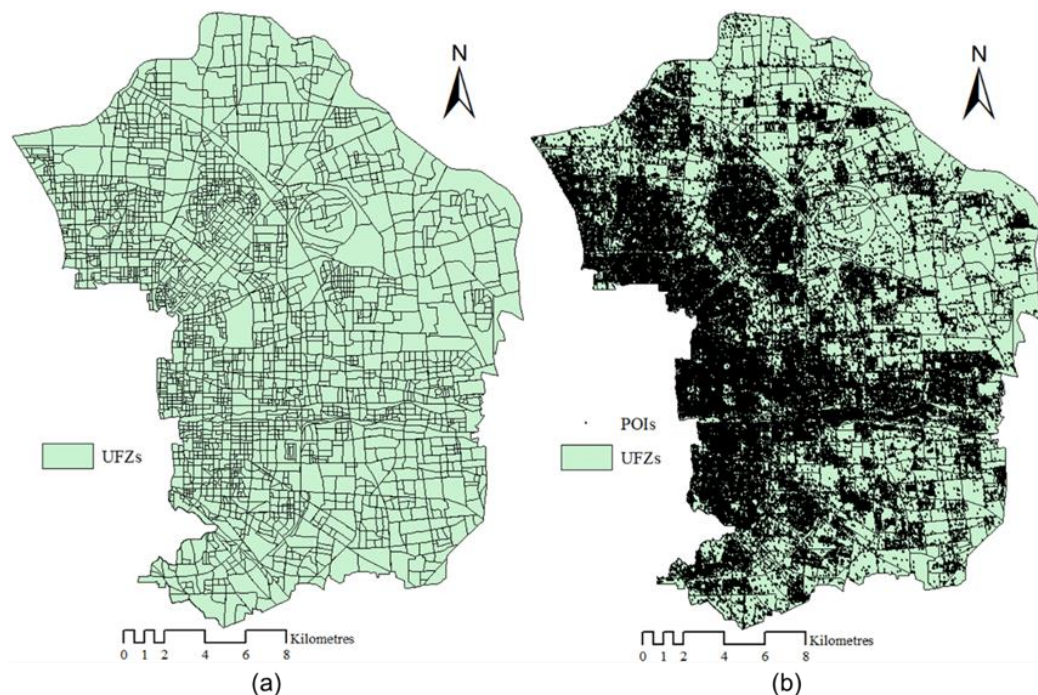

Figure 2. (**a**) UFZs in Chaoyang District; (**b**)POIs in Chaoyang District.

The POI dataset used in this research was fetched via application programming interfaces (APIs) provided by Gaode Map Services (https://www.amap.com/, accessed on 17 April 2020). We fetched 177165 records of POIs in Chaoyang district, the spatial distribution of POIs is shown in Figure 2b. In our POI dataset, there are 23 labels in the top-level category and 906 labels in the final level category. In order to ensure the richness of node types in Deepwalk model, the final level category of POI was chosen as its category.

Figure 3 displays the classification results of UFZs, which is annotated by volunteers with background knowledge of urban planning on the basis of urban land use planning

data, POIs and high spatial resolution remote sensing image. In the following research, the manual annotated results is token as the actual urban functional zone map. The urban functional type of Chaoyang district contains 11 different types of UFZs, which are Villages (V), Commercial Regions (C), Leisure and Entertainment Regions (LE), Residential Regions (R), Residential and Commercial Regions (RC), Foreign Embassy and Consulate Regions (F), Science, Education& Cultural Regions (RCE), Industrial Regions (I), Building materials and Hardware Regions (BH), Auto Service Regions (A), Village and Leisure and Entertainment Regions (VLE).
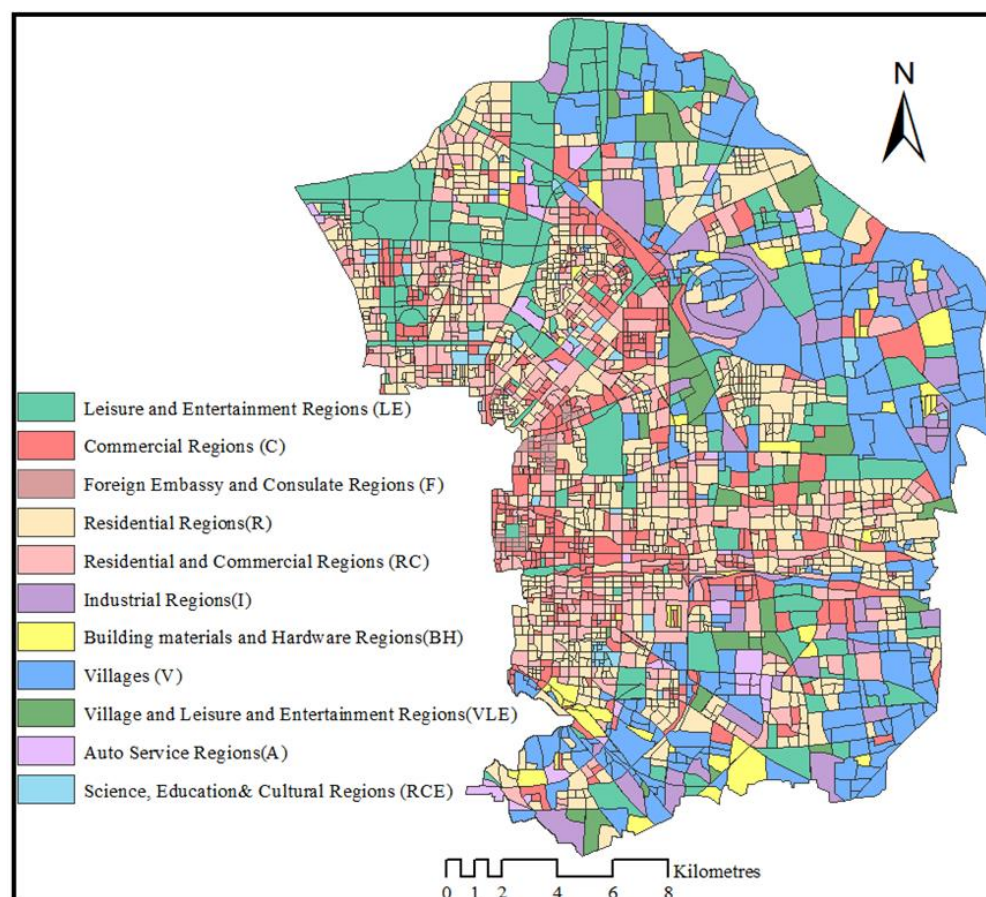

Figure 3. The manual classification results of UFZs in Chaoyang district.

3.2. Result

3.2.1. Parameters Sensitivity Analysis

In the proposed model, there are four key parameters (the number of neighbors K, embedding dimension d, walk count γ and walk length t), which play significant roles in urban functional zone classification. Hence, sensitivity analyses between them and the overall accuracy (OA) of urban functional zone classification were carried out.

Analyzing four uncertain parameters at the same time is a very large task. In those four parameters, the number of neighbors K is a parameter related to the homogeneity of POIs. The number of neighbors K and the embedding dimension d are not only present in the proposed model, but also important parameters of the urban functional zone classification model based on the homogeneity of POIs. To simplify the analysis process, we first use the Place2vec model to analyze the impact of the number of neighbors on the overall accuracy. As description in Section 1, Place2vec is a representative urban functional zone classification model based on the homogeneity of POIs. Place2vec model has the only two paraments. In the experiments, the number of neighbors was set to 4–50, the embedding dimension was set to 50, 100, 200, 300, and 400. The relationship between the number of

neighbors, the embedding dimension and the overall accuracy of classification is shown in Figure 4.
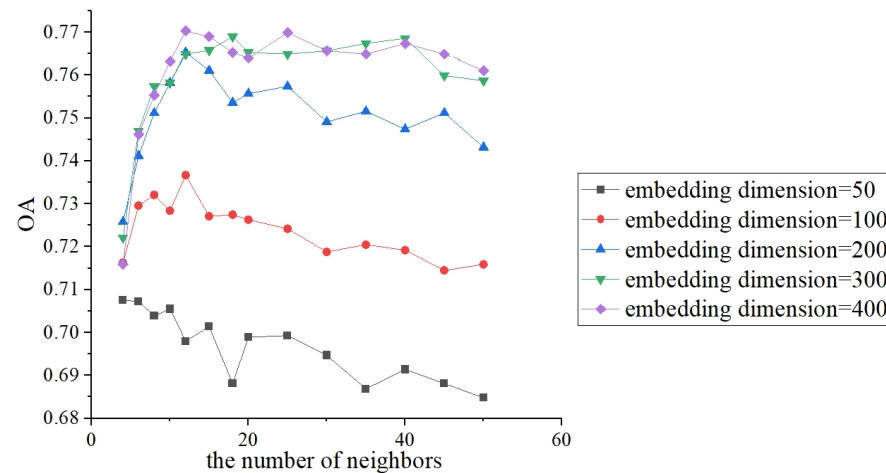

Figure 4. Accuracy assessment of urban functional zone classification using different numbers of neighbors and different embedding dimensions.

When the embedding dimension is 200, 300 and 400, the overall accuracy reaches a higher value; when the number of neighbors is set to 12, the difference in the overall accuracies is not significant. However, as the embedding dimension increases, the computational load of the program also increases significantly. Therefore, in the following experiments, the number of neighbors is set to 12 and the embedding dimension is set to 200.

After determine the number of neighbors and the embedding dimension, we discuss the impact of the walk count γ and the walk length t on the overall accuracy, respectively. Due to the randomness of corpus generated by random walk, even if the parameters are fixed, the overall accuracy of urban functional zone classification also has some randomness. Therefore, for each pair of parameters, repeated experiments are needed, and the overall accuracy of classification results would be in a range. In this paper, experiments for each pair of parameters were repeated 30 times. the relationship between walk count and overall accuracy of classification results, and the relationship between the walk length and overall accuracy of classification results are displayed in Figure 5.

From Figure 5a, it can be seen that with the increase of walk length, the average overall accuracy of urban functional zone classification gradually increases, and then tends to be stable. When the walk length increases to a certain value, the average overall accuracy begins to decrease. When the walk length is between 8 and 18, the average overall accuracy is relatively high and the difference is small; but the range of the overall accuracy first decreases and then expands, that is, the classification accuracy first stabilizes and then fluctuates. Figure 5b shows that increasing the walk count leads to a decrease in the overall accuracy range, which also means that the classification accuracy tends to stabilize. As the walk count increases, the average overall accuracy of the classification results increases first and then decreases slowly. Therefore, we can usually obtain more stable classification results by increasing the walk count appropriately. Combine these two figures, when the walk count is equal to 20 and the walk length is equal to 12, the proposed model could have better classification capability and higher classification accuracy.

Through the above experiments, we choose to set the number of neighbors K to 12, the embedding dimension d to 200, the walk count γ to 20, and the walk length t to 12.

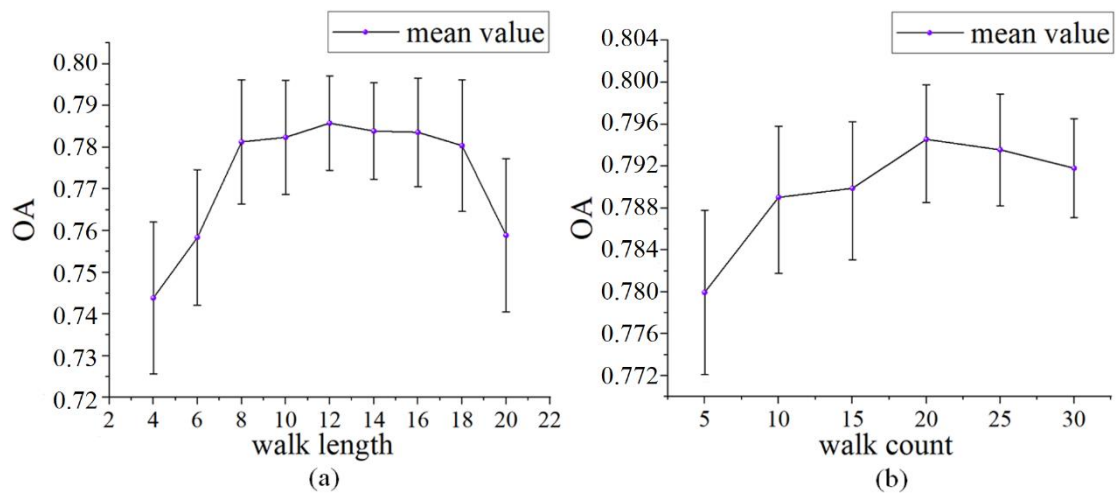**Figure 5.** (**a**) Accuracy assessment of urban functional zone classification using different walk lengths, while the walk count is set to 20; (**b**) Accuracy assessment of urban functional zone classification using different walk counts, while the walk length is set to 12.

3.2.2. Result Analysis

Setting the parameters according to the conclusion of the last section ($K = 12$, $d = 200$, $\gamma = 20$, $t = 12$), we repeated the urban functional zone classification of Chaoyang district using the proposed model for 30 times. Figure 6 shows the urban functional zone classification results whose overall accuracy is closest to the average overall accuracy.
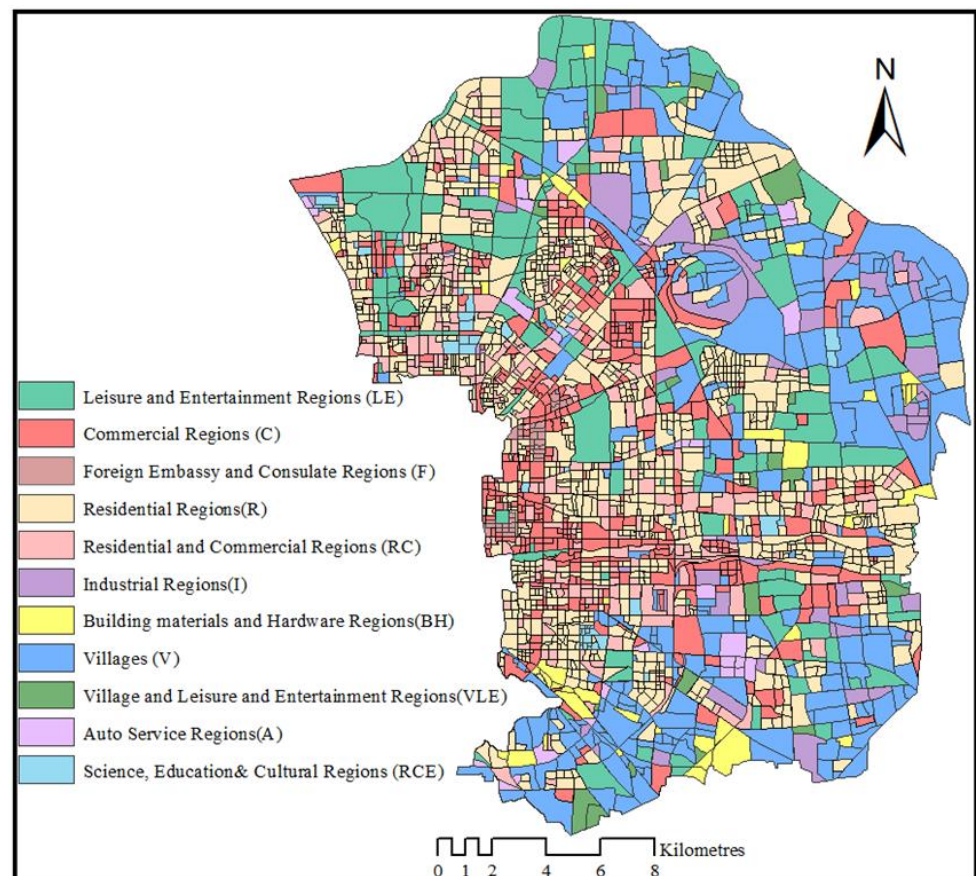

Figure 6. Urban functional zone classification result using the proposed method.

The visual comparison shows that the classification results of the proposed method are closer to the manual interpretation results. The spatial distribution of functional zones is closely related to the geographical location of Chaoyang District in Beijing. Leisure and Entertainment Regions are evenly distributed. Villages, Industrial Regions, Building materials and Hardware Regions are mainly distributed in the east, north and south of Chaoyang District. These coincide with the fact that these parts of Chaoyang District are adjacent to Tongzhou, Shunyi and Daxing district, which are not core districts of Beijing. Commercial Regions, Residential Regions, Residential and Commercial Regions, Science, Education & Cultural Regions are mainly distributed in the west and central Regions; and Foreign Embassy and Consulate Regions are located in the west. These coincide with the fact that the west of Chaoyang District is adjacent to Dongcheng district, a core district of Beijing.

Following preliminary validation by visual comparison, the accuracy of the classification result was analyzed quantitatively. In this process, the classification results and the manually interpreted urban functional zone map were compared zone by zone. Table 1 illustrates the confusion matrix of the classification result based on the proposed model. In this table, the rows represent the classification result of the proposed model, and the columns represent the manual classification result. It can be seen from the table, Residential Regions (R) and Foreign Embassy and Consulate Regions (F) have highest accuracy rates, 0.917 and 0.973 respectively; Villages (V), Industrial Regions (I), Commercial Regions (C), Leisure and Entertainment Regions (LE), Education & Cultural Regions (RCE)and Auto Service Regions (A) have higher accuracy rates, 0.791, 0.709, 0.745, 0.789, 0.719 and 0.732 respectively; Village and Leisure and Entertainment Regions (VLE), Residential and Commercial Regions (RC) and Building materials and Hardware Regions (BH) have lower accuracy rates, 0.469, 0.553 and 0.586 respectively. For the three functional types with lower accuracy rates, 28.1% and 12.5% of Village and Leisure and Entertainment Regions (VLE) are divided into Villages (V) and Commercial Regions (C); 29% and 11.8% of Residential and Commercial Regions (RC) are divided into Residential Regions (R) and Commercial Regions (C); 13.8%, 10.3% and 10.3% of Building materials and Hardware Regions (BH) are divided into Villages (V), Industrial Regions (I) and Commercial Regions (C). In terms of the confusion functional types, the main reason for the low classification accuracy is that their POI quantity distributions are similar to their confusion functional types.

Table 1. The confusion matrix of this method.

	V	VLE	R	RC	BH	I	C	LE	RCE	A	F
V	**0.791**	0.281	0.017	0.025	0.138	0.127	0.054	0.023	0.053	0.098	0.000
VLE	0.005	**0.469**	0.001	0.003	0.000	0.000	0.002	0.023	0.000	0.000	0.000
R	0.104	0.000	**0.917**	0.290	0.034	0.000	0.106	0.058	0.140	0.000	0.000
RC	0.005	0.031	0.037	**0.553**	0.017	0.000	0.052	0.012	0.053	0.024	0.027
BH	0.014	0.031	0.000	0.005	**0.586**	0.018	0.004	0.006	0.018	0.000	0.000
I	0.019	0.000	0.002	0.000	0.103	**0.709**	0.019	0.012	0.000	0.000	0.000
C	0.043	0.125	0.020	0.118	0.103	0.127	**0.745**	0.047	0.018	0.122	0.000
LE	0.019	0.063	0.002	0.003	0.017	0.000	0.004	**0.789**	0.000	0.024	0.000
RCE	0.000	0.000	0.002	0.000	0.000	0.000	0.009	0.023	**0.719**	0.000	0.000
A	0.000	0.000	0.001	0.003	0.000	0.018	0.004	0.006	0.053	**0.732**	0.000
F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.000	**0.973**

3.3. Comparison

To validate the proposed model, a comparative study was carried out by using the Word2vec model and the Place2vec model, which classify urban functional zones based on the spatial relationship of POIs and regard the homogeneity of POIs as their spatial relationship. The feasibility and effectiveness of the Word2vec model in urban functional zone classification were proved by Yao et al. [15]. And it has been shown that the Place2vec model performs better than other semantic models in urban functional zone classification [28]. In

the Word2vec model, the building method of corpus refers to paper [15]. In the Word2vec model and the Place2vec model, after derived POI vectors, each zone vector was got by computing the average vector of all POIs inside the zone. And then, SVM was used with zone vectors to classify UFZs. Parameters in the Place2vec were set to the same as those in the proposed model ($K = 12$, $d = 200$).

Figure 7 shows the classification results of urban functional zones in Chaoyang district using the Word2vec model and the Place2vec model. Comparing Figures 6 and 7a, it is obviously that the proposed model has higher classification accuracy than the Word2vec model. In the urban functional zone map classified using the Word2vec model, the accuracy of Villages (V), Leisure and Entertainment Regions (LE), and Residential Regions (R) are lower. Comparing Figures 6 and 7b, it is difficult to directly judge which of the two models has the higher classification accuracy. Then, the classification results of the three methods were compared and analyzed quantitatively, as shown in Table 2.
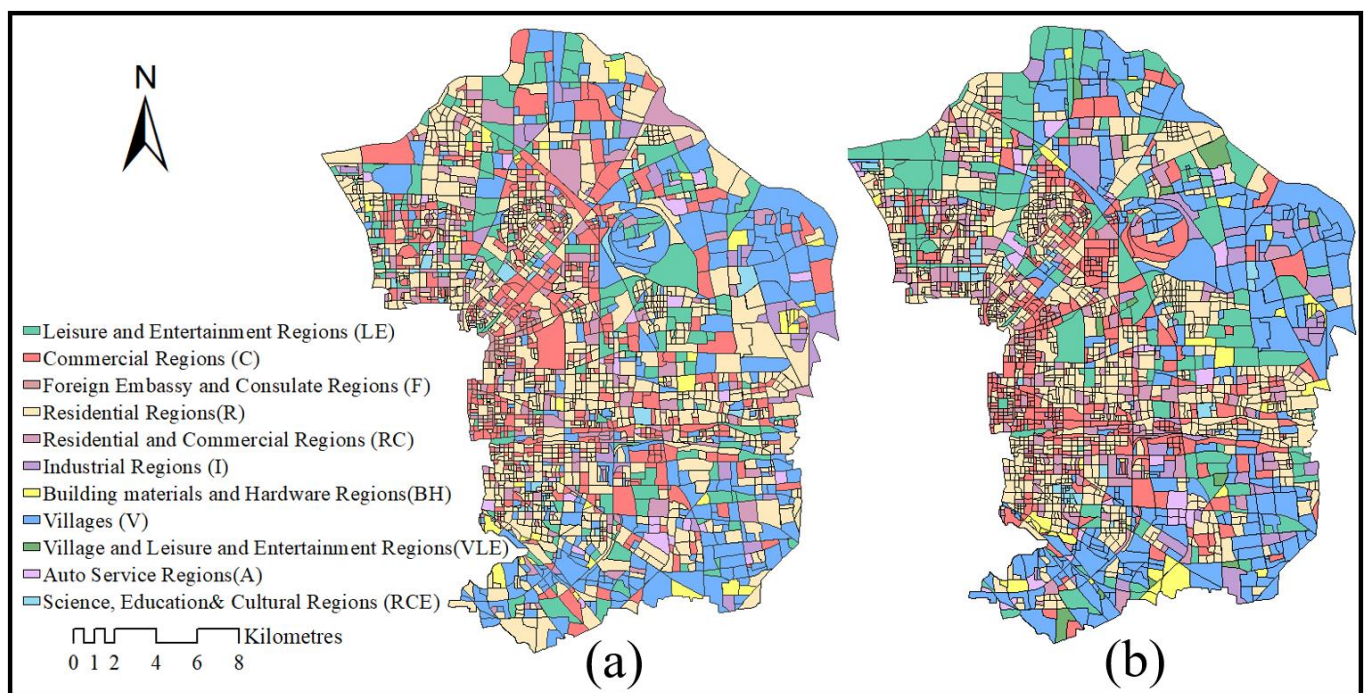

Figure 7. Urban functional zone classification results using (**a**) the Word2vec model, (**b**) the Place2vec method.

Table 2. Comparison of accuracy between two methods.

	Word2vec Model	Place2vec Model	The Proposed Method
V	0.649	0.787	0.791
VLE	0.000	0.281	0.469
R	0.887	0.914	0.917
RC	0.411	0.512	0.553
BH	0.466	0.500	0.586
I	0.273	0.618	0.709
C	0.674	0.732	0.745
LE	0.696	0.766	0.789
RCE	0.632	0.754	0.719
A	0.610	0.707	0.732
F	0.973	0.973	0.973
OA	0.694	0.765	0.784
Kappa	0.598	0.694	0.718

Table 2 shows the accuracies of the three methods in the classification results of urban functional zones in the study area, including the classification accuracy rate of each functional type, overall accuracy and Kappa index. Overall, the overall accuracy and kappa index of the Word2vec model are 0.694 and 0.598, which are significantly lower than the other two models; the overall accuracy and kappa index of the Place2vec method are 0.765 and 0.694, respectively, while the overall accuracy and kappa index of the proposed method are 0.784 and 0.718, indicating that the accuracy of the proposed model is higher than the Place2vec model. From the specific functional types, the urban functional zones classified by the Word2vec model, except for F zones, have lower accuracy rate than the functional zones classified by the other two models. The urban functional zones classified by the proposed method, except for RCE and F, have higher accuracy rate than the functional zones classified by the Place2vec; the F zones classified by the proposed method have equal accuracy rate with the F zones classified by Place2vec. It can be seen that the proposed method is superior to the Word2vec model and the Place2vec method. It is proved that considering both the homogeneity and structural similarity of geographical elements, the proposed model has higher accuracy than the model only considering the homogeneity of geographical elements.

4. Conclusions

The rapid development of urbanization has intensified the complexity of the spatial structure of urban functional zone, and also brings great challenges to the identification of urban functional zones. Embedding geographic elements based on their spatial relationship, represented by Place2vec, is an important method for urban functional zone classification. In previous studies, the spatial relationship of geographic elements is considered as the homogeneity. However, the spatial relationship of geographic elements contains not only homogeneity but also structural similarity. To solve the problem, this paper carries out two aspects of work: (1) Propose a new urban functional classification method based on POIs and Deepwalk to model both the homogeneity and structural similarity of geographical elements. (2) Apply the proposed method to classify urban functional zones of Chaoyang District, Beijing; and the classification results were compared with the Word2vec model and the Place2vec model. The results indicated that the proposed model has higher accuracy, and it is confirmed that considering both the homogeneity and structural similarity of geographical elements can help improve the classification accuracy of urban functional zones.

This model still has the limitation that the accuracy is not very stable. In order to ensure the reliability and stability of the classification results, the classification process must be repeated many times to find a suitable result. The randomness is due to the fact that the proposed model is based on Deepwalk, which examines the spatial relationships of geographic elements from their spatial contexts generated by random walks. The quality of the spatial contexts determines the accuracy of the model. In future studies, we will add walking direction control to reduce the randomness of the spatial contexts of geographical elements and improve the stability of classification results.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

References

1. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.Z.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 712–725. [CrossRef]
2. Liu, B.H.; Deng, Y.B.; Li, M.; Yang, J.; Liu, T. Classification Schemes and Identification Methods for Urban Functional Zone: A Review of Recent Papers. *Appl. Sci.* **2021**, *11*, 9968. [CrossRef]
3. Wang, Y.D.; Gu, Y.Y.; Dou, M.X.; Qiao, M.L. Using spatial semantics and interactions to identify urban functional regions. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 130. [CrossRef]
4. Cai, L.; Zhang, L.Q.Y.; Liang, Y.; Li, J. Discovery of urban functional regions based on Node2vec. *Appl. Intell.* **2022**, *52*, 16886–16899. [CrossRef]
5. Zhang, Y.; Li, Q.; Tu, W.; Mai, K.; Yao, Y.; Chen, Y.Y. Functional urban land use recognition integrating multi-source geospatial data and cross-correlations. *Comput. Environ. Urban Syst.* **2019**, *78*, 101374. [CrossRef]
6. Zhang, X.Y.; Du, S.H.; Wang, Q.; Zhou, W.Q. Multiscale Geoscene Segmentation for Extracting Urban Functional Zones from VHR Satellite Images. *Remote Sens.* **2018**, *10*, 281. [CrossRef]
7. Huang, X.; Hu, T.; Li, J. Mapping urban areas in China using multisource data with a novel ensemble SVM method. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *56*, 4258–4273. [CrossRef]
8. Xu, S.Y.; Qing, L.B.; Han, L.M.; Liu, M.; Peng, Y.H.; Shen, L.F. A New Remote Sensing Images and Point-of-Interest Fused (RPF) Model for Sensing Urban Functional Regions. *Remote. Sens.* **2020**, *12*, 1032. [CrossRef]
9. Wu, J.J.; Zhang, J.; Zhang, H.X. Urban Functional Area Recognition Based on Unbalanced Clustering. *Math. Probl. Eng.* **2022**, *2022*, 7245407. [CrossRef]
10. Huang, S.Z. Evolution of City Functional Layout of Beijing in the 1990S. *Beijing Plan. Constr.* **2003**, *3*, 22–25.
11. Andrade, R.; Alves, A.; Bento, C. POI Mining for Land Use Classification: A Case Study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 493. [CrossRef]
12. Wang, Z.Y.; Ma, D.B.; Sun, D.Q.; Zhang, J.X. Identification and analysis of urban functional area in Hangzhou based on OSM and POI data. *PLoS ONE* **2021**, *16*, e0251988. [CrossRef]
13. Chen, Y.; Chen, X.; Liu, Z. Understanding the spatial organization of urban functions based on co-location patterns mining: A comparative analysis for 25 Chinese cities. *Cities* **2020**, *97*, 102563. [CrossRef]
14. Hu, Y.F.; Han, Y.Q. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. *Sustainability* **2019**, *11*, 1385. [CrossRef]
15. 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]
16. Chen, Y.; Qian, H.Z.; Wang, X.; Wang, D.; Han, L.J. A GloVe Model for Urban Functional Area Identification Considering Nonlinear Spatial Relationships between Points of Interest. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 498. [CrossRef]
17. Chen, Z.L.; Zhou, L.L.; Yu, W.H. Identification of the urban functional regions considering the potential context of interest points. *Acta Geod. Cartogr. Sin.* **2020**, *49*, 907–920.
18. Crooks, A.; Pfoser, D.; Jenkins, A. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [CrossRef]
19. Liu, X.; He, J.; Yao, Y. Classifying urban land use by integrating remote sensing and social media data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1675–1696. [CrossRef]
20. Du, S.J.; Du, S.H.; Liu, B.; Zhang, X.Y.; Zheng, Z.J. Large-scale urban functional zone mapping by integrating remote sensing images and open social data. *GISci. Remote Sens.* **2020**, *57*, 411–430. [CrossRef]
21. Cao, K.; Guo, H.; Zhang, Y. Comparison of Approaches for Urban Functional Zones Classification Based on Multi-Source Geospatial Data: A Case Study in Yuzhong District, Chongqing, China. *Sustainability* **2019**, *11*, 660. [CrossRef]
22. Lu, W.P.; Tao, C.; Li, H.F.; Qi, J.; Li, Y.S. A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. *Remote Sens. Environ.* **2022**, *270*, 112830. [CrossRef]
23. Bao, H.Q.; Ming, D.P.; Guo, Y.; Zhang, K.; Zhou, K.Q.; Du, S.G. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data. *Remote Sens.* **2020**, *12*, 1088. [CrossRef]
24. Feng, Y.; Huang, Z.; Wang, Y.L.; Wan, L.; Liu, Y.; Zhang, Y.; Shan, X. An SOE-based learning framework using multisource big data for identifying urban functional zones. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 7336–7348. [CrossRef]
25. Hu, S.; He, Z.; Wu, L.; Yin, L.; Xu, Y.; Cui, H. A framework for extracting urban functional regions based on multiprototype word embeddings using points-of-interest data. *Comput. Environ. Urban Syst.* **2020**, *80*, 101442. [CrossRef]
26. Niu, H.F.; Elisabete, A.S. Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Comput. Environ. Urban Syst.* **2021**, *88*, 101651. [CrossRef]

27. Yan, B. From ITDL to Place2Vec–Reasoning About Place Type Similarity and Relatedness by Learning Embeddings from Augmented Spatial Contexts. In Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–10 November 2017; Volume 35.

28. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.R.; Gu, C. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [CrossRef]

29. Sun, Z.H.; Jiao, H.Z.; Wu, H.; Peng, Z.H.; Liu, L.B. Block2vec: An Approach for Identifying Urban Functional Regions by Integrating Sentence Embedding Model and Points of Interest. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 339. [CrossRef]

30. Yang, K.; Zhu, J.H. Next POI Recommendation via Graph Embedding Representation From H-Deepwalk on Hybrid Network. *IEEE Access* **2019**, *7*, 171105–171113. [CrossRef]

31. Li, Q.; Xiao, F.; An, L.; Long, X.Z.; Sun, X.C. Semantic Concept Network and Deep Walk-based Visual Question Answering. *ACM Trans. Multimed. Comput. Commun. Appl.* **2019**, *15*, 1–19. [CrossRef]

32. Berahmand, K.; Nasiri, E.; Rostami, M.; Forouzandeh, S. A modified DeepWalk method for link prediction in attributed social network. *Computing* **2021**, *103*, 2227–2249. [CrossRef]

33. Chen, Y.F.; Wang, L.; Qi, D.H.; Ma, T.H.; Zhang, W. Community Detection Based on DeepWalk Model in Large-Scale Networks. *Secur. Commun. Networks* **2020**, *2020*, 8845942. [CrossRef]

34. Yu, H.T.; Ma, R.; Chao, J.B.; Zhang, F.Z. An Overlapping Community Detection Approach Based on Deepwalk and Improved Label Propagation. IEEE Trans. *Comput. Soc. Syst.* **2023**, *10*, 311–321.

35. Yang, J.; Li, J.H.; Wei, L.; Gao, L.; Mao, F.Q. Spatiotemporal DeepWalk Gated Recurrent Neural Network: A Deep Learning Framework for Traffic Learning and Forecasting. *J. Adv. Transp.* **2022**, *2022*, 4260244. [CrossRef]

36. Chen, Z.H.; You, Z.H.; Guo, Z.H.; Yi, H.C.; Luo, G.X.; Wang, Y.B. Prediction of Drug-Target Interactions from Multi-Molecular Network Based on Deep Walk Embedding Model. *Front. Bioeng. Biotechnol.* **2020**, *8*, 338. [CrossRef]

37. Zhao, P.; Luo, A.; Liu, Y. Where to go next: A spatio-temporal gated network for next poi recommendation. *IEEE Trans. Knowl. Data Eng.* **2020**, *99*, 2512–2524.