A Method for Extracting Functional Areas Specialized for Bicycle Usage

Jiwon Lee [1], Wonjong Gong [2], Hyeongwoo Kim [3], Jiyoung Kim [4]

[1] Social Eco Tech Institute, Konkuk University, Seoul 05022, Republic of Korea – vert802@konkuk.ac.kr
[2] Social Eco Tech Institute, Konkuk University, Seoul 05022, Republic of Korea – gwjgwj@konkuk.ac.kr
[3] Social Eco Tech Institute, Konkuk University, Seoul 05022, Republic of Korea – kimhyeongwoo@konkuk.ac.kr
[4] Social Eco Tech Institute, Konkuk University, Seoul 05022, Republic of Korea – elliekim@konkuk.ac.kr

Keywords: Bicycle Usage, Functional Area Extraction, POI Embedding, Network Voronoi Diagram.

Abstract

The extraction of urban functional areas plays a critical role in data-driven policymaking. While previous studies have primarily focused on general-purpose functional area extraction, this study proposes a novel methodology for identifying bicycle-specialized functional areas by integrating spatial network analysis with semantic POI embedding. Using Seoul, South Korea, as a case study, we first constructed linear spatial units by applying a network Voronoi algorithm to the city's bicycle road network and public bicycle station data. Next, POI data were classified according to bicycle trip purposes and embedded using a Word2Vec-based approach to capture high-dimensional semantic features. These features were then aggregated for each spatial unit, with weights assigned based on POI type and proximity to bicycle roads. Finally, K-means clustering was conducted to extract distinct functional areas optimized for bicycle usage. The experimental results identified four unique cluster types, including residential-centered and park-oriented zones, demonstrating the effectiveness of the proposed methodology in supporting bicycle-friendly urban planning. This approach offers valuable insights for public bicycle redistribution, infrastructure deployment, and sustainable mobility policy.

1. Introduction

1.1 Research Background

The extraction of urban functional areas is crucial as foundational data for decision-making across various fields such as housing, transportation, and environmental policy. In recent years, with the increasing accessibility of data such as Points of Interest (POI) and satellite imagery, along with advances in AI technologies, research on automatically identifying urban functional areas has grown significantly. However, previous studies have rarely extracted functional areas for specific purposes. Most have focused on delineating the entire urban landscape from a generalized perspective. Yet, cities are complex and organically structured systems, and the perception of space can vary depending on one's perspective. For instance, functional characteristics emphasized in a neighbourhood may differ depending on whether an individual is a car user or a public transit user. An area physically close to a subway station may be seen as highly accessible by public transit users but may be perceived as congested and inconvenient by car users due to heavy pedestrian traffic.

Taking these research questions into account, this study proposes a methodology for extracting functional areas specialized for bicycle usage. As bicycles continue to gain attention as an environmentally friendly mode of transport—particularly within the context of the growing sharing economy—their societal importance is increasing. Therefore, the extraction of functional areas tailored to bicycle usage can be highly valuable for policymaking, such as bicycle redistribution and the installation of related infrastructure.

To achieve this, we propose a method that defines spatial units from the perspective of bicycle users and assigns weights to POIs that are closely related to bicycle usage. Whereas previous research has commonly used polygon-based spatial forms—such as administrative boundaries or grids—this study introduces linear spatial units based on the bicycle road network. Additionally, POI types that promote bicycle use are weighted according to findings from prior studies to ensure they are appropriately reflected in the extraction process.

1.2 Literature Review

Traditional methods for extracting urban functional areas primarily rely on direct surveys, including field investigations and resident interviews (Zhai *et al*., 2019). However, these approaches are time-consuming and costly, and they are limited in capturing the dynamic and rapidly changing nature of urban environments. With the recent proliferation of urban-related data and advancements in AI technologies capable of processing such data, there has been a growing trend toward utilizing diverse data sources for urban functional area extraction.

The data used for this purpose can be broadly categorized into remote sensing data and geospatial big data. Remote sensing data can reflect the physical appearance of a city in near real-time, and the rapid progress of image-related AI technologies has facilitated numerous studies in this domain. Nevertheless, urban functional areas cannot be fully explained by mere land use distribution alone. They are often closely associated with internal socio-economic activities within a space. As a result, there has been an increasing focus on studies utilizing geospatial big data, which include not only physical vector data but also unstructured data such as trajectory data from various modes of transportation and social media data.

Due to this wide applicability, recent studies have also explored the integration of remote sensing data and geospatial big data for urban functional area extraction (Qian *et al*., 2020; Wang and Feng, 2024). The extraction methodologies differ depending on the type of data employed. Among studies using geospatial big data—such as POI data—methods commonly include density-based approaches (Deng and He, 2022; Luo *et al*., 2023) and semantic information extraction based on Word2Vec (Zhai *et al*., 2019; Zhang *et al*., 2021; Niu and Silva, 2021; Qin *et al*., 2022;

Wang and Feng, 2024). In the latter case, semantic features are first extracted and then similar types are grouped using clustering methods to delineate urban functional areas.
Thus, recent research trends focus less on the novelty of algorithms themselves and more on improving and integrating algorithms to extract more precise semantic information. Table 1 summarizes and compares related studies in terms of data sources, methodologies, and spatial units used for urban functional area extraction.

Author	Data	Method	Spatial Unit
Zhai *et al.* (2019)	POI	Place2vec K-means	Administrative district
Qian *et al.* (2020)	RS image Taxi GPS	YOLO V3 K-means++ KNN	Road block
Zhang *et al.* (2021)	POI	Glove K-means++	TAZ
Niu and Silva (2021)	POI	Doc2vec K-means	Administrative district
Hu *et al.* (2021)	Taxi GPS POI	GCNN	Road segment
Deng and He (2022)	Building POI	LDA SVM	Voronoi diagram
Qin *et al.* (2022)	POI	Word2vec Random forest	Road block
Luo *et al.* (2023)	POI	Kernel density K-star	Grid
Liu *et al.* (2023)	Taxi GPS POI	CA-RFM	Road block
Wang and Feng (2024)	RS image Building POI Social media	VGG16 BERT Random forest	Road block (Parcel)

Table 1. Comparison of Related Work

As shown in Table 1, the spatial units employed for urban functional area extraction vary across studies. Given that urban spaces are continuous and complex, it is essential to subdivide them into discrete spatial units for analysis. In this context, accurately defining spatial units and assigning appropriate functions are critical for robust analytical results. In previous studies on urban functional area extraction, administrative units (Zhai *et al*., 2019; Zhang *et al*., 2021; Niu and Silva, 2021) and block units delineated by road boundaries (Qian *et al*., 2020; Qin *et al*., 2022; Liu *et al*., 2023; Wang and Feng, 2024) have been widely adopted. More recently, however, diverse spatial units such as Voronoi diagrams (Deng and He, 2022), road segments (Hu *et al*., 2021), and grids (Luo *et al*., 2023) have also been utilized.
Each type of spatial unit has its own advantages and disadvantages depending on the data used and the analytical methodologies applied. Administrative units, for example, have the advantage of compatibility with official demographic and statistical data collected at the national level. However, they may fail to capture actual human mobility patterns and overlook activities occurring near administrative boundaries.
Road-block-based units, which are subdivided based on road networks, reflect the natural boundaries of urban spaces. They are advantageous in that they can better represent urban activity structures and are intuitive for spatial interpretation. Nevertheless, their effectiveness depends heavily on the accuracy and resolution of road network data (Du *et al*., 2024).
Other units, such as Voronoi diagrams and grids, are structurally simple and easy to construct, making them suitable for large-scale datasets. However, they have limitations in that they do not necessarily account for the specific characteristics of the data or actual human activity areas.
To address these challenges, Deng and He (2022) proposed a method for deriving spatial units that considers these limitations. They pointed out that previous studies often relied on overly large spatial units, such as parcels or traffic analysis zones (TAZs), which tend to overlook the fine-grained characteristics of urban areas and may introduce biases into research findings. To mitigate this issue, they introduced a spatial partitioning approach using Voronoi diagrams aimed at capturing human activity-centered functional areas. Their proposed method divides space using Voronoi diagrams but also takes into account buildings and roads—critical elements of human activity—when defining analytical units. This approach enables the extraction of functional areas that more accurately reflect actual zones of human activity.
These issues related to spatial units are also closely linked to the data utilized in functional area extraction. In particular, some studies have employed trajectory data of transportation modes to better account for actual human mobility when delineating functional areas (Qian *et al*., 2020; Hu *et al*., 2021; Luo *et al*., 2023). In these cases, spatial units are primarily derived based on roads. However, most studies have not directly used individual road segments but instead have divided the space into blocks based on the road network.
Hu *et al.* (2021), however, addressed this research gap by classifying functional areas at the road segment level. They analogized road segments to "words" and taxi trajectories to "documents," and trained a Word2Vec-based model to semantically embed each segment into high-dimensional vectors. Furthermore, they labeled the function of each road segment using surrounding POI data, and finally utilized a Graph Convolutional Neural Network (GCNN) model to predict the function of each segment. This study is significant in that it extracted functional areas directly at the road segment level, thereby reflecting the characteristics of actual taxi trajectory data where the data are generated. Nevertheless, the study has limitations, as it relied solely on taxi trajectory data for extracting semantic information from road segments. The POI data they used served only as labeling data for road function prediction, and even then, the POIs were simply categorized into three types: commercial, public, and transportation. Such coarse categorization is insufficient to fully capture the complexity of urban functions.
In this context, the present study introduces several contributions that differentiate it from prior research.
First, unlike previous studies that aimed to classify functional areas for the entire city from a general perspective, this study focuses on identifying functional areas specialized for bicycle usage. The classification of urban functional areas is highly valuable as foundational data for various policy decisions. Specifically, the identification of bicycle-specialized functional areas can directly support decision-making related to bicycle policies. For instance, areas classified as residential-centered bicycle usage zones could be expected to have high bicycle demand during morning commuting hours and bicycle surpluses in the evening. Such insights could inform strategies for bicycle redistribution and station rebalancing.
Second, to extract bicycle-specialized functional areas, this study proposes a novel methodological framework distinct from existing approaches. While most previous studies have used block-based spatial units, this study introduces a linear, network-based spatial unit that considers bicycle road accessibility. By applying a network Voronoi algorithm based on the locations of

public bicycle stations, we subdivided the accessible road segments to define spatial units. Additionally, when extracting semantic information from POI data, we improved the algorithm by assigning weights to POI types that promote bicycle usage, thereby enabling the derivation of semantic features that more accurately reflect bicycle-related functions.

2. Methodology

This study aims to extract functional areas specialized for bicycle usage by first defining road-based spatial units that reflect bicycle usage patterns. To identify the function of each spatial unit, POI data with socially meaningful classifications were embedded using a Word2Vec-based method, from which high-dimensional semantic features were extracted. Finally, POI data were aggregated at the spatial unit level to extract functional areas tailored to bicycle usage. The overall research framework is illustrated in Figure 1.

2.1 Road-Based Spatial Units Reflecting Bicycle Usage

To extract urban functional areas specialized for bicycle usage, it is essential to define spatial units based on the actual roads where bicycle activity occurs. To achieve this, the locations of bicycle stations were first geocoded, and a network Voronoi diagram was applied to partition the accessible road segments for each station. While a conventional Voronoi diagram divides a plane by identifying regions closest to a set of seed points using Euclidean distance, a network Voronoi diagram instead uses network distances to determine proximity. Okabe *et al*. (2008) defined the network Voronoi diagram as the union of network Voronoi subsets, which can be formally expressed as Eq. (1).

$$Vor = \{Vor_1, Vor_2, Vor_3, \dots \dots, Vor_n\} \tag{1}$$
$$Vor_i = \{p \mid d_s(p, p_i) < d_s(p, p_j), j \neq i, j = 1, \dots, n\}$$

where Vor = network Voronoi diagram
Vor_i = network Voronoi subset
p = random point
p_i = general point
$d_s(p, p_i)$ = shortest distance on the network between points p and p_i
n = number of generated points on the network

Based on this equation, the bicycle road network was divided using the positions of bicycle stations as reference points. The bicycle road network comprises nodes and edges. First, for each bicycle station, the closest node was identified and set as the center node. Subsequently, edges connected to these nodes were classified. If an edge was connected to nodes belonging to the same Voronoi cell, it was assigned a single value corresponding to that cell. However, if an edge connected nodes belonging to different cells, the edge was split, and each segment was assigned the corresponding cell value.

2.2 Semantic Feature Extraction Using POI Data

To identify the functional characteristics of each spatial unit, it is necessary to extract semantic information from POI data that contain rich attribute information. For this purpose, we constructed a Word2Vec-based learning model following the approach proposed by Zhai *et al.* (2019). First, POI types were categorized according to bicycle trip purposes, based on publicly available data. Next, to create a training dataset, we formed pairs of central POIs and neighboring POIs. Specifically, for each central POI, the k nearest POIs were designated as neighboring POIs, forming context pairs. To enhance the model's ability to capture semantic similarity, the dataset was augmented so that POI types located closer in actual distance were more likely to be recognized as semantically similar.

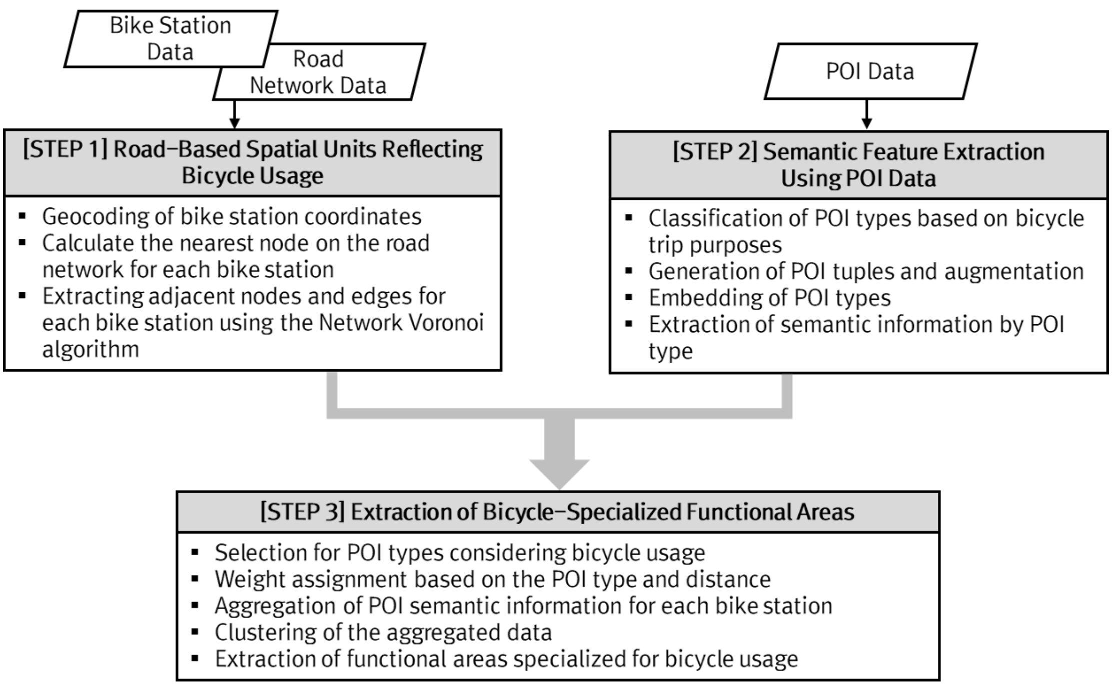

Figure 1. Research flow

This training dataset was then used to train a Skip-gram model of Word2Vec, which predicts surrounding words (neighboring POIs) from a given central word (central POI). The model was trained using a cross-entropy loss function, which measures the difference between the predicted probability distribution and the true distribution. Finally, a softmax function was applied to normalize the output values between 0 and 1, converting them into probabilities. As a result, each POI type was represented as a vector, and these vectors encapsulate the semantic information of each POI type.

2.3 Extraction of Bicycle-Specialized Functional Areas

Finally, to extract functional areas specialized for bicycle usage, each spatial unit was aggregated by assigning weights based on distance and POI type. To match point-based POIs to linear spatial units, the distance from each POI to the spatial unit was considered, with higher weights assigned to POIs located closer to the spatial unit. In addition, following the findings of Zhao *et al*. (2020), POIs classified as "Park" and "Transportation" were identified as having a positive influence on bicycle usage and were thus assigned additional weights. Ultimately, the weighted semantic information of POI types was aggregated for each spatial unit, and K-means clustering was performed to classify the functional areas.
To characterize the identified functional areas, each cluster was further analyzed using POI density and POI enrichment factor. POI density indicates the dominant POI type within a given cluster, while the POI enrichment factor enables comparisons of whether a certain POI type appears in relatively higher or lower proportions compared to other areas. The calculation methods for these indices are provided in Eq. (2) and Eq. (3).

$$PD = N_i^q / A_i \tag{2}$$

where N_i^q = number of POIs in category q in the cluster i
A_i = total area of the cluster i

$$EF = (N_i^q / N_i)/(N^q / N) \tag{3}$$

where N_i^q = number of POIs in category q in the cluster i
N_i = total POI count in cluster i
N_q = total POI count in cluster q
N = total count of POIs throughout the city

3. Experiments and Results

3.1 Study Area and Data

This study focuses on Seoul, South Korea, as the primary study area to extract functional areas specialized for bicycle usage. For this purpose, data on 2,628 public bicycle stations (Seoul Bike “Ddareungi”) were collected as of May 2022. To construct linear-based spatial units, only roads classified with a network type of “bike” from OpenStreetMap (OSM) were selected and used. In the case of POI data, about 780,000 data of 20 types were collected by referring to the travel purpose used by the Household Travel Survey. All datasets used in this study were obtained from the Seoul Open Data Plaza(data.seoul.go.kr). The entire algorithm was implemented using Python version 3.10.6. For network-based analysis, libraries such as NetworkX and OSMnx were utilized, and TensorFlow 2.0 was used for extracting semantic features of POI types.

3.2 Experimental Results

To establish linear, bicycle-specific spatial units, the locations of bicycle stations were first matched to the nearest nodes of bicycle roads provided by OpenStreetMap (OSM), which were then designated as central nodes. To apply the network Voronoi diagram to these central nodes, Dijkstra's algorithm was employed to classify sets of nodes closest to each central node. The resulting node classifications were stored as dictionary-type data. Next, edges connected to the classified nodes were subdivided. In cases where an edge connected nodes with different classification values, the Shapely library was used to split the edge accordingly. Through this process, a total of 2,628 bicycle stations were classified into linear bicycle road segments.
To identify the function of each spatial unit, semantic information was extracted from POI data. The POI data were categorized into 20 types and embedded using a Word2Vec-based approach. For this, K-nearest neighbors were used to select the 10 closest POIs to each central POI to form tuples. The tuples were then augmented by incorporating distance information before performing embedding. Referring to previous studies, the hyperparameters were set as follows: embedding dimension = 70 and number of iterations = 10,000.
Finally, the functional areas were aggregated and classified based on the previously defined spatial units. During aggregation, weights were assigned to individual POI data to emphasize bicycle-specialized functional characteristics. Following prior studies, POIs categorized as "Park" and "Transportation" were assigned a weight of 2, while all other POI types were assigned a weight of 1. Additionally, the distance from each POI to the bicycle road segment within each spatial unit was taken into account. POIs were assigned differentiated weights based on their proximity as follows: within 10 meters, 10–50 meters, 50–100 meters, 100–500 meters, and beyond 500 meters.
These weighted values were then combined with the previously extracted 70-dimensional semantic vectors to compute the final functional value for each spatial unit. Based on these aggregated results, K-means clustering was performed to classify the functional areas. The optimal number of clusters was determined using the elbow method, resulting in K = 4 for the final functional area classification.
The results of the bicycle-specialized functional area extraction are shown in Figure 2. To analyze the identified functional areas, POI density and POI enrichment factor were calculated for each cluster(Table 2 and Table 3). The findings revealed: (Cluster 1) Bicycle usage areas centered around detached housing, (Cluster 2) Bicycle usage areas centered around villa-type housing, (Cluster 3) Bicycle-friendly areas with a concentration of parks and schools, (Cluster 4) Areas used for commuting or shopping by bicycle
The upper part of Figure 2 provides a zoomed-in view of how each cluster is spatially differentiated. Notably, since the spatial units were established based on roads accessible from bicycle stations, the aggregation of POIs differs from that of previous studies, offering a unique and refined approach.

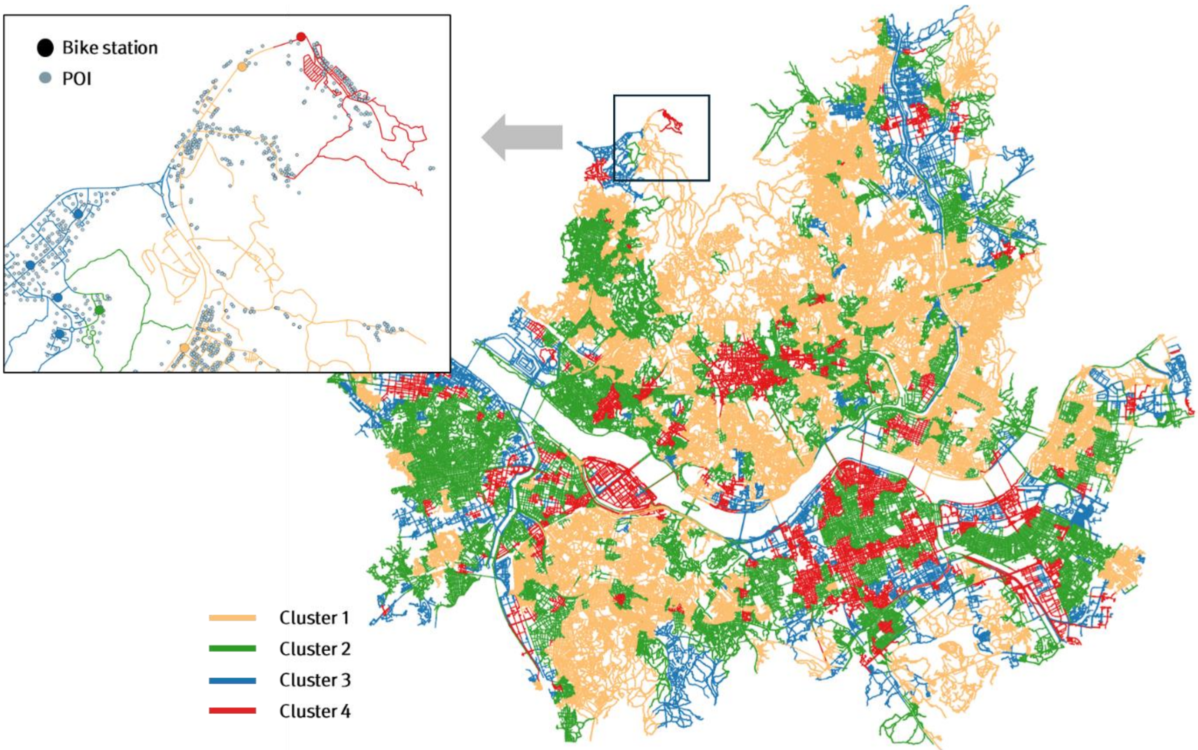

Figure 2. Functional area extraction results

POI category	Cluster1	Cluster2	Cluster3	Cluster4
Home_single house	0.566	0.293	0.060	0.097
Home_apartment	0.015	0.031	0.250	0.023
Home_others	0.114	0.173	0.053	0.036
Work_public	0.001	0.001	0.004	0.002
Work_private	0.081	0.146	0.047	0.186
Work_factory	0.005	0.010	0.018	0.014
School	0.006	0.009	0.044	0.005
Academy	0.007	0.007	0.009	0.004
Job-related service	0.006	0.007	0.014	0.008
Shopping	0.030	0.040	0.034	0.071
Leisure	0.011	0.013	0.016	0.017
Park	0.004	0.006	0.024	0.005
Dining_restaurant	0.078	0.135	0.113	0.289
Dining_café	0.016	0.031	0.029	0.064
Dining_bar	0.014	0.024	0.007	0.055
Medical_hospital	0.003	0.008	0.013	0.032
Medical_pharmacy	0.004	0.006	0.010	0.013
Transport_bus	0.035	0.049	0.231	0.055
Transport_subway	0.003	0.008	0.020	0.021
Transport_parkinglot	0.002	0.002	0.003	0.003

Table 2. POI density by cluster

POI category	Cluster1	Cluster2	Cluster3	Cluster4
Home_single house	0.626	0.323	0.067	0.107
Home_apartment	0.174	0.354	2.871	0.267
Home_others	0.381	0.579	0.179	0.121
Work_public	0.232	0.328	1.088	0.609
Work_private	0.304	0.553	0.179	0.701
Work_factory	0.215	0.454	0.804	0.617
School	0.248	0.338	1.693	0.196
Academy	0.405	0.447	0.529	0.256
Job-related service	0.324	0.404	0.770	0.428
Shopping	0.350	0.466	0.397	0.836
Leisure	0.315	0.373	0.459	0.491
Park	0.795	1.107	4.452	0.872
Dining_restaurant	0.265	0.457	0.384	0.982
Dining_café	0.247	0.471	0.446	0.972
Dining_bar	0.278	0.482	0.140	1.085
Medical_hospital	0.398	1.058	1.625	4.151
Medical_pharmacy	0.254	0.411	0.677	0.861
Transport_bus	1.105	1.563	7.376	1.766
Transport_subway	0.645	1.660	4.036	4.153
Transport_parkinglot	0.896	1.261	1.588	1.742

Table 3. POI enrichment factor by cluster

4. Conclusion

This study proposed a novel methodology for extracting functional areas specialized for bicycle usage by subdividing spatial units based on actual bicycle road networks and aggregating high-dimensional semantic information derived from POI location and type data. The experiment was conducted using all public bicycle stations in Seoul, South Korea.

First, road networks accessible from each station were divided by applying a network Voronoi diagram based on station locations. Next, POI types were categorized considering bicycle trip purposes, and these were used as training data to extract semantic vectors for each type. Finally, weights were assigned to individual POIs based on their distance and type, and semantic information was aggregated for each spatial unit. K-means clustering was then applied to classify the functional areas. As a result, four bicycle-specialized functional area clusters were identified, and the characteristics of each cluster were analyzed using POI density and POI enrichment factor metrics.

This study is distinct from previous research in that it subdivided spatial units into linear segments to better reflect actual bicycle usage patterns, rather than relying on traditional block-based or administrative units for the entire city. Furthermore, by incorporating both distance and type-based weights in POI aggregation, the proposed approach provides a more nuanced method for identifying bicycle-specific functional areas.

However, this study has limitations in that it only utilized spatial semantic information derived from POI data when extracting functional areas. In the context of bicycle-specific functional areas, not only the physical urban functions but also the dynamic flow of movements are critical factors. Therefore, future research aims to propose a spatiotemporal functional area extraction method that incorporates temporal information related to bicycle movements in addition to spatial characteristics.

Acknowledgement

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143804).

References

Deng, Y., & He, R., 2022: Refined Urban Functional Zone Mapping by Integrating Open-Source Data. *ISPRS International Journal of Geo-Information*, 11(8), 421.

Du, S., Zhang, X., Lei, Y., Huang, X., Tu, W., Liu, B., & Du, S., 2024: Mapping urban functional zones with remote sensing and geospatial big data: a systematic review. *GIScience & Remote Sensing*, 61(1), 2404900.

Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., & Gong, X., 2021: Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems*, 87, 101619.

Liu, T., Cheng, G., & Yang, J., 2023: Multi-Scale Recursive Identification of Urban Functional Areas Based on Multi-Source Data. *Sustainability*, 15(18), 13870.

Luo, G., Ye, J., Wang, J., & Wei, Y., 2023: Urban Functional Zone Classification Based on POI Data and Machine Learning. *Sustainability*, 15(5), 4631.

Niu, H., & Silva, E. A., 2021: Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Computers, Environment and Urban Systems*, 88, 101651.

Okabe, A., Satoh, T., Furuta, T., Suzuki, A., & Okano, K., 2008; Generalized network Voronoi diagrams: Concepts, computational methods, and applications. *International Journal of Geographical Information Science*, 22(9), 965-994.

Qian, Z., Liu, X., Tao, F., & Zhou, T., 2020: Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. *Remote Sensing*, 12(15), 2449.

Qin, Q., Xu, S., Du, M., & Li, S., 2022: Identifying urban functional zones by capturing multi-spatial distribution patterns of points of interest. *International Journal of Digital Earth*, 15(1), 2468-2494.

Wang, Z., Bai, J., & Feng, R., 2024: A Multi-Feature Fusion Method for Urban Functional Regions Identification: A Case Study of Xi'an, China. *ISPRS International Journal of Geo-Information*, 13(5), 156.

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

Zhang, C., Xu, L., Yan, Z., & Wu, S., 2021: A GloVe-Based POI Type Embedding Model for Extracting and Identifying Urban Functional Regions. *ISPRS International Journal of Geo-Information*, 10(6), 372.

Zhao, Y., Ke, S., Lin, Q., & Yu, Y., 2020: Impact of land use on bicycle usage. *Journal of Transport and Land Use*, 13(1), 299-316.