

Article

A Spatial Semantic Feature Extraction Method for Urban Functional Zones Based on POIs

Xin Yang *  and Xi'ang Ma

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

Abstract: Accurately extracting semantic features of urban functional zones is crucial for understanding urban functional zone types and urban functional spatial structures. Points of interest provide comprehensive information for extracting the semantic features of urban functional zones. Many researchers have used topic models of natural language processing to extract the semantic features of urban functional zones from points of interest, but topic models cannot consider the spatial features of points of interest, which leads to the extracted semantic features of urban functional zones being incomplete. To consider the spatial features of points of interest when extracting semantic features of urban functional zones, this paper improves the Latent Dirichlet Allocation topic model and proposes a spatial semantic feature extraction method for urban functional zones based on points of interest. In the proposed method, an assumption (that points of interest belonging to the same semantic feature are spatially correlated) is introduced into the generation process of urban functional zones, and then, Gibbs sampling is combined to carry out the parameter inference process. We apply the proposed method to a simulated dataset and the point of interest dataset for Chaoyang District, Beijing, and compare the semantic features extracted by the proposed method with those extracted by the Latent Dirichlet Allocation. The results show that the proposed method sufficiently considers the spatial features of points of interest and has a higher capability of extracting the semantic features of urban functional zones than the Latent Dirichlet Allocation.



Citation: Yang, X.; Ma, X. A Spatial Semantic Feature Extraction Method for Urban Functional Zones Based on POIs. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 220. <https://doi.org/10.3390/ijgi13070220>

Academic Editors: Wolfgang Kainz, Jiangfeng She, Min Yang and Jun Zhu

Received: 8 May 2024

Revised: 19 June 2024

Accepted: 21 June 2024

Published: 25 June 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: urban functional zone; spatial topic model; spatial semantic features; point of interest (POI); latent dirichlet allocation (LDA)

1. Introduction

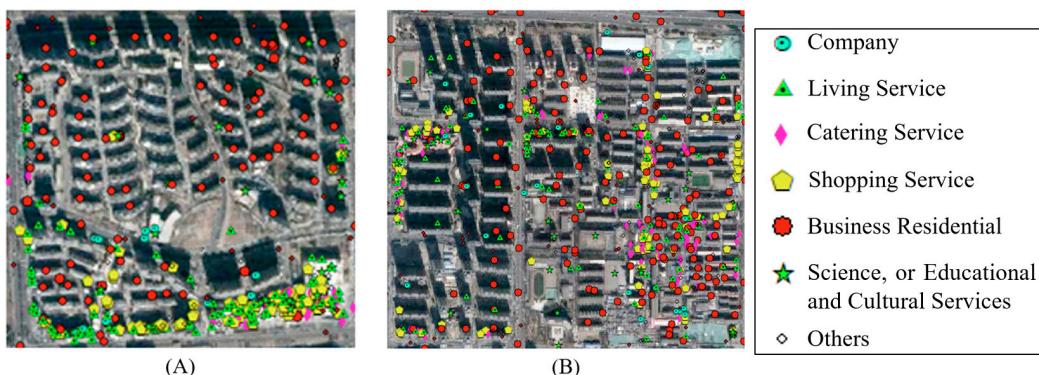
In recent years, the spatial structure of urban functions has become more complex and diverse due to rapid urbanization. This has posed significant challenges for urban planning and management. Research on the classification method of urban functional zones and explorations into urban functional spatial structures have become current hotspots in the field [1–4]. Urban functional zones are spatial units formed by the spatial aggregation of different geographic elements. Extracting the semantic features of urban functional zones from geographic elements is an important method to research the classification of urban functional zones [5–7]. Furthermore, semantic features are high-level features that can bridge the semantic gap between geographic element data and human cognition [8–12], enabling a better understanding of the functional features of urban functional zones.

Points of interest (POIs) are a type of geographic point data that represent geographic elements and contain information about their spatial locations and socio-economic attributes. They are widely used for research on extracting the semantic features of urban functional zones due to their wide coverage, fast updating, and easy accessibility [12–16]. In this type of research, urban functional zones are treated as documents in natural language; POIs are treated as words; and then, topic models (such as Latent Dirichlet Allocation [17] and probabilistic Latent Semantic Analysis) in natural language processing are used to extract the semantic features of urban functional zones. For example, Xing et al. [10,18,19]

used Latent Dirichlet Allocation (LDA) to extract semantic features of urban functional zones from POIs. Du [20] and Sun et al. [21] used LDA to extract the semantic features of urban functional zones from high-spatial-resolution remote sensing imagery and POIs. Gao et al. [22] used Embedded Topic Model to extract the semantic features of urban functional zones from high-spatial-resolution remote sensing imagery and POIs. Du et al. [23] combined taxi trajectory data, bicycle stock data, and POIs to extract the semantic features of urban functional zones by using probabilistic Latent Semantic Analysis (pLSA) and LDA. Liu et al. [9] used pLSA and LDA, while Zhang et al. [24] used LDA, combined with high-spatial-resolution remote sensing images, POIs, and real-time Tencent user data, to extract the multi-factor semantic features of urban functional zones. Zhang et al. [25] used Dirichlet Multinomial Regression (DMR) to extract the semantic features of urban functional zones from POIs and bicycle rental records. Yu et al. [26] used DMR to extract the spatiotemporal semantic information of urban functional zones from POIs and the Sina Weibo check-in data. Yuan [1] and Chen et al. [27] combined POIs and the GPS trajectories of floating cars to extract the semantic features of user movement patterns within urban functional zones using LDA.

Topic model is a text mining technology based on a probabilistic model. By analyzing the co-occurrence relationships among words in documents, it extracts the topics of the documents and generates a probability distribution of words for each topic and a probability distribution of topics for each document. Using topic model to extract semantic features of urban functional zones from POIs, the type and quantitative features of POIs in urban functional zones can be fully considered. However, topic model cannot consider the spatial location relationship of POIs due to its treatment of documents as bags-of-words. This limitation reduces the accuracy of extracting semantic features of urban functional zones. The spatial location attribute is a crucial feature of geographic elements, which is different from words in natural language processing.

The semantic features of an urban functional zone are not only reflected in the quantitative combination of geographic elements but also reflected in the spatial combination of geographic elements. Firstly, two urban functional zones with the same types and quantities of POIs may have different semantic features due to the different spatial distribution patterns of their POIs (such as spatial aggregation and uniform distribution). As shown in Figure 1A, the urban functional zone contains Shopping Service POIs, Catering Service POIs, Living Service POIs, and Company POIs. These POIs are spatially clustered and spatially separated from Business Residential POIs. Thus, the urban functional zone in Figure 1A can be considered a mixed zone with both residential and commercial functions. As shown in Figure 1B, the urban functional zone also contains Shopping Service POIs, Catering Service POIs, Living Service POIs, Company POIs, and Business Residential POIs. But, these POIs are evenly distributed. Thus, the urban functional zone in Figure 1B can be considered a mature residential zone with complete service functions. Secondly, two urban functional zones with the same types and quantities of POIs may have different semantic features due to different heterogeneity of their POIs. Thirdly, two urban functional zones with different quantities of POIs may have the same semantic features due to the same types of POIs. Ignoring the spatial features of POIs and only considering the type and quantitative features of POIs would reduce the accuracy of the semantic features of urban functional zones.



When ignoring spatial features of POIs : (A) and (B) contain similar types and quantities of POIs and have similar semantic features.

When considering spatial features of POI : (A) and (B) have different spatial semantic features.

Figure 1. Two urban functional zones with similar types and quantitative semantic features but different spatial distribution patterns.

Currently, some researchers have recognized the limitations of topic models in solving spatial feature problems such as computer vision and have improved upon topic models by integrating the spatial features of their research objects. For example, when studying computer images, Wang [28], Pan [29], and Li et al. [30] found that it was difficult to use topic model to model the spatial correlation between image patches in images and proposed three different spatial topic models for image studies. The Spatial Latent Dirichlet Allocation (SLDA) proposed by Wang et al. [28] designed cross-overlapping image subregions as documents and used image patches in the image subregions as visual words, thus encoding the spatial information of the visual words in the documents. The Markov Topic Random Field (MTRF) proposed by Pan et al. [29] treats the parts of visual words as topics and uses Markov Random Fields to establish the relationship between neighboring topics to reflect the relationships of visual words. The Space-LDA proposed by Li et al. [30] designed two regional attributes, namely “topic popularity” and “topic content”, to describe the spatial region information of high-spatial-resolution remote sensing images. Two attributes served as the potential topics of image patches and the prior distributions of image patches for each topic, thus constraining the generation process of image documents. When studying the microenvironment of biological cells, Chen et al. [31] found that LDA had difficulty reflecting the consistency of the microenvironment of neighboring cells and proposed the Spatial-LDA for biological cell studies. This model improved the inference process of LDA by introducing adjacent relationships between cells to affect prior probabilities that a cell belongs to a certain topic. When extracting and analyzing regional communities from social network data, Canh et al. [32] found that LDA failed to mine the available information about geographic location relationships between users, so they extended the SLDA [28]. In the extended model, messages containing the geographic locations of users were treated as visual words, regions were treated as documents, and a collection of postings containing the geolocations of users within specific time intervals were treated as images. It was applied successfully to extract the hidden regional communities from social network data.

Unfortunately, the aforementioned spatial topic models are not suitable for this study due to the differences in data types. The above studies improved topic models for specific domains and specific data to construct spatial topic models. The data utilized in those studies were image and network structure data, while the data utilized in our research are discrete point data. The differences in data types determine the variations in spatial feature extraction methods, as well as the diverse approaches to integrating spatial features into topic models. Therefore, the aforementioned spatial topic models cannot be directly applied to extract spatial semantic features for urban functional zones based on POIs. To address this issue, this study proposes to improve the LDA model and to build a spatial semantic feature extraction method for urban functional zones based on POIs.

This paper is organized as follows: Section 2 provides a detailed description of the spatial semantic feature extraction method for urban functional zones based on POIs. In Section 3, the proposed method is applied to extract the spatial semantic features of urban functional zones from simulated datasets. Then, the results are compared with the results of LDA to verify the effectiveness of the proposed method. In Section 4, the proposed method is applied to extract the spatial semantic features of urban functional zones from Chaoyang District, Beijing, and to classify the urban functional zones. The classification result was compared with the urban functional zone classification result of LDA to validate the accuracy of the proposed method. Section 5 discusses the advantages and limitations of the proposed method and presents the conclusions and future work.

2. Methodology

The core idea of the proposed method is to assume that POIs belonging to the same semantic feature are spatially correlated, and the spatial features of semantic features are reflected in the spatial separation of POIs. And, at the same time, to reduce the complexity and computation time, the proposed method ignores the spatial relationship of POIs within the same semantic feature. Based on the core idea, we improve the document generation process and the parameter inference process of LDA. In the proposed method, the urban functional zones are treated as documents and POI types are treated as words. In the generation process of urban functional zones, POIs belonging to the same semantic feature within the urban functional zones are put into the same bag-of-words, and the spatial relationship of POIs within the same semantic feature are ignored; POIs belonging to different semantic features are put into different bags-of-words, and the spatial features of the semantics are represented by the separation relationship of the bags-of-words. In the parameter inference process, the topic probabilities of a given POI are updated using the topic probabilities of POIs that are spatially clustered with the specified POI. This ensures spatial separation of the semantic features. The generation process of urban functional zones and the parameter inference process in the proposed method are described in detail in the following sections.

2.1. The Generation Process of Urban Functional Zones

LDA is a model that generates topics for documents. It is also called a three-level Bayesian probabilistic model, consisting of a structure of documents–topics–words. In LDA, the topic distribution of each document and the word distribution of each topic are first determined by two Dirichlet priors. Then, the following process is looped to generate all the words in each document: a topic is selected from the topic list according to the topic probability distribution of the document, and then, according to the word probability distribution of the topic, a word is selected from the word list. In LDA, the location of each word in a document is not generated, and a document is considered as an unordered bag-of-words.

In the proposed method, urban functional zones are treated as documents and POIs as words, the functional topic (functional semantic) distribution of each urban functional zone and the POI distribution of each topic are first determined by two Dirichlet priors. Then, the following process is looped to generate all the POIs in each urban functional zone: a topic is selected from the topic list according to the topic probability distribution of the urban functional zone, and then, according to the word probability distribution of the topic, a POI is selected from the word list. During the loop execution, POIs of the same topic within an urban functional zone are put into the same bag-of-words. The proposed method does not generate the locations of all bags-of-words but uses the separation relationship of the bags-of-words within an urban functional zone to represent the spatial features of the semantics. At the same time, the proposed method does not generate the spatial locations of the POIs within a bag-of-words and ignores the spatial features of POIs within the bag-of-words. Different from LDA, in the proposed method, the spatial features of POIs within an urban functional zone are represented by the spatial separation of bags-of-words, and an urban functional zone is not considered as an unordered bag-of-words.

The Bayesian network graph of the proposed method is shown in Figure 2. The urban functional zone is denoted as $t, t \in \{1, 2, \dots, T\}$, where T is the quantity of urban functional zones in the corpus. Topic is denoted as $k, k \in \{1, 2, \dots, K\}$, where K is the quantity of topics in the corpus. N is the quantity of POIs in the corpus. The urban functional zone t has D_t bags-of-words, the bag-of-words d in the urban functional zone t is denoted as $Doc(t, d)$, and there are $N_{Doc(t, d)}$ POIs in $Doc(t, d)$. $W_{t,d,n}$ denotes the POI n of the bag-of-words d in the urban functional zone t . α and β are Dirichlet prior hyperparameters. $\theta_{t,d}$ is the topic distribution of the bag-of-words d in the urban functional zone t , $Z_{t,d,n}$ is the topic of the POI n of the bag-of-words d in the urban functional zone t , and $\varphi_{Z_{t,d,n}}$ is the POI distribution of the topic $Z_{t,d,n}$. The following are detailed steps of the generation process of urban functional zones in the proposed method:

- (1) For a topic k , a multinomial parameter φ_k is sampled from a Dirichlet prior $\varphi_k \sim Dirichlet(\beta)$.
- (2) For a bag-of-words d in the urban functional zone t , a multinomial parameter $\theta_{t,d}$ over the K topics is sampled from a Dirichlet prior $\theta_{t,d} \sim Dirichlet(\alpha)$.
- (3) For a POI n of the bag-of-words d in the urban functional zone t , a topic label $Z_{t,d,n}$ is sampled from the multinomial distribution $Z_{t,d,n} \sim Multinomial(\theta_{t,d})$.
- (4) The POI $W_{t,d,n}$ of the bag-of-words d in the urban functional zone t is sampled from the multinomial distribution of topic $Z_{t,d,n}$, $W_{t,d,n} \sim Multinomial(\varphi_{Z_{t,d,n}})$.

Thus, the joint distribution of all visible variables and the hidden variables in the proposed method is as follows:

$$p(\theta, \varphi, z, w | \alpha, \beta) = \prod_{d=1}^{D_t} p(\theta_{t,d} | \alpha) \prod_{k=1}^K p(\varphi_k | \beta) \prod_{n=1}^{N_{Doc(t,d)}} p(Z_{t,d,n} | \theta_{t,d}) p(W_{t,d,n} | \varphi_{Z_{t,d,n}}) \quad (1)$$

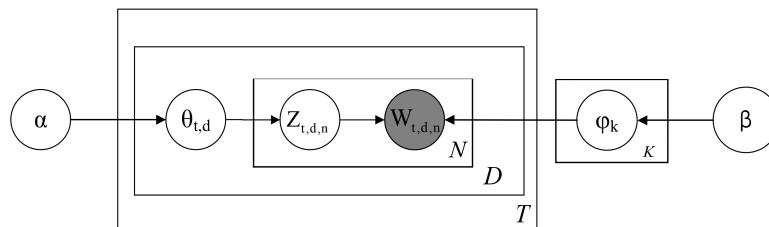


Figure 2. Bayesian network diagram of the proposed method.

2.2. The Parameter Inference Process

In this paper, the hidden variables θ and φ in the proposed method are inferred by Gibbs sampling. According to the principle of Gibbs sampling and the generation process of urban functional zones in Section 2.1, the parameter inference process of the proposed method is designed as follows. Firstly, a topic for each POI is randomly assigned. Then, for each POI, the probability of belonging to each topic is calculated based on the proportion of current POI types in each topic and the proportion of POIs in each topic within the spatial proximity range of the current POI; the roulette wheel method is then used to update the topic of the POI; and this process iterates for each POI until a termination condition is met. Finally, the semantic features of the urban functional zones are computed. In the parameter inference process of LDA, the probability of the current POI belonging to each topic is updated based on the proportion of POIs in each topic within the urban functional zone where the current POI is located. Different from LDA, the proposed method updates the probability of the current POI belonging to each topic based on the proportion of POIs in each topic within the spatial proximity range of the current POI. This improvement ensures that the topics of POIs are only related to those within their spatial proximity range, thereby guaranteeing that the POIs belonging to the same semantic feature are spatially correlated.

Detailed steps of the parameter inference process of the proposed method are as follows:

(1) Initialize the topics of all POIs: Randomly assign a topic for each POI in urban functional zones, that is $Z_{t,d,n} = k \sim \text{Multnomial}(1/K)$.

(2) Cluster POIs within each urban functional zone: Due to the different spatial distribution patterns of POIs within each urban functional zone, a uniform number of clusters is unable to be set for all urban functional zones. Therefore, an algorithm that does not need to pre-specify the number of clusters is needed to perform spatial clustering for POIs within each urban functional zone. In the proposed method, the Agglomerative Clustering algorithm is employed to aggregate POIs within a functional zone into multiple clusters, with each cluster representing the spatial proximity range of its internal POIs. It is important to note that the clusters mentioned in this section differ from the concept of “bags-of-words” in Section 2.1. In Section 2.1, POIs belonging to the same topic within an urban functional zone are put into the same “bag-of-words.” Therefore, POIs within a bag-of-words belong to a single topic. However, POIs belonging to different topics may be clustered into the same cluster due to spatial proximity range. Consequently, a cluster may contain one or more bags-of-words.

(3) Perform iterative sampling to update the topics of POIs: The probability values for each POI belonging to all topics are calculated using the following formula:

$$p(Z_{t,d,n} = k | Z_{-(t,d,n)}, w) \propto \frac{n_{k,-(t,d,n)}^{(v)} + \beta_v}{\sum_{v=1}^V (n_{k,-(t,d,n)}^{(v)} + \beta_v)} \cdot \frac{n_{t,c,-(t,d,n)}^{(k)} + \alpha_k}{\sum_{k=1}^K (n_{t,c,-(t,d,n)}^{(k)} + \alpha_k)} \quad (2)$$

where the corpus contains V POI types, $n_{k,-(t,d,n)}^{(v)}$ is the quantity of POI types v for topic k in the corpus (excluding $W_{t,d,n}$), $W_{t,d,n}$ belongs to cluster c of the urban functional zone t , and $n_{t,c,-(t,d,n)}^{(k)}$ is the quantity of POIs of topic k in cluster c of the urban functional zone t (excluding $W_{t,d,n}$).

Then, the roulette wheel method is employed to generate a random number. This random number is then combined with the obtained probability values to assign a new topic for each POI.

(4) Repeat step (3) until the termination condition is met: Upon reaching the specified number of iterations, the topics assigned to all POIs within each urban functional zone are utilized to compute the topic distributions of each cluster and the POI distribution of each topic. The calculation formulas are as follows:

$$\varphi_{k,v} = \frac{n_k^{(v)} + \beta_v}{\sum_{v=1}^V (n_k^{(v)} + \beta_v)} \quad (3)$$

$$\theta_{t,c,k} = \frac{n_{t,c}^{(k)} + \alpha_k}{\sum_{k=1}^K (n_{t,c}^{(k)} + \alpha_k)} \quad (4)$$

(5) Calculate the semantic features of urban functional zones: The semantic features of an urban functional zone are computed using the arithmetic mean of the topic distributions of all clusters within the urban functional zone. This feature considers the spatial separation of different semantic features within the urban functional zone. It represents the spatial semantic features of the urban functional zone.

3. Experiments Using Simulated Datasets

To demonstrate the effectiveness and superiority of the proposed method, the proposed method is used to extract spatial semantic features of urban functional zones from simulated datasets. The performance of the proposed method is compared with that of LDA. In this paper, we generate three datasets, respectively, containing urban functional zones with different spatial distribution patterns of POIs, urban functional zones

with different spatial locations of POIs, and urban functional zones with different quantities of POIs between topics. The two methods are implemented in Python 3.8 with the gensim.models.ldamodel package; the proposed method also uses the agglomerativeClustering method from the sklearn.cluster package. Some of the experiments require the SVM algorithm, so the SVM classifier was implemented using the SVC method in the sklearn.svm package.

3.1. Generation of Simulated Datasets

Simulation datasets were generated from a 20×20 km area according to the following steps:

(1) First, the area is divided into 400 1×1 km zones. Then, for each zone, two rectangles, Rec1 and Rec2, are generated, centered on Point1 and Point2, and the relative positions of Point1, and Point2 are set as $(0.25, 0.75)$ and $(0.75, 0.25)$. The area of the rectangles can be controlled by two parameters (rec_length , rec_width). Figure 3 shows the two rectangles generated for a zone.

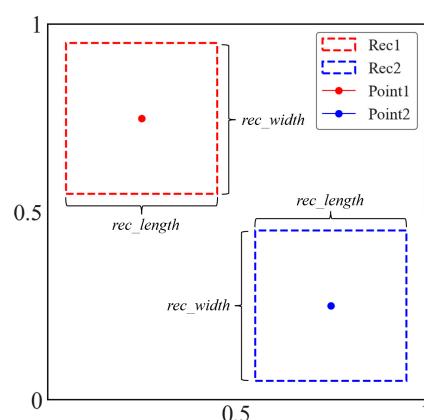


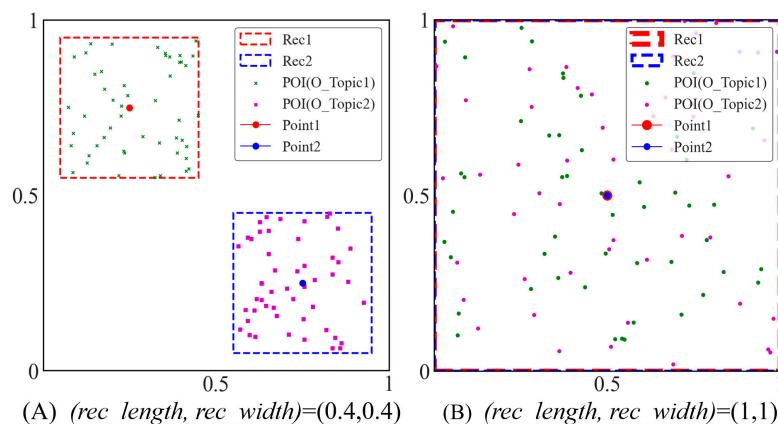
Figure 3. Two rectangles generated for a zone.

(2) POIs are generated within the rectangles of each zone. For each zone, the quantity of POIs in Rec1 and Rec2 is controlled by parameters $rec_poiNum1$ and $rec_poiNum2$, the spatial positions of POIs within the rectangles are randomly generated, and the POI types are selected based on the word probability distribution from predefined O_Topic1 and O_Topic2 (shown as Table 1). The percentage of topics for POIs within the two rectangles is controlled by parameter $topic1_percentage$. Specifically, if $topic1_percentage$ is set to 0.4, the proportions of POIs belonging to O_Topic1 are 40% in Rec1 and 60% in Rec2, the proportions of POIs belonging to O_Topic2 are 60% in Rec1 and 40% in Rec2, and so forth.

To generate zones with different spatial distribution patterns of POIs, a group of five datasets, named Dataset1, is generated. Each dataset contains two categories of urban functional zones with different spatial distribution patterns of POIs. In each dataset, $rec_poiNum1$ and $rec_poiNum2$ of all zones are set to 50, and $topic1_percentage$ is set to 1 (meaning Rec1 only contains POIs of O_Topic1, and Rec2 only contains POIs of O_Topic2). The zone where (rec_length, rec_width) are set to $(0.4, 0.4)$ is considered to be an urban functional zone with a cluster distribution pattern of POIs, as illustrated in Figure 4A. The zone where Point1, Point2 are all changed to $(0.5, 0.5)$ and (rec_length, rec_width) are set to $(1, 1)$ is considered to be an urban functional zone with a mixed distribution pattern of POIs, as shown in Figure 4B. The proportion of these two categories of urban functional zones can be adjusted by a parameter p , and p takes values of 0.1, 0.2, 0.3, 0.4, and 0.5 for the five datasets, respectively. Specifically, when $p = 0.1$, the proportion of the urban functional zone with a cluster distribution pattern of POIs is 0.1, and the proportion of the urban functional zone with a mixed distribution pattern of POIs is 0.9, and so forth.

Table 1. The probability distributions of the POI type of O_Topic1 and O_Topic2.

Topic	POI Type	Probability
O_Topic1 (Daily Life Service)	Travel Agency	0.3
	Ticket Office	0.2
	Post Office	0.1
	Logistics Service	0.1
	Telecom Office	0.05
	Lottery Store	0.05
	Job Center	0.05
	Repair Store	0.05
	Photo Finishing	0.05
	Laundry	0.05
O_Topic2 (Science/Culture and Education Service)	School	0.5
	Museum	0.1
	Exhibition Hall	0.05
	Art Gallery	0.05
	Library	0.05
	Training Institution	0.05
	Planetarium	0.05
	Archives Hall	0.05
	Cultural Palace	0.05
	Research Institution	0.05

**Figure 4.** Two categories of urban functional zones with different POI spatial distribution patterns.

To generate zones with heterogeneity of POIs, a group of five datasets, named Dataset2, is generated. Each dataset contains two categories of urban functional zones with different heterogeneity of POIs. In each dataset, (rec_length, rec_width) of all zones are set to $(4, 4)$, and $rec_poiNum1$ and $rec_poiNum2$ are set to 50. Because the locations of the POIs in each rectangle are randomly generated, the heterogeneity of POIs is reflected in the percentage of topics for POIs within the two rectangles. In each dataset, there are 200 urban functional zones with $topic1_percentage$ set to 1. The difference between these five datasets is that the $topic1_percentage$ values of the other 200 urban functional zones in each dataset correspond to five different settings: 0.5, 0.6, 0.7, 0.8 and 0.9. Figure 5A shows an urban functional zone with $topic1_pertange = 1$, and Figure 5B shows an urban functional zone with $topic1_pertange = 0.6$.

To generate zones with differences in the quantity of POIs between topics, a dataset, named Dataset3, is generated. In this dataset, $topic1_percentage$ of all the zones is set to 1, and (rec_length, rec_width) are set to $(4, 4)$; the 400 zones are equally divided into five parts. The total quantity of POIs in each urban functional zone is 100; then, the proportion of POIs can be changed only by changing the variation in $rec1_poiNum$, and the $rec1_poiNum$ of the urban functional zones in each part takes values of 10, 20, 30, 40, and 50. Figure 6 shows the two urban functional zones with $rec1_poiNum = 10$ and $rec1_poiNum = 50$.

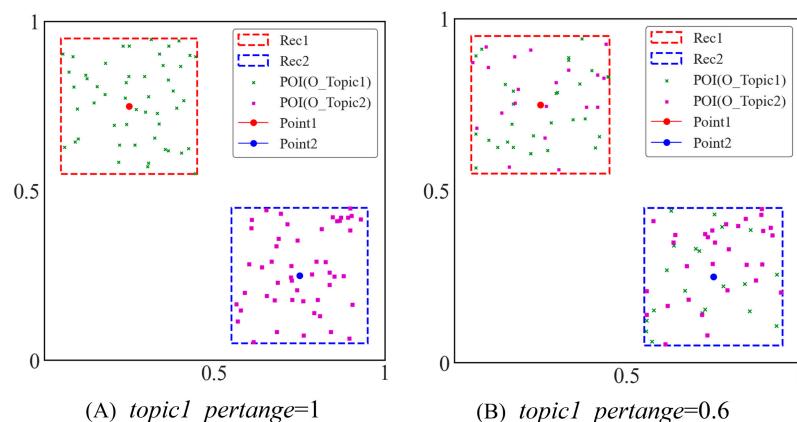


Figure 5. Two categories of urban functional zones with heterogeneity of POIs.

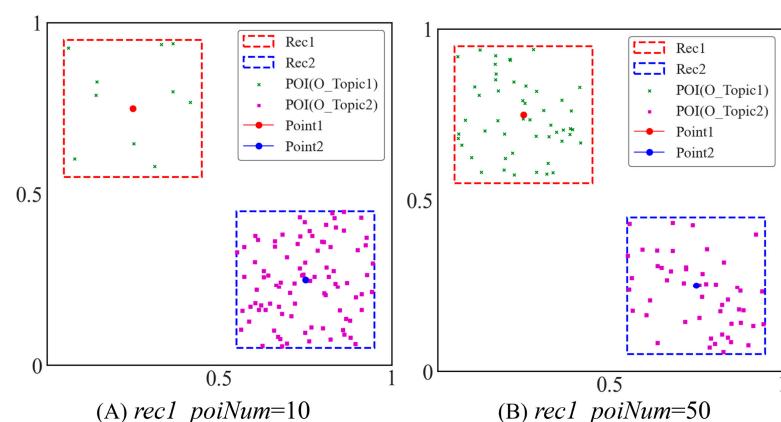


Figure 6. Two categories of urban functional zones with differences in the quantity of POIs between topics.

3.2. Parameter Settings

In the process of extracting topics from the simulated datasets using those two methods, the number of iterations of the parameters inference process is set to 500, the number of topics generated by the method is set to 5, and α and β are set to “auto” to allow the methods to automatically learn the optimal values.

In the proposed method, the different *distance_threshold* values of the Agglomerative Clustering algorithm may lead to different results; the optimal result is selected from the results with a *distance_threshold* value of 0.5~2 km as the final result.

3.3. Experimental Results

3.3.1. The Influence of Different Spatial Distribution Patterns of POIs on Topic Extraction

LDA and the proposed method were used to extract the semantic features of urban functional zones in Dataset1; then, the urban functional zones were classified using the extracted results and the Support Vector Machine (SVM) algorithm. Figure 7 shows the overall accuracy (OA) of the urban functional zone classification results using the two methods. It can be noted that at $p = 0.1$, the quantity of the urban functional zones with a mixed distribution pattern of POIs in the dataset is much smaller than the quantity of urban functional zones with a cluster distribution pattern of POIs; the OAs of the proposed method and LDA are approximate. As p increases, the quantity of urban functional zones in these two categories gradually approaches, the OAs of LDA gradually decreases, and the OAs of the proposed method are maintained at 100%. Those indicate that LDA fails to distinguish the urban functional zones with the same type and quantity of POIs but different spatial distribution patterns of POIs, but the proposed method performs well in this regard.

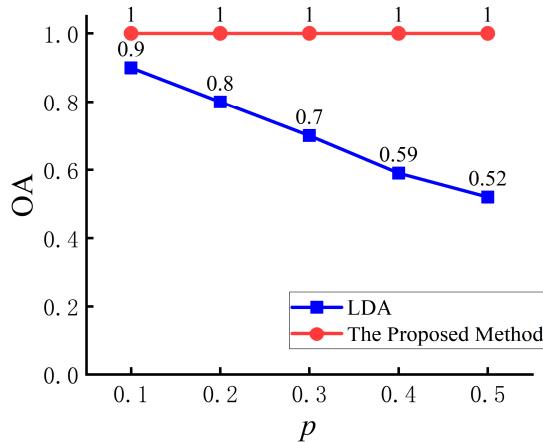


Figure 7. The OAs of the two method using Dataset1.

To further illustrate the superiority of the proposed method, the topics extracted from the dataset with $p = 0.5$ are utilized to analyze the performance of the two methods. Figure 8 shows the word distributions of the five topics obtained by LDA and the arithmetic mean of the topic distributions for all urban functional zones. Figure 9 shows the word distributions of the five topics obtained by the proposed method and the arithmetic mean of the topic distributions for all urban functional zones. In Figure 8F, it can be found that in the extraction results of LDA, all the urban functional zones only contain two topics, L_Topic1 and L_Topic4. Combining Figure 8B,E, we can see that the word distributions in these two topics are very similar; therefore, they can be regarded as the same topic and can be regarded as a mixed topic of O_Topic1 and O_Topic2 (seen in Section 3.1). In Figure 9, it can be found that in the extraction results of the proposed method, the urban functional zones mainly contain three topics (P_Topic0, P_Topic1, and P_Topic4), P_Topic0 can be regarded as O_Topic2, P_Topic1 can be regarded as O_Topic1, and P_Topic4 can be regarded as a mixed topic of O_Topic1 and O_Topic2. This shows that LDA cannot consider the influence of different spatial distribution patterns of POIs on topic extraction; the proposed method can consider the influence of different spatial distribution patterns of POIs on topic extraction.

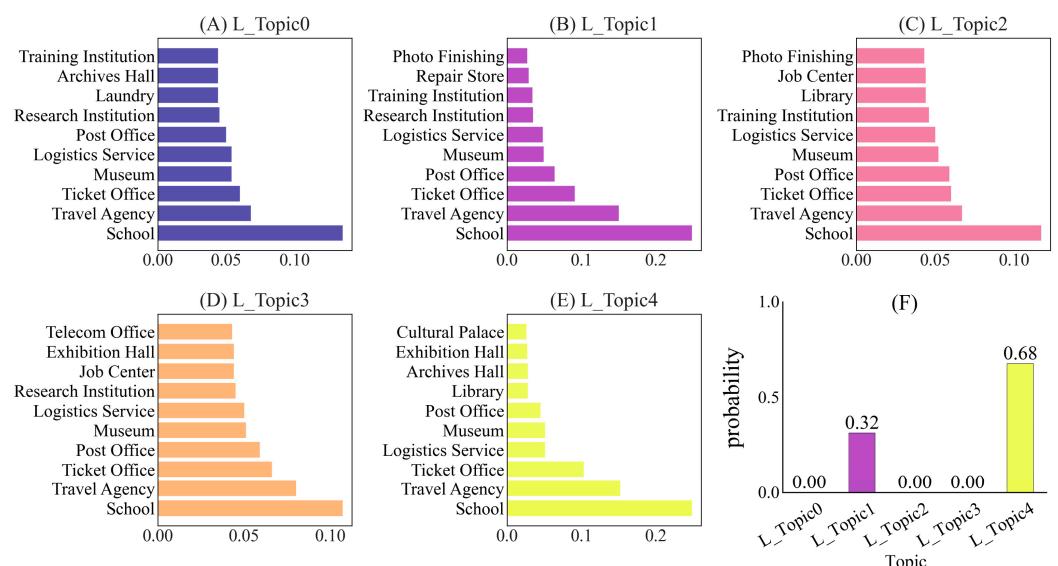


Figure 8. (A–E) show the word distributions of the five topics obtained by LDA from the dataset with $p = 0.5$; (F) shows the arithmetic mean of the topic distributions obtained by LDA from the dataset with $p = 0.5$.

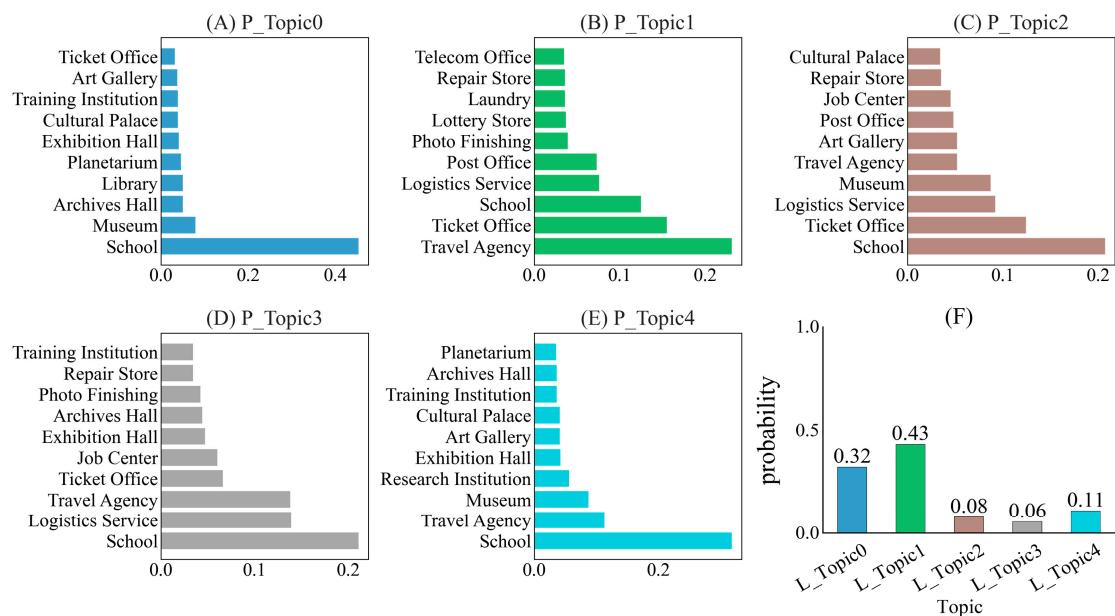


Figure 9. (A–E) show the word distributions of the five topics obtained by the proposed method from the dataset with $p = 0.5$; (F) shows the arithmetic mean of the topic distributions obtained by the proposed method from the dataset with $p = 0.5$.

3.3.2. The Influence of Different Heterogeneity of POIs on Topic Extraction

LDA and the proposed method were used to extract the semantic features of urban functional zones in Dataset2; then, the urban functional zones were classified using the extracted results and the SVM algorithm. Figure 10 shows the OAs of the two methods. For the five datasets with different values of *topic1_percentage*, the OAs of LDA are all 0.53, while the OAs of the proposed method are all between 0.98 and 1. Those indicate that LDA is unable to distinguish urban functional zones with the same types and quantities of POIs but heterogeneity of POIs, but the proposed method performs well in this regard.

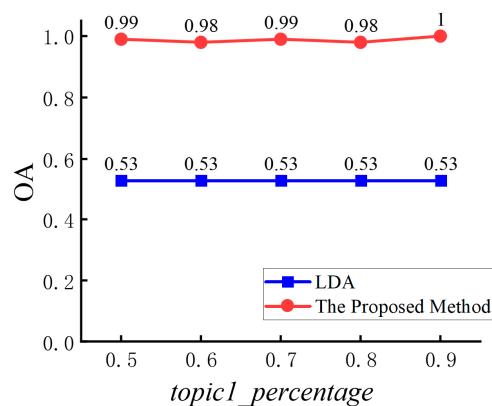


Figure 10. The OAs of the two method using Dataset2.

To further illustrate the superiority of the proposed method, the topics extracted from the dataset with *topic1_percentage* = 0.6 are utilized to analyze the performance of the two methods. Figure 11 shows the word distributions of the five topics obtained by LDA and the arithmetic mean of the topic distributions for all urban functional zones. Figure 12 shows the word distributions of the five topics obtained by the proposed method and the arithmetic mean of the topic distributions for all urban functional zones. In Figure 11F, it can be found that in the extraction results of LDA, all the urban functional zones only contain two topics, L_Topic1 and L_Topic4. Combining Figure 11B,E, we can see that the word distributions in these two topics are very similar; therefore, they can be regarded

as the same topic and can be regarded as a mixed topic of O_Topic1 and O_Topic2. In Figure 12, it can be found that in the extraction results of the proposed method, the urban functional zones mainly contain three topics (P_Topic0, P_Topic3, and P_Topic4), P_Topic0 can be regarded as O_Topic2, P_Topic3 can be regarded as O_Topic1, and P_Topic4 can be regarded as a mixed topic of O_Topic1 and O_Topic2. This shows that LDA cannot consider the influence of heterogeneity of POIs on topic extraction; however, the proposed method can consider the influence of heterogeneity of POIs on topic extraction.

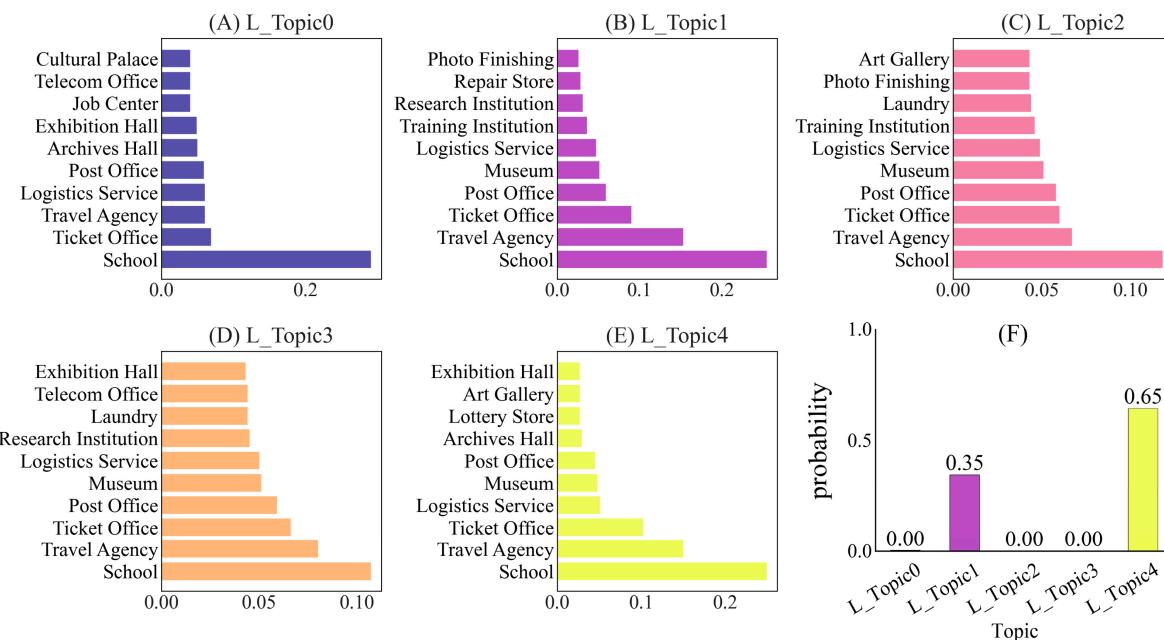


Figure 11. (A–E) show the word distributions of the five topics obtained by LDA from the dataset with $\text{topic1_percentage} = 0.6$; (F) shows the arithmetic mean of the topic distributions obtained by LDA from the dataset with $\text{topic1_percentage} = 0.6$.

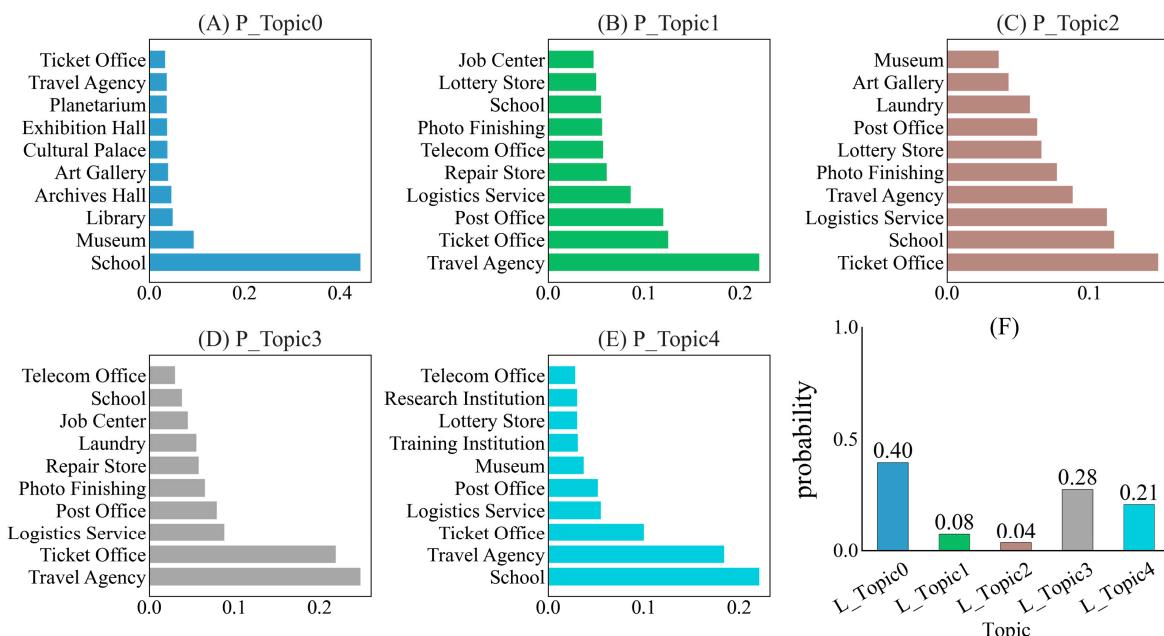


Figure 12. (A–E) show the word distributions of the five topics obtained by the proposed method from the dataset with $\text{topic1_percentage} = 0.6$; (F) shows the arithmetic mean of the topic distributions obtained by the proposed method from the dataset with $\text{topic1_percentage} = 0.6$.

3.3.3. The Influence of Different Quantities of POIs in Topics on Topic Extraction

The proposed method and LDA are applied on Dataset3 to extract topics of the urban functional zones. Figures 13 and 14 show the word distributions of topics and the arithmetic mean with different *rec1_poiNum* values of the topic distributions obtained by two methods. As illustrated in Figure 13F, under different *rec1_poiNum* values, there are actually only three topics (L_Topic1, L_Topic2, and L_Topic3) in the extracted results of the urban functional zones by LDA. From the type and percentage of POIs in the topics in Figure 13B–D, it can be seen that the three topics can be regarded as mixed topics of O_Topic1 and O_Topic2. L_Topic2 and L_Topic3 have small percentages of POI types that belong to O_Topic1, and they are similar and can be regarded as the same topic. L_Topic1 has a relatively large percentage of POI types that belong to O_Topic1, and it can be regarded as another topic. From Figure 13F, it can be seen that when *rec1_poiNum* is set to 10 and 20, indicating a lower proportion of O_Topic1 in the urban functional zones, the urban functional zones only contain L_Topic2 and L_Topic3; when *rec1_poiNum* is set to 30, there are three topics (L_Topic1, L_Topic2, and L_Topic3) in the urban functional zones; and as *rec1_poiNum* increases to 40 and 50, the urban functional zones mainly contain L_Topic1. That is, when the quantities of POIs in the topics change, the topic distribution of the urban functional zones extracted by LDA changes. As illustrated in Figure 14F, under different *rec1_poiNum* values, there are mainly three topics (P_Topic0, P_Topic1, and P_Topic4) in the extracted results of the urban functional zones by the proposed method. From the type and percentage of POIs in the topics in Figure 14A,B,E, it can be seen that P_Topic0 and P_Topic4 can be regarded as O_Topic2, and P_Topic1 can be regarded as O_Topic1. From Figure 14F, it can be seen that in all the urban functional zones with five different *rec1_poiNum* values, the proportion of P_Topic1 is close to 50%, and the proportion of P_Topic0 and P_Topic4 is close to 50%; that is, no matter how the quantities of POIs in topics changes, the topic distribution of urban functional zones extracted by the proposed method remains unchanged. This shows that LDA cannot consider the influence of different quantities of POIs in topics on topic extraction; however, the proposed method can consider the influence of different quantities of POIs in topics on topic extraction.

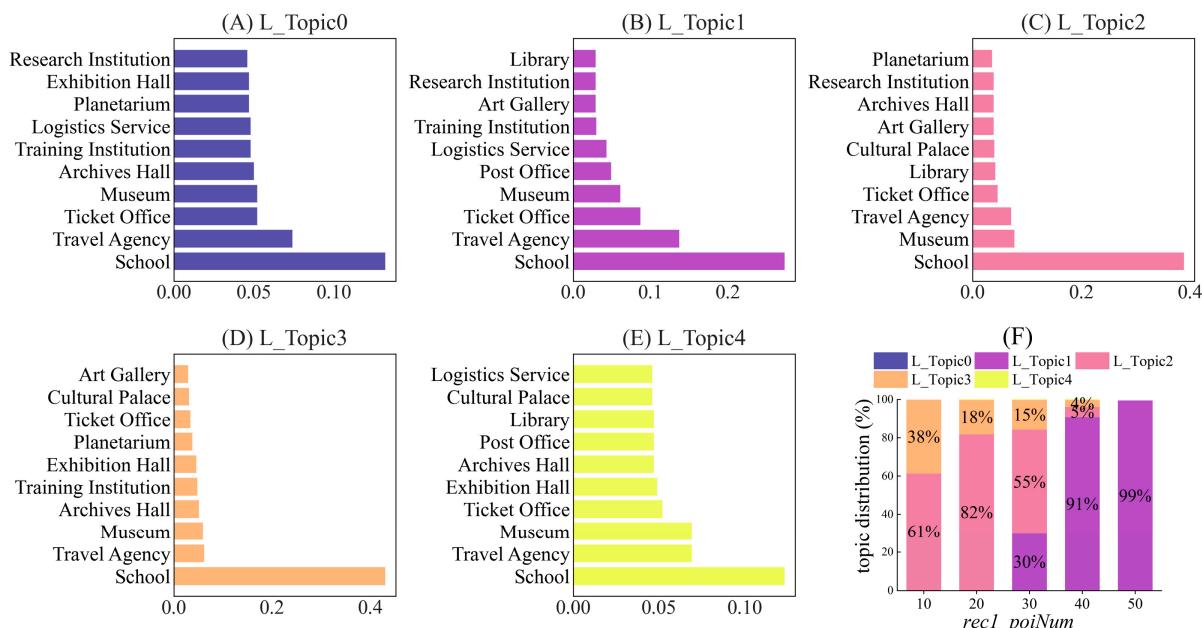


Figure 13. (A–E) show the word distributions of the five topics obtained by LDA from Dataset3; (F) show the arithmetic mean with different *rec1_poiNum* values of the topic distributions obtained by LDA from Dataset3.

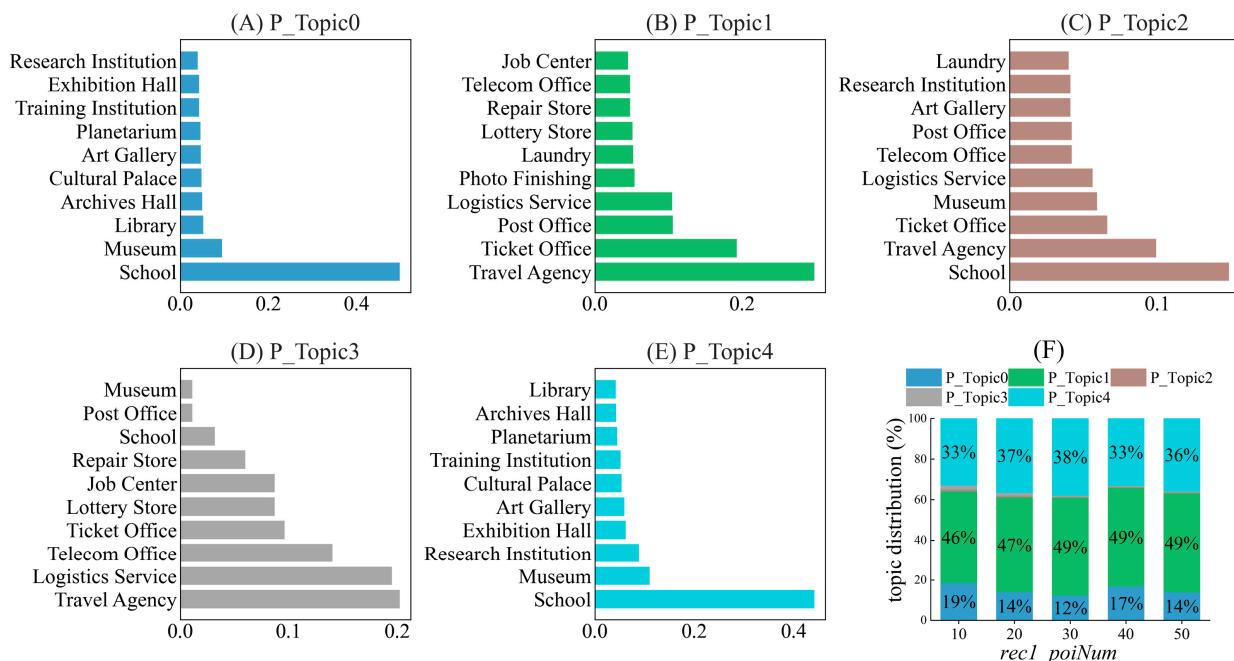


Figure 14. (A–E) show the word distributions of the five topics obtained by the proposed method from Dataset3; (F) show the arithmetic mean with different *rec1_poiNum* values of the topic distributions obtained by the proposed method from Dataset3.

4. Case Study Using a Chaoyang POI Dataset

In order to verify the performance of the proposed method more sufficiently, we selected a real city as the study area for a case study. We first extracted the semantic information of urban function zones in Chaoyang District using the proposed method. The SVM algorithm has the characteristics of strong adaptability, fast learning speed, and limited requirements on sample sizes [33,34], and it is often used in the field of urban functional zone classification [9,20,23], so we used the SVM algorithm combined with the extracted semantic information to classify the urban functional zones in the study area. Finally, the LDA was used as a comparison of the proposed method.

4.1. Study Area and Data

The study area selected for the case study was Chaoyang District, situated in the south-central part of Beijing. As one of the city's six main districts, Chaoyang District boasts a high population density, a thriving economy, and an extensive road network. It is adjacent to the central urban area, with the Dongcheng and Xicheng Districts on the west; Haidian District, where Beijing's high-tech industrial base and university cluster, to the northwest; the Shunyi and Changping Districts, which have developed science and technology industries and beautiful natural environments, on the north; the subsidiary-center Tongzhou District on the east; and the Fengtai and Daxing Districts, where traditional industry coexists with modernization, on the south. Thus, the urban functional zones within Chaoyang District show a complex structure composed of various categories. In addition, Chaoyang City has rich functional types of urban functional zones such as commercial centers, diplomatic hubs, residential communities, and areas dedicated to scientific research and education.

The traffic analysis zone (TAZ) served as a basic unit of urban functional zones in this study, and the study area was divided into 592 TAZs by merging the administrative area boundary data (<http://www.gadm.org/>, accessed on 17 April 2020) with the first three levels of road network data, as shown in Figure 15. The road network data of Chaoyang District was download from Open Street Map (<https://www.openstreetmap.org/>, accessed on 17 April 2020) in 2020. OSM is an open-source map provider that aims to provide users with free and easily accessible digital map resources and is considered to be the most successful and prevailing volunteered geographic information at this stage [35].

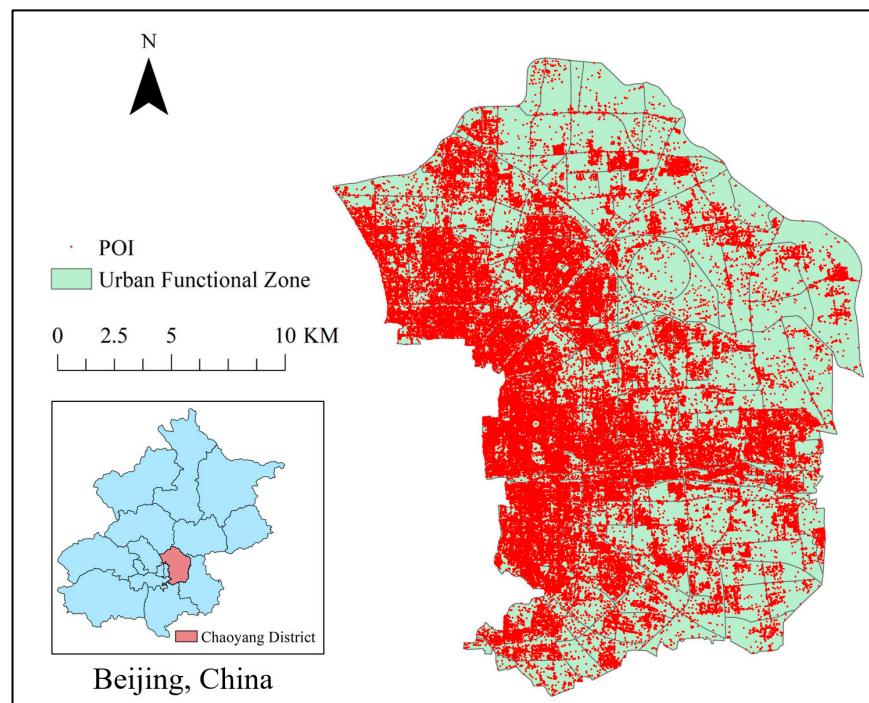


Figure 15. POIs and urban functional zones in Chaoyang District.

The POI dataset of Chaoyang District was acquired in April 2020 through the API provided by the Gaode Mapping Service (<https://www.amap.com/>, accessed on 17 April 2020); the dataset consists of 177,164 POIs. In this dataset, there are 26 big categories and 906 subcategories; the subcategory were selected as the POI types. Figure 15 shows the POIs in Chaoyang District.

In addition, urban planning and land use planning data (from the Chaoyang District People's Government of Beijing Municipality, <http://www.bjchy.gov.cn/>, accessed on 17 April 2020), high-resolution remote sensing images (<https://www.gscloud.cn/>, accessed on 17 April 2020), and POIs were used to annotate 592 urban functional zones in Chaoyang District. Figure 16 shows the classification result map of Chaoyang District annotated by volunteers with urban planning background knowledge based on these data. In this study, the manually interpretation results served as the actual urban functional zone map.

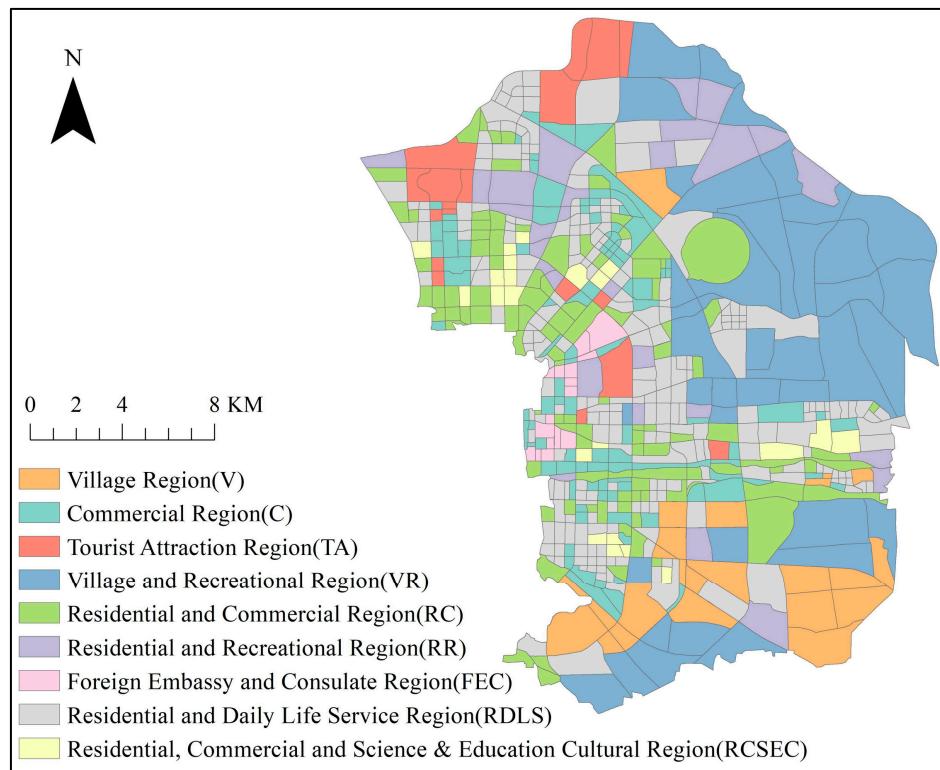


Figure 16. The urban functional zone map annotated by manual interpretation.

4.2. Experiment and Result Analysis

4.2.1. Parameter Settings

In the proposed method, the number of iterations of the parameter inference process was set to 500, and the number of topics generated by the method was set to 200. The parameters α and β were set to “auto”. Furthermore, it is worth noting that different values of the *distance_threshold* in the Agglomerative Clustering algorithm will produce different results.

For this study, the *distance_threshold* value was explored within the range of 0 to 0.1 degrees latitude and longitude (equivalent to approximately 0 to 11.112 km), with an increment of 0.005 degrees (approximately 0.556 km). Figure 17 shows the classification accuracy with different *distance_threshold* values. It can be observed that the overall accuracy (OA) was highest when the *distance_threshold* value was set to 0.02 degrees (approximately 2.224 km). Therefore, 0.02 degrees (approximately 2.224 km) was selected as the optimal *distance_threshold* value for further analysis and validation. The SVM algorithm uses the Radial Basis Function (RBF) kernel because the RBF function provides good flexibility and performance in nonlinear problems [36]. When training the SVM with the RBF kernel, two parameters must be considered: C and γ . We set $C = [2^{-10}, 2^{10}]$ and $\gamma \in [0, 1]$, and searched the optimum parameters using a grid-search method; the optimization objective was to maximize Kappa. In total, 60% of the data was randomly selected as the training set, and the remaining 40% was used as the test set.

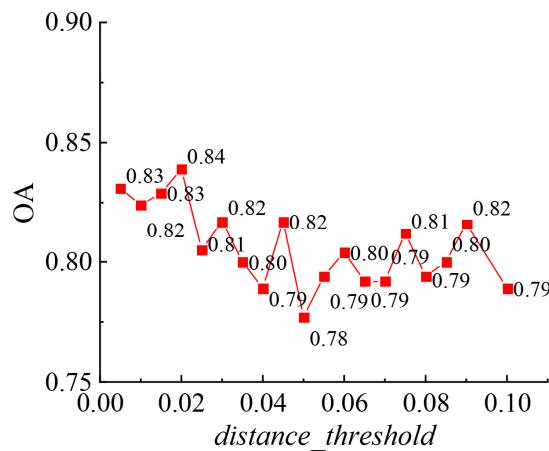


Figure 17. The OA of the proposed method for different *distance_threshold*.

4.2.2. Experimental Result Analysis

The results map of the urban functional zone classification of Chaoyang District using the proposed method is shown in Figure 18. As can be seen from Figure 18, in the western area, which is adjacent to the Dongcheng and Xicheng districts, the main types of functional zones distributed are Commercial (C); Residential and Commercial (RC); Residential and Daily Life Service (RDLS); and Residential, Commercial, and Science and Education and Cultural (RCSEC) regions. Additionally, a Foreign Embassy and Consulate (FEC) region is also situated in the western part of Chaoyang District. In the eastern part of Chaoyang District, the landscape is primarily characterized by a Village and Recreational (VR) region. The northern part mainly consists of a Tourist Attraction (TA) region and a Residential and Recreational (RR) region. And, the southern part mainly has a distribution of Village (V), and Village and Recreational (VR) regions. This classification is generally consistent with the actual distribution and actual geographical location of urban functional zones in Chaoyang District.

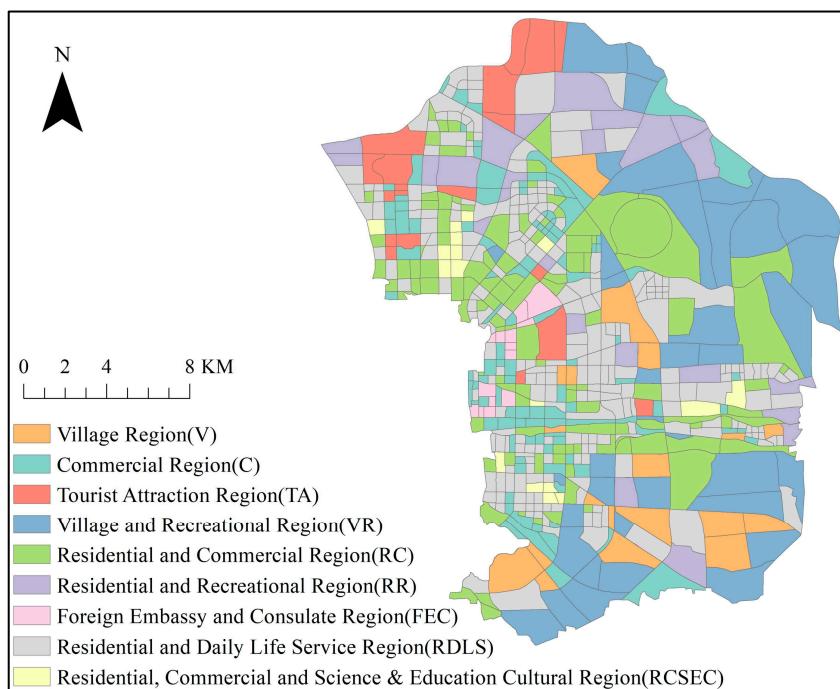


Figure 18. The results map of the urban functional zone classification of Chaoyang District using the proposed method.

Figure 19A demonstrates the percentage confusion matrix for the classification results of the proposed method. It shows the classification performance of the method on different categories of urban functional zones. The bolded values on the diagonal of the rectangle are the percentage of correct classifications. From the diagonal elements, this proposed method has the highest classification accuracy of 92% in identifying the Residential and Daily Life Service (RDLS) region. It also performs well in classifying Residential and Commercial (RC), Commercial (C), Tourist Attraction (TA), and Foreign Embassy and Consulate (FEC) regions, with classification accuracies exceeding 80%. The classification accuracy of the Village and Recreational (VR) and Residential, Commercial, and Science and Education Cultural (RCSEC) regions is between 70% and 80%. However, the classification accuracy for the Village (V) and Residential and Recreational (RR) regions is relatively low, both at 61%. In these two categories of relatively poorly categorized urban functional zones: in the Village (V) region, 22% are misclassified as a Village and Recreational (VR) region, 13% as a Commercial (C) region, and the remaining 4% as a Residential and Daily Life Service (RDLS) region; in the Residential and Recreational (RR) region, 14% are misclassified as a Residential and Daily Life Service (RDLS) region, while 7% are misclassified as a Village and Recreational (VR) region, 7% as a Commercial (C) region, 4% as a Residential and Commercial (RC) region, and 4% as a Tourist Attraction (TA) region. The poor classification accuracy for Village (V) and Residential and Recreational (RR) regions can be attributed to the high similarity in their spatial features and POI statistics with other.

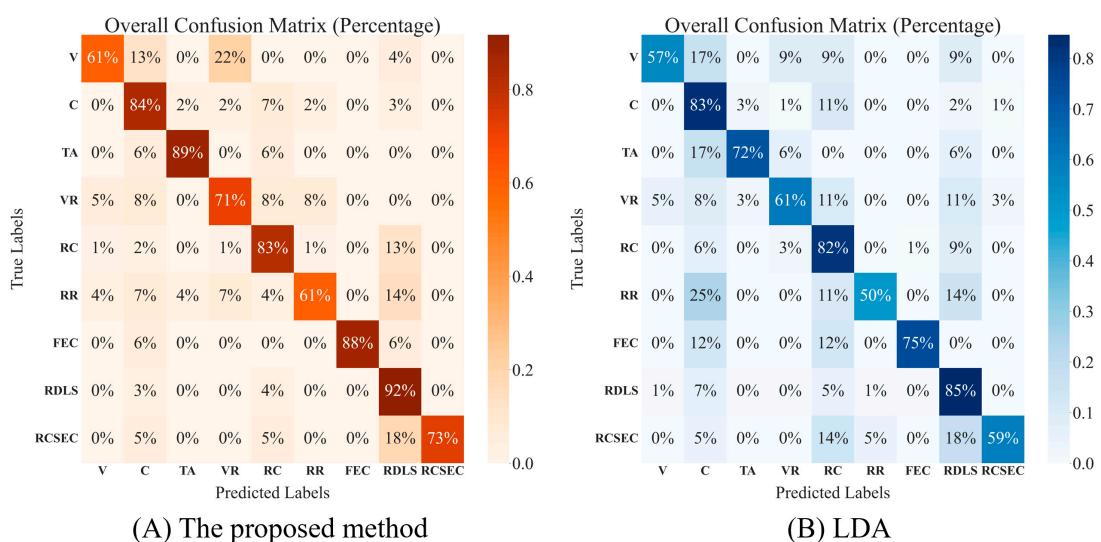


Figure 19. Heat map of classification confusion matrix.

4.3. Comparison

Due to the proposed method being an improvement of the LDA, in order to validate our hypothesis that, when classifying urban functional zones, considering the spatial features of semantic information can improve the accuracy of the classification, we conducted an experiment using the LDA method as a comparison method. To avoid the influence of different parameter settings on the classification results, the parameters in LDA were kept consistent with the corresponding parameters in the proposed method. Only observing the extracted semantic features of the two methods does not help us to intuitively infer that the proposed method is superior to LDA on real datasets, so we needed to choose some evaluation metrics to assess the performance of the two methods. We chose the overall accuracy (OA), kappa coefficient, and confusion matrix as the evaluation metrics of the results of the two methods.

Figure 20 shows the results map of the urban functional zone classification of Chaoyang District using LDA.

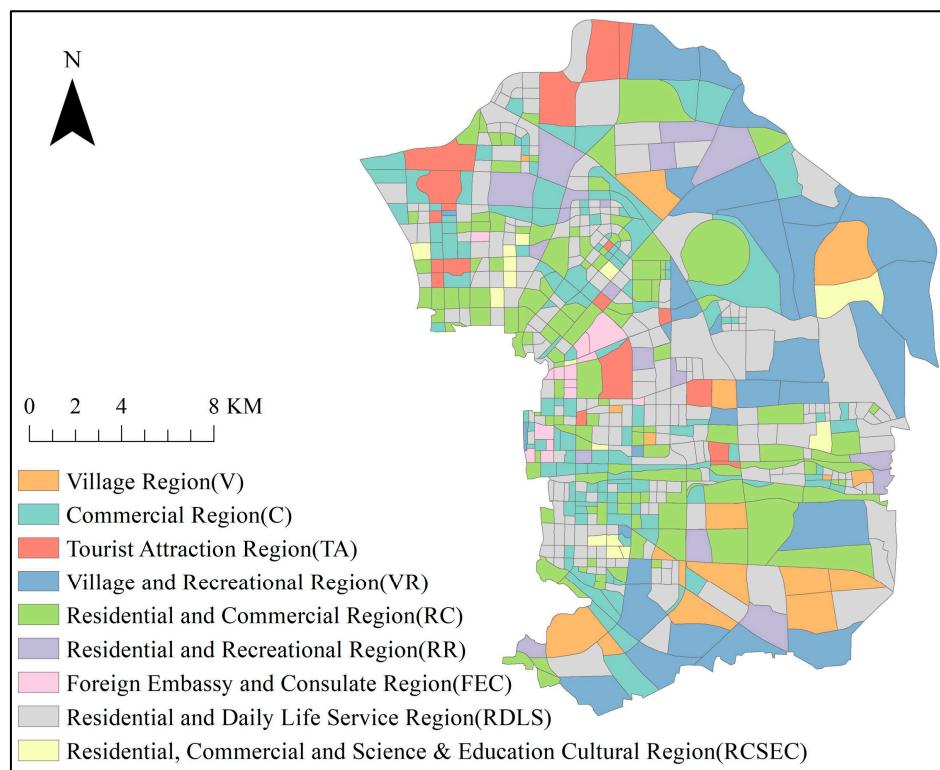


Figure 20. The results map of the urban functional zone classification of Chaoyang District using LDA.

Table 2 presents the overall accuracy (OA) and kappa coefficient obtained through the SVM classification of the semantic information extracted by the two methods on the Chaoyang District dataset. From Table 2, it can be seen that the OA of the proposed method is improved by 6% and the Kappa coefficient by 8% compared to LDA.

Table 2. Comparison of evaluation indexes between LDA and the proposed method.

Method	Overall Accuracy (OA)	Kappa Coefficient
LDA	0.78	0.71
The proposed method	0.84	0.79

Figure 19 illustrates a comparison between the confusion matrix heatmaps of the LDA and the proposed method. From the comparison, it is evident that the proposed method consistently outperforms the LDA in terms of classification accuracy across various urban functional zone categories. For single urban functional zones, the classification accuracies of the Village (V), Commercial (C), Tourist Attraction (TA), and Foreign Embassy and Consulate (FEC) regions are, respectively, improved by 4%, 1%, 17%, and 13% compared to LDA when the proposed method was applied. Overall, the classification accuracy for single urban functional zones improved by approximately 8.8% using the proposed method. For mixed urban functional zones, the classification accuracies of the Village and Recreational (VR); Residential and Daily Life Service (RDLS); Residential and Commercial (RC); Residential and Recreational (RR); and Residential, Commercial, and Science and Education Cultural (RCSEC) regions are, respectively, improved by 10%, 7%, 1%, 11%, and 14% compared to LDA when the proposed method was applied. Overall, the classification accuracy for mixed urban functional zones improved by approximately 8.6% using the proposed method.

5. Discussion and Conclusions

Extracting accurate semantic features of urban functional zones is important for understanding urban functional zones and exploring urban spatial structure. It is useful and convenient to extract the semantics of urban functional zones using topic models and POIs. However, topic models can only extract statistical information about POIs and ignore spatial information about POIs. Therefore, the extracted semantic features are incomplete. In this paper, we improve the LDA model (a typical topic model) and propose a novel method to extract the spatial semantic features of urban functional zones from POI data.

The proposed method is applied to simulated datasets and a real case study. Experimental results on the simulated datasets show that the proposed method effectively considers the influence of different spatial distribution patterns of POIs, the heterogeneity of POIs, and different quantities of POIs in topics on topic extraction. This indicates that the proposed method successfully considers the spatial features of POIs within urban functional zones. Experimental results on the Chaoyang POI dataset show that in terms of urban functional zone classification, the spatial semantic features obtained by the proposed method are more accurate than the semantic features obtained by LDA.

However, the proposed method can be further improved. The method treats the spatial proximity range as a bag-of-POIs and limits the extraction of semantic features to each bag-of-POIs. This will lead to two problems: (1) ignoring the impact of the spatial distances among POIs in a bag-of-POIs on semantic features; (2) POIs in different bags-of-POIs within an urban functional zone cannot completely belong to the same semantic features. According to common sense, this is unreasonable. Therefore, this method does not fully consider the impact of spatial distances among POIs on semantic features. In further research, to overcome this shortcoming, we will establish a novel spatial semantic feature extraction method for urban functional zones based on POIs. In the novel method, the extraction of topics is not limited in spatial proximity range but in an entire urban functional zone. The distance between POIs is used to determine the probability of them belonging to the same topic. The POIs with a closer spatial distance have a higher probability of belonging to the same topic, and the POIs with further spatial distances have a lower probability of belonging to the same topic.

Author Contributions: Conceptualization, Xin Yang; methodology, Xin Yang and Xi'ang Ma; software, Xin Yang and Xi'ang Ma; validation, Xi'ang Ma; formal analysis, Xi'ang Ma; investigation, Xi'ang Ma; resources, Xin Yang; data curation, Xi'ang Ma; writing—original draft preparation, Xin Yang and Xi'ang Ma; writing—review and editing, Xin Yang and Xi'ang Ma; visualization, Xi'ang Ma; supervision, Xin Yang; project administration, Xin Yang; funding acquisition, Xin Yang. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42201506, and the Visiting Research Fund for Teachers of General Undergraduate Universities in Shandong Province.

Data Availability Statement: The original data and codes that support the findings of this study are openly available at https://github.com/maxiang98/UFZs_POIs_Semantic-Feature-Extraction (accessed on 17 June 2024).

Acknowledgments: We are particularly grateful to the academic editors and all reviewers for their critical comments or suggestions, which have had a significant impact on improving the quality of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [[CrossRef](#)]
- Du, S.; Du, S.; Liu, B.; Zhang, X. Mapping large-scale and fine-grained urban functional zones from VHR images using a multi-scale semantic segmentation network and object based approach. *Remote Sens. Environ.* **2021**, *261*, 112480. [[CrossRef](#)]

3. 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](#)]
4. Xue, B.; Xiao, X.; Li, J.; Zhao, B.; Fu, B. Multi-source Data-driven Identification of Urban Functional Areas: A Case of Shenyang, China. *Chin. Geogr. Sci.* **2022**, *33*, 21–35. [[CrossRef](#)]
5. Wang, Y.; Gu, Y.; Dou, M.; Qiao, M. Using Spatial Semantics and Interactions to Identify Urban Functional Regions. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 130. [[CrossRef](#)]
6. Zhang, X.; Du, S.; Wang, Q. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 170–184. [[CrossRef](#)]
7. Bao, H.; Ming, D.; Guo, Y.; Zhang, K.; Zhou, K.; Du, S. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data. *Remote Sens.* **2020**, *12*, 1088. [[CrossRef](#)]
8. Zhong, Y.; Zhu, Q.; Zhang, L. Scene Classification Based on the Multifeature Fusion Probabilistic Topic Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6207–6222. [[CrossRef](#)]
9. Liu, X.; He, J.; Yao, Y.; Zhang, J.; Liang, H.; Wang, H.; Hong, Y. Classifying urban land use by integrating remote sensing and social media data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1675–1696. [[CrossRef](#)]
10. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [[CrossRef](#)]
11. Bratasanu, D.; Nedelcu, I.; Datcu, M. Bridging the Semantic Gap for Satellite Image Annotation and Automatic Mapping Applications. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *4*, 193–204. [[CrossRef](#)]
12. 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](#)]
13. 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.* **2016**, *31*, 825–848. [[CrossRef](#)]
14. Xiao, Y.; Chen, X.; Li, Q.; Yu, X.; Chen, J.; Guo, J. Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 358. [[CrossRef](#)]
15. Hu, Y.; Han, Y. 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](#)]
16. Niu, H.; Silva, E.A. 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](#)]
17. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022. [[CrossRef](#)]
18. Xing, H.; Meng, Y. Integrating landscape metrics and socioeconomic features for urban functional region classification. *Comput. Environ. Urban Syst.* **2018**, *72*, 134–145. [[CrossRef](#)]
19. Xu, N.; Luo, J.; Wu, T.; Dong, W.; Liu, W.; Zhou, N. Identification and Portrait of Urban Functional Zones Based on Multisource Heterogeneous Data and Ensemble Learning. *Remote Sens.* **2021**, *13*, 373. [[CrossRef](#)]
20. Du, S.; Du, S.; Liu, B.; Zhang, X.; Zheng, Z. Large-scale urban functional zone mapping by integrating remote sensing images and open social data. *GIScience Remote Sens.* **2020**, *57*, 411–430. [[CrossRef](#)]
21. Sun, Z.; Li, P.; Wang, D.; Meng, Q.; Sun, Y.; Zhai, W. Recognizing Urban Functional Zones by GF-7 Satellite Stereo Imagery and POI Data. *Appl. Sci.* **2023**, *13*, 6300. [[CrossRef](#)]
22. Gao, Z.; Sun, W.; Cheng, P.; Yang, G.; Meng, X. Identify Urban Functional Zones Using Multi Feature Latent Semantic Fused Information of High-spatial Resolution Remote Sensing Image and POI Data. *Remote Sens. Technol. Appl.* **2021**, *36*, 618–626. [[CrossRef](#)]
23. Du, Z.; Zhang, X.; Li, W.; Zhang, F.; Liu, R. A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou City, China. *Trans. GIS* **2019**, *24*, 123–141. [[CrossRef](#)]
24. Zhang, Y.; Li, Q.; Tu, W.; Mai, K.; Yao, Y.; Chen, Y. Functional urban land use recognition integrating multi-source geospatial data and cross-correlations. *Comput. Environ. Urban Syst.* **2019**, *78*, 101374. [[CrossRef](#)]
25. Zhang, X.; Li, W.; Zhang, F.; Liu, R.; Du, Z. Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 459. [[CrossRef](#)]
26. Yu, L.; He, X.; Liu, J. Discovering urban functional regions based on semantic mining from spatiotemporal data. *J. Sichuan Univ. (Nat. Sci. Ed.)* **2019**, *56*, 246–252.
27. Chen, S.; Tao, H.; Li, X.; Zhuo, L. Discovering urban functional regions using latent semantic information: Spatiotemporal data mining of floating cars GPS data of Guangzhou. *Acta Geogr. Sin.* **2016**, *71*, 471–483. [[CrossRef](#)]
28. Wang, X.; Grimson, E. Spatial latent dirichlet allocation. *Adv. Neural Inf. Process. Syst.* **2007**, *20*, 1577–1584.
29. Pan, Z.; Liu, Y.; Liu, G.; Guo, M.; Li, P. MTRF: A topic model with spatial information. *J. Comput. Appl.* **2015**, *35*, 2715.
30. Li, Y.; Shao, H.; Jiang, N.; Shi, e.; Ding, Y. Classification of land cover in high-resolution remote sensing images based on Space-LDA model. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 177–183. [[CrossRef](#)]
31. Chen, Z.; Soifer, I.; Hilton, H.; Keren, L.; Jojic, V. Modeling Multiplexed Images with Spatial-LDA Reveals Novel Tissue Microenvironments. *J. Comput. Biol.* **2020**, *27*, 1204–1218. [[CrossRef](#)] [[PubMed](#)]
32. Canh, T.V.; Gertz, M. A spatial LDA model for discovering regional communities. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Niagara, ON, Canada, 25–29 August 2013; pp. 162–168.

33. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
34. Huang, C.; Davis, L.; Townshend, J. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [[CrossRef](#)]
35. Hong, Y.; Yao, Y. Hierarchical community detection and functional area identification with OSM roads and complex graph theory. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1569–1587. [[CrossRef](#)]
36. Patle, A.; Chouhan, D.S. SVM kernel functions for classification. In Proceedings of the 2013 International Conference on Advances in Technology and Engineering, Mumbai, India, 23–25 January 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–9. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.