WILEY | Hindawi

Research Article

Planning Flexible Bus Service as an Alternative to Suspended Bicycle-Sharing Service: A Data-Driven Approach

Yujia Wang [1] **Yuan Tian** [1,2] **Binyu Yang** [1] **Jian Wang** [1,2] **Xiaowei Hu** [1] **and Shi An**[1,2]

[1]*School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China*
[2]*School of Management, Harbin Institute of Technology, Harbin 150001, China*

Correspondence should be addressed to Yuan Tian; ytian.phd@hotmail.com

Many free-floating bicycle-sharing (FFBS) operators in cold region cities will put bicycles in warehouses and suspend services in winter due to factors such as safety and maintenance costs, resulting in the corresponding travel demand no longer be met. Considering the short-distance and green travel characteristics of FFBS, the flexible bus, as a sustainable demand-responsive transit service, is a suitable alternative transportation mode. To operate such a flexible bus system, service area and operation time planning is a key stage, however, the planning methods in relevant studies are not suitable for this research scenario. In view of the above, this paper proposed a data-driven method to determine the service area and operation time of flexible buses based on FFBS data. Firstly, an FFBS trip path reconstruction algorithm consists of fine-grained road network modelling and trajectory matching is proposed. Then, in the defined time slice, incorporating the idea of ride-sharing routes generation, according to the density-based clustering principle and considering topology between trajectories, a path clustering algorithm PATHSCAN is developed to generate the one-day path clusters. After that, a frequent pattern mining algorithm is applied to the multiday path clusters, and frequent pattern results with spatio-temporal correlation will be merged into the final service area. The generated planning results will cover ride-matching trips and high-frequency riding paths. Detailed application analysis and verification are carried out by using the real data from Tianjin, China. Through the evaluation and verification under the relatively limited experimental data set, the proposed data-driven method shows ideal planning results. Flexible bus service can supplement the green short-distance travel mode after the suspension of FFBS and can avoid FFBS travel demands switched to unsustainable transportation modes to a large extent. This study will contribute to urban sustainable transportation development and improving greenness.

1. Introduction

In recent years, an active, low-emission, and sustainable transportation mode, the free-floating bicycle-sharing (FFBS) system, has sprung up all over the world. Based on the background of sharing economy and the wide adoption of smartphones, the FFBS system comes into fashion in more and more major cities in China [1]. Recently, the FFBS system has been in operation in more than 300 cities in China. However, the operations are greatly affected by seasonal changes and weather conditions [2, 3]. In the winter months, especially in cold regions with heavy snow, it can be expected that bicycle-sharing trips will decrease and bicycle

idle time will increase [4]. In addition to the reduction of revenue, bicycle-sharing operators will have to face even higher depreciation costs for bicycle damage. Besides, considering the safety factor of bicycle travel in winter, most bicycle-sharing service providers take bicycles back to the warehouse and do not provide operation services in winter in many cold regions of northern China. The suspension of FFBS will lead to the lack of flexible and convenient travel options for urban short-distance travel in winter.

The traditional public transit with fixed routes and schedules has low frequency and is less flexible; the traditional public transit usually cannot cover the short-distance travel undertaken by FFBS from the spatio-temporal dimension.

Taxis or ride-hailing services do provide greater flexibility, but the price is extremely high for most short-distance travelers. Besides, taxis and ride-hailing services tend to aggravate urban road traffic congestion and increase vehicle exhaust emissions, which is not conducive to the sustainable development of the transport industry. Last-mile demand-responsive transportation system is the closest to FFBS in terms of target demand, which is also the same as FFBS to solve the last-mile travel. The last-mile transportation system provides travelers with travel services from the nearest metro station or bus stop to a passenger's home or other destination [5]. People who travel between public transportation nodes and other destinations are the main target population for the last-mile service. The elderly, the disabled, commuters, and other special groups are the main service groups of the last-mile service [6]. There are currently some last-mile transportation modes, such as shuttle bus, community bus, and minibus. However, people who use FFBS to travel are not limited to the above groups. There is also a large demand for using FFBS for short-distance commuting, shopping, and other special-purpose trips that are unrelated to transit stations, but they are not within the scope of the last-mile transportation modes mentioned above. The FFBS trip duration distributions are much skewed towards shorter trips and have different daily and hourly characteristics in the distribution of use. Flexible bus is also one kind of demand-responsive transit services, which has no fixed routes and timetables, and its service mode can be consistent with the short-distance and flexible trip characteristics. Therefore, planning a flexible bus service according to the potential FFBS demand is considered a suitable and sustainable alternative after FFBS service is suspended in cold regions in winter. Different from the current last-mile transportation modes, service planning results of the flexible bus completely generated and driven by the potential FFBS demands will cover more general travel demand patterns, not limited to a specific travel demand pattern such as shuttle travel connecting public transit stations.

Because of high flexibility and benign travel experience, the flexible bus will play a principal role in advocating people to travel green. Guiding public transit travel is a key step for the transport industry to achieve carbon neutrality. In China, public transport priority has become a well-known national strategy. After the suspension of the bicycle-sharing service in winter, many short-distance travels using FFBS before may switch to some unsustainable transportation modes, such as private cars, taxis, or ride-hailing services. In addition, some long-distance trips completed by connecting FFBS and public transit may also have to be switched to unsustainable transportation modes. It is against the principle of green and sustainable transportation development. As a green transport mode, alternative flexible bus service planning generated driven by potential FFBS demand can avoid this problem to a large extent. The reasonably planned flexible bus service can fill the gap of green short-distance travel services in cold regions in winter and contribute to urban sustainable transportation development and improving the greenness.

Previous research on demand-responsive transit system focused on developing algorithms to optimize routes,

schedules, and fleet assignment [7–9]. Yet, the current research on how to determine the service area and operation time of a flexible bus is insufficient, which is also the vitally important section to determine whether the practical operation can succeed. Most existing studies lacked consideration of travel demand and are usually carried out from the perspective of spatial only, resulting in many cities still suffering from where to provide such flexible services in practice. For flexible services, the temporal dimension is an equally important perspective to the spatial dimension.

In view of the above, this paper aims to develop a data-driven method to determine the service area and operation time of flexible bus based on FFBS travel data. Moreover, the flexible bus service is suggested to operate within the determined service area to serve the travel demand after the FFBS is suspended in winter. Compared with previous studies on the flexible bus without considering the actual service area or representing the service area as geometric shapes, the service area determination in this paper will be completely driven by potential demand, as well as fully consider personal path selection preference and travel demand aggregation degree. The result will cover not only the spatial dimension but also the temporal dimension. In detail, the methodology consists of four stages: bicycle-sharing travel path reconstruction, path clustering based on the PATHSCAN algorithm, frequent pattern mining, and final service area determination based on clusters merging. The first stage refers to breaking the original road segments into subroad segments with a uniform length according to the travel characteristics of flexible bus service and then matching the bicycle-sharing trajectories with the subsegments. The core stage is a spatio-temporal trajectory clustering algorithm. Based on the network topology, the algorithm can detect bicycle-sharing trips which can share a common flexible bus service and generates service routes for one-day based on the detected trips. The frequent pattern mining algorithm is then used to process the multiday path clustering results. At last, frequent patterns with spatio-temporal correlation will be further merged into the proposed service area.

The proposed methodology will be applied to a real-world FFBS dataset, including 1,015,986 trips in a complete week collected from the Tianjin FFBS system. The planning process of flexible bus will be shown and analyzed in detail. In addition, three days' FFBS data close to winter will be further used to conduct a verification experiment, which can verify the substitution effect of the generated spatio-temporal service area for bicycle-sharing trips. This paper will creatively introduce the idea of carpooling routes generation into the bicycle-sharing trajectory clustering, and integrate frequent pattern mining to find the ride-matching FFBS trips and high-frequency riding paths, based on which the generated spatio-temporal planning result of flexible bus service can solve a large number of relatively clustered short-distance travel demands after the suspension of bicycle-sharing service in winter in cold regions. In addition, the work of this paper is advantageous to advocate green travel for the public and promote sustainable development of the transportation industry.

The remainder of this paper is organized as follows. Section 2 reviews the related research works. Section 3 presents the integrated spatio-temporal service determining method composed of four stages. A case study using the bicycle-sharing trajectory and order data of Tianjin is given in Section 4, and a verification experiment is carried out to verify the generated results. Section 5 concludes the paper and discusses future research directions.

2. Literature Review

2.1. Service Area and Operation Time Planning of Demand-Responsive Transit. It is vitally important to determine operation time and service area when planning and designing a demand-responsive transit service. However, most studies on demand-responsive transit do not pay enough attention to this stage but focus on optimizing routes [10, 11], schedules [12, 13], etc., under the assumption of ideal spatio-temporal demand distribution.

The few existing studies involving the determination of service area or operation time can be divided into two types. One assumes the service area as a geometric shape, and the other type usually generates a flexible transit service area based on potential travel demand.

As many cities have a square or rectangular grid street mode, the service area of demand-responsive transit is usually modeled as a geometric shape, mainly the rectangular [14, 15], and the travel requests are assumed to be uniformly distributed in the spatio-temporal dimension [16]. For example, Edward Kim et al. [14] modeled the service area when studying the service area design and route plan of a flexible bus feeder system connecting rail transit lines. Subsequently, Kim and Roch [15] proposed a mathematical model for the joint optimization of the flexible bus service area and headways, in which the service area does not change due to different time periods. Nourbakhsh and Ouyang [17] proposed the concept of "bus tube" to indicate the predetermined area served by the flexible-route bus, and the tube's shape was rectangular.

Some data-driven studies obtain the potential travel demand from personal mobility data first, and then flexible transit service area determined based on potential demands, such as Pan et al. [18], took the flexible feeder transit system serving irregular shapes and enclosed communities as the research object and proposed a gravity-based method of service area selection chain. The area containing the pick-up points is considered to be the service area of the flexible feeder transit system. Shu et al. [19] identified the potential last-mile travel demand from the multisource transportation data first and then generated shuttle service including bus stops and routing-scheduling solutions.

After reviewing these studies, it is found that there is a lack of reasonable and effective methods to determine the service area for a specific scenario. Despite some studies suggest that service area can be modeled as a geometric shape because of the pattern of the network, the road network layout of cities often has various patterns, and more importantly, the service area determination lacks consideration of travel demand. In addition, most research on flexible transit is only carried out from the spatial dimension, lack of consideration of the time dimension. The previous planning methods are not suitable for the application scenario of this study. How to determine the service area and operation time of flexible bus considering the potential travel demand is the problem to be solved in this study.

2.2. Trajectory Clustering Studies Related to Transportation Planning. Trajectory clustering is a process of classifying highly similar trajectories into one cluster, which mainly has two categories: trajectory clustering of moving objects in free space and road network environment. Gaffney and Smyth [20] proposed a probabilistic regression model to address the problem of clustering trajectories, which is the first trajectory clustering algorithm designed for free space. However, due to the large scale, time-aware, and lane change characteristics of vehicles, clustering vehicle trajectory is still a challenge. A great deal of trajectory clustering algorithms has been put forward and applied in the field of transportation planning [21–23]. For example, Hong et al. [24] proposed a TOPOSCAN trajectory clustering algorithm, which can be used in directed networks. Firstly, each trajectory was mapped to the corresponding path by map matching. Then, the path-segments were clustered together considering the path-segment density. Afterward, Hong et al. [25] proposed the shortest path distance measurement approach considering time dimension, ST-TOPOSCAN, which detected shared subpaths among trajectories based on the topology of predefined road networks and clustered them. Chen et al. [26] focused on the sensitive features of the road network and proposed a trajectory clustering method, TCRNC, which not only solved the problem of similarity measurement among trajectories but also effectively combined the local and global similarity features. The research results can be applied to many aspects of transportation, such as urban road planning and public transport planning.

However, the transportation planning results obtained by related trajectory clustering studies do not contain complete trajectories, but only parts of one trajectory participate in clustering, the generated planning result is a combination of hot road segments, such as the ride-sharing path recommendation results in Hong et al. [24]. Thus, the existing clustering methods are not appropriate to apply in this paper to determine the service area.

3. Methodology

In this section, the description of the research problem is given firstly. Then, the research framework of the proposed method is introduced. Finally, the proposed method is presented in detail.

3.1. Problem Description. People living in cold regions lack a flexible and convenient way to travel after the suspension of bicycle-sharing in winter. Whether considering the convenience and economic benefits of travelers or the sustainable development of the transport industry, the flexible bus is a suitable substitute. Unlike the traditional transit system,

where bus service routes, stops, and schedules are fixed, flexible bus picks up and drops off passengers within predetermined scopes without fixed stops.

This paper aims to mine the potential travel demand for flexible bus service from bicycle-sharing order and trajectory data and generate service area and operation time planning results driven by demand. The generated service area can vary with working days, weekends, and different periods of time to cope with the travel characteristics of large fluctuations.

3.2. Research Framework.
This paper mainly uses the path clustering and frequent pattern mining methods to find the ride-matching FFBS trips and high-frequency riding paths, to determine the substituted flexible bus service from the spatial and temporal dimensions. Figure 1 shows the research framework with four separate sections: data, method, results, and verification. The road network data and FFBS data within the studied area are used as input data. After preprocessing the input data, the spatio-temporal path clustering is performed. The path clustering algorithm (PATHSCAN) is developed according to the density-based clustering principle and considers topology between trajectories and is actually a data-driven ride-matching method, similar to ride-sharing routes generation, which is used to generate the one-day path clusters. The frequent pattern mining is then developed to process the multiday path clustering results. At last, after merging the frequent patterns, the final generated results reveal the service area and operation time.

3.3. Bicycle-Sharing Travel Path Reconstruction.
Before clustering, it is necessary to reconstruct the travel path of bicycle-sharing. Through fine-grained road network modelling and trajectory mapping, the trajectory vector is transformed into the path.

3.3.1. Fine-Grained Road Network Modelling considering the Topological Relationship

(1) Road Segment Division. A road network is represented by a directed graph $G = (V, E)$, where nodes $V = \{n_1, n_2, \ldots, n_m\}$ indicate intersections, and edges $E = \{(r_{id}, n_i, n_j) \mid n_i, n_j \in V\}$ indicate road segments between two intersections in the road network. It should be noted that one road segment from n_i to n_j has a different identifier from its opposite road segment from n_j to n_i. The original road segments will be divided into smaller units at the intersection firstly. Then, to achieve better results in the follow-up clustering, the road segments are further evenly interrupted. The interrupted segment is called the subroad segment, which is denoted as $s = (s_{id}, s_i, s_j) \in E$. It is assumed that the midpoint of the subroad segment is the potential getting on and off point for travelers. Obviously, if the divided subroad segment is too long, the walking distance will be also long. Moreover, if the subroad segment is too short, the flexible bus will start and stop frequently. Therefore, this paper will comprehensively consider the above factors to determine a reasonable length interval for the subroad segment.

(2) Road Network Adjacency Relationship Construction. The adjacency relationship between subroad segments is detected and recorded in a topology attribute table. Adjacency relationship is embodied by incoming subroad segments and outgoing subroad segments. If $s = (s_{id}, s_i, s_j) \in E$ is the object subroad segment, the incoming subroad segments are $s_k = (s_{id}^k, s_i^k, s_j^k \mid s_j^k = s_i) \in E$. In other words, the end node of the incoming subroad segment is the start node of the object subroad segment. On the contrary, the start node of the outgoing subroad segment is the end node of the object subroad segment. For instance, the road network shown in Figure 2 includes 16 nodes and 11 directional road segments denoted as different identifiers. The topology attribute table expressed in Table 1 shows the subroad segment ID, road segment ID, start and end nodes, and incoming and outgoing subroad segments.

3.3.2. Trajectory Mapping.
Trajectory mapping is to map the trajectory points to the corresponding road segments, so as to reduce the positioning error caused by a variety of external factors. In the past decades, the algorithms of trajectory mapping have been fully studied [27]. Trajectory mapping comprises three steps: the choice of candidate subroad segments for each trajectory point, the determination of matched subroad segments, and the checking of the adjacency relationship.

(1) The Candidate Road Segment Selection. The trajectory data used in this paper are a bicycle-sharing trajectory, which are a series of GPS points sorted in time order. It is expressed as a vector $TR = \{\text{routeCode}, p_1, p_2, \ldots, p_n\}$, where routeCode is the trip identifier meaning which trip the trajectory pertains to. p_i represents the time and position information when p_i is recorded. In this step, we take the trajectory point as the center of the circle and draw the circle with γ as the radius. The subroad segments within the circle and intersecting with the circle are regarded as the initial candidate subroad segments of the trajectory point. As shown in Figure 3, in the set of initial candidate subroad segments obtained in the first step, there are segments that do not meet the trajectory. The above segments are defined as redundant candidate subroad segments.

According to the adjacency relationship between subroad segments, when a candidate subroad segment is not the segment of the original and destination position of a trajectory, and only one endpoint has an adjacency relationship with other candidate segments, it will be seen as redundant. Deleting redundant candidate subroad segment is an iterative process. We filter out the redundant candidate segments in all initial candidate segments and delete them.

As shown in Figure 4, p_1 denotes a trajectory point. S_1, S_2, and S_3 denote the candidate subroad segments. Then, we determine the candidate points for the trajectory points. There are two cases of candidate points: one is to make a vertical line from the trajectory point to the candidate subroad segment. If there is a projection point on the
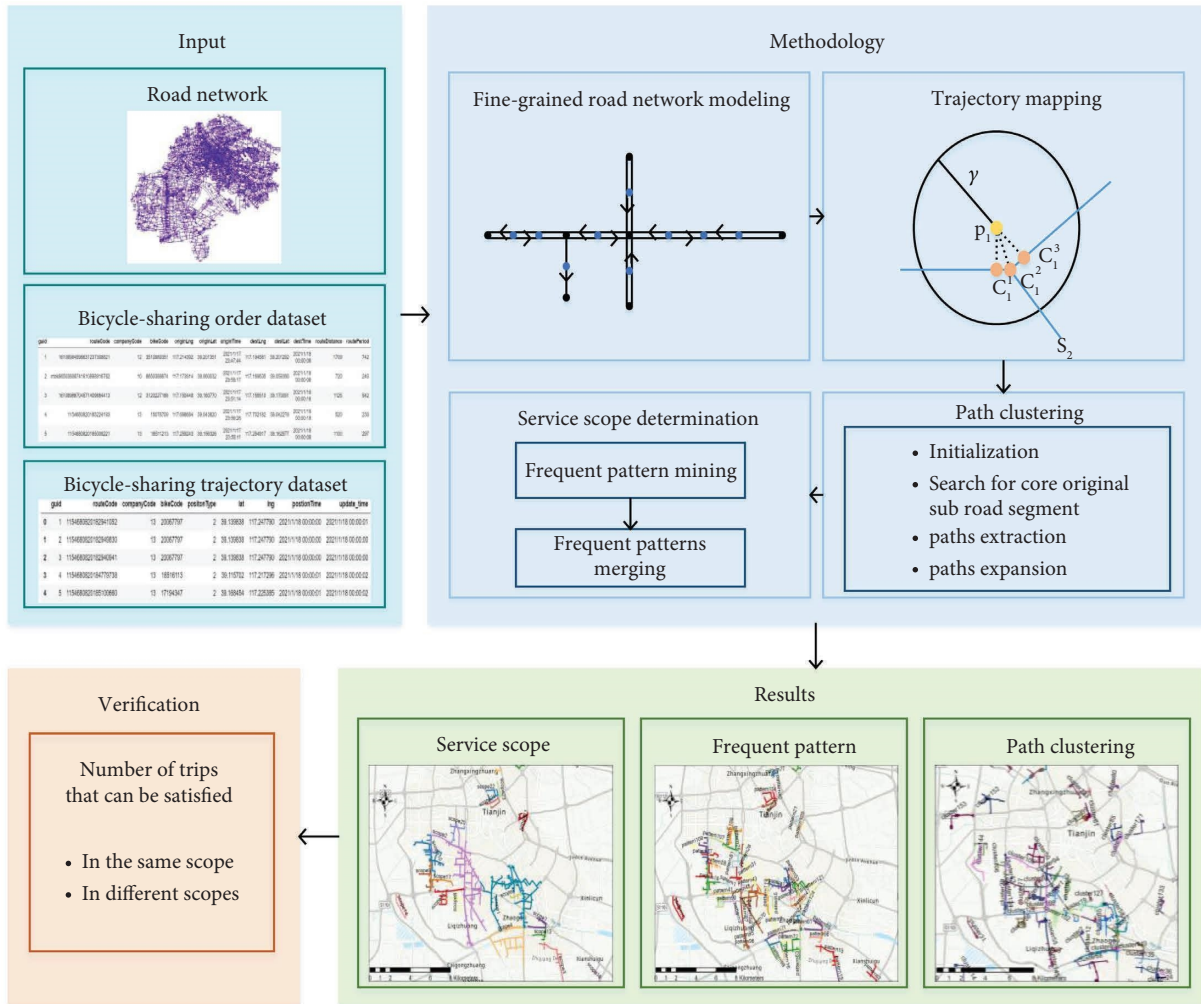
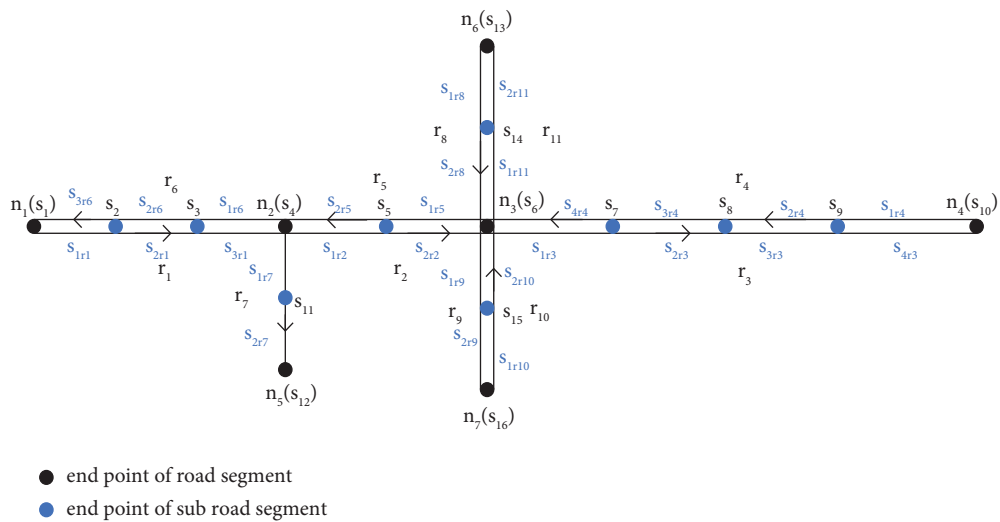FIGURE 1: The research framework of the methodology.

FIGURE 2: Illustrative diagram of fine-grained road network modelling.

TABLE 1: Illustrative table of road network topology.

Subroad segment ID	Road segment ID	Start node	End node	Incoming subroad segment	Outgoing subroad segment
s_{1r1}	r_1	n_1 (s_1)	s_2	s_{3r6}	s_{2r1}
s_{2r1}	r_1	s_2	s_3	s_{1r1}	s_{3r1}
s_{3r1}	r_1	s_3	n_2 (s_4)	s_{2r1}	s_{1r2}, s_{1r7}

FIGURE 3: Redundant candidate subroad segments.
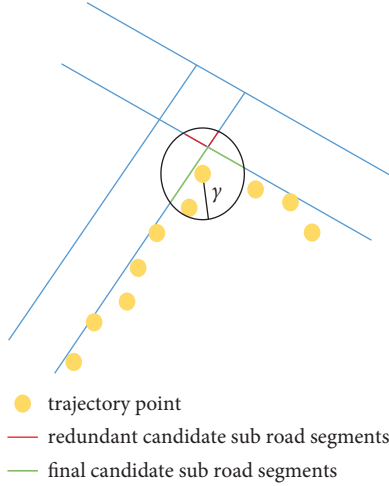
- trajectory point
- redundant candidate sub road segments
- final candidate sub road segments

candidate segment, the projection point is the candidate point, as shown in C_1^1 and C_1^3 in Figure 4; in another case, if there is no projection point on the candidate subroad segment, the point closest to the trajectory point on the candidate segment is taken as the candidate point, as shown in C_1^2 in Figure 4.

(2) The Matched Subroad Segment Determination. According to the hidden Markov model, the observation probability of candidate points and the transition probability between adjacent candidate segments are calculated. The hidden state sequence with the largest probability calculated by the Viterbi algorithm is the matched subroad segment sequence.

(3) The Adjacency Relationship Checking. For each trajectory point, it is necessary to judge whether there is an adjacency relationship between the matched subroad segment and the previous subroad segment. If not, suitable intermediate subroad segments should be inserted between the previous and current mapping segments. A path is a group of time-ordered, connected subroad segments passed by a trip routeCode, denoted as $P = \{(routeCode, S_i, S_j, \ldots, S_m) \mid S_i, S_j, \ldots, S_m \in E\}$. Matching will be completed between trajectories and paths with a shared routeCode. Only one consecutive subroad segment sequence is persisted to obtain the final path P. Thus, the group of trajectories can be regarded as the group of paths.

3.4. Path Clustering Based on PATHSCAN Algorithm. In order to find riding trajectories with high similarity in spatio-temporal dimension, this paper proposes a

PATHSCAN clustering algorithm. The procedures of path clustering are displayed in Figure 5.

After map matching procedures, bicycle-sharing GPS trajectory data are transformed into one-dimensional path data. In this section, the clustering objects are paths. The core neighborhood is the outgoing subroad segments connected to the core subroad segment. The proposed clustering algorithm follows the density-based principle and, considering the topology, takes advantage of a path set, a network topology attribute table, and a predetermined threshold parameter to generate the path clusters. It includes four main steps: initialization, core original subroad segment determination, path extraction, and path expansion.

  (i) Initialization

 (ii) Core original subroad segment determination: we define time slice and determine the core original subroad segment s_{cor}. A core original sub road segment is the segment with the largest number of bicycle-sharing trip starting points during the time slice

(iii) Path extraction: paths starting from the current core original subroad segment and meeting the density threshold will be extracted into path clusters

(iv) Path expansion: we search for the next consecutive subroad segment (i.e., the outgoing subroad segments connected to the current core subroad segment). We repeat the path extraction step.

Firstly, an empty set for the path clusters and core original subroad segment clusters are initialized, respectively, and define the iteration count.

Secondly, a time slice is predetermined and the number of trip starting points contained by each subroad segment with the time slice are counted. In the core original subroad segment search step, for the ith iteration, the ith cluster is initialized as an empty set. The subroad segment with the largest number of bicycle-sharing trip starting points is selected as the core original subroad segment. The core original subroad segment is inserted in the S_{cor}^i. If there are paths starting from the core original subroad segment with the number of trajectories contained larger than the threshold α, continue to the next step; or else, the iterations end, showing that no more cluster could be detected given that threshold during the time slice.

Thirdly, in the path extraction step, any path starting from the current core original subroad segment with the number of trajectories contained larger than the threshold is inserted into the cluster set C_i. In addition, the subset of paths that meet the above conditions is also inserted into C_i, whether they meet the threshold condition or not. For
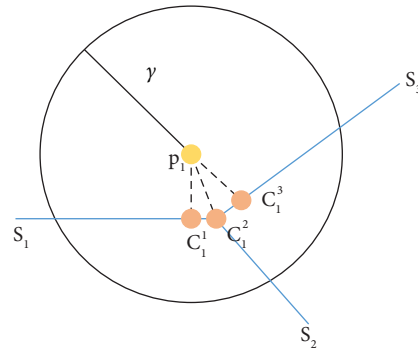
FIGURE 4: Candidate subroad segments and candidate points of trajectory point.
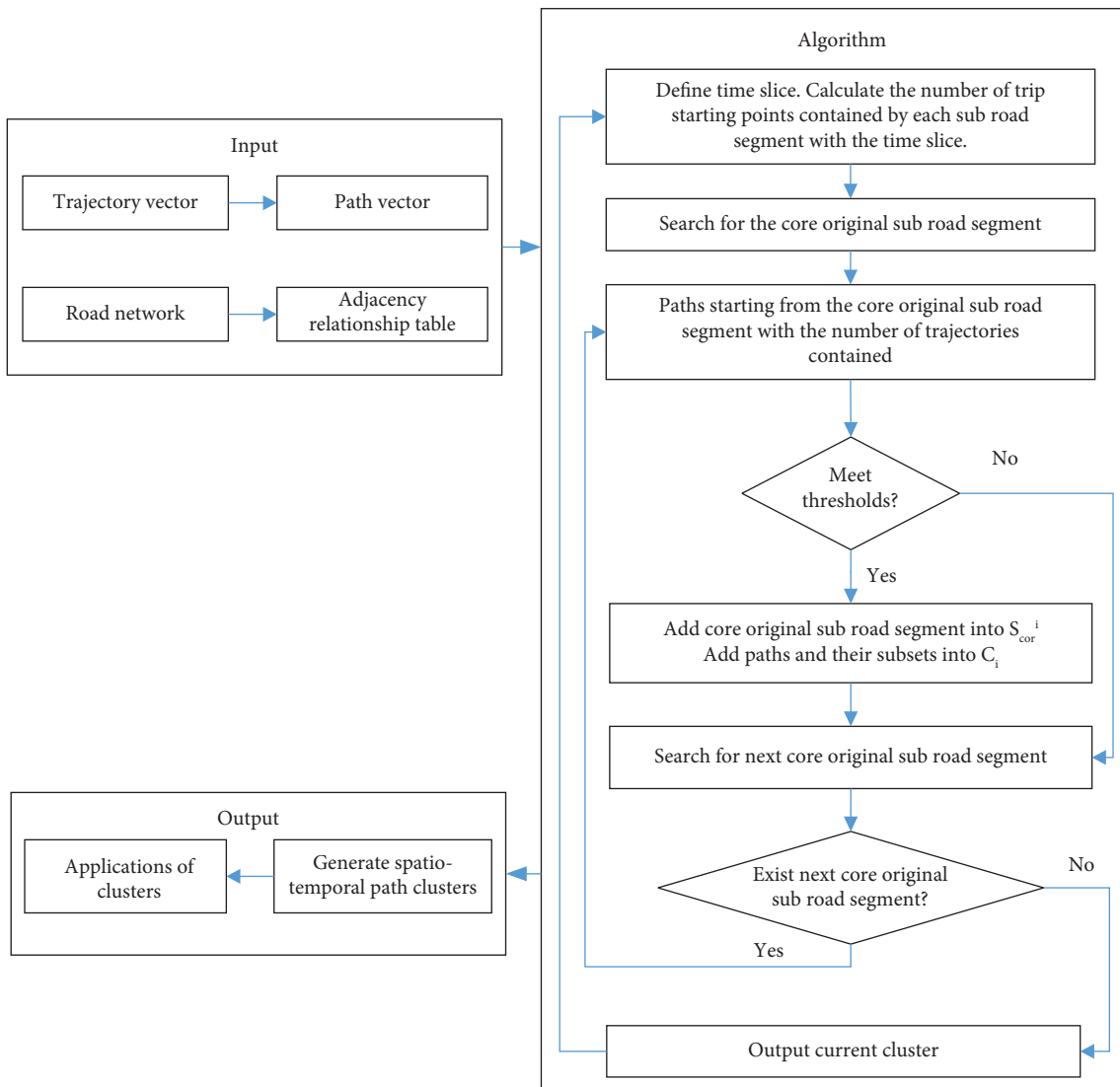

FIGURE 5: Flow chart of spatio-temporal path clustering.

example, as illustrated in Figure 2 and Table 2, assuming s_{2r1} is the current core original subroad segment, P_1, P_2, P_3, P_4, and P_5 meet the threshold condition. Obviously, P_2 is the subset of P_1. P_4 and P_5 are subsets of P_3, P_1, and P_3 and their subsets are all inserted into path clusters.

Finally, the outgoing subroad segment which connected to the current core original subroad segment is considered the next core original subroad segment. We repeat the path expansion step. It should be noted that the paths which have been inserted in C_i in the previous

Table 2: Example diagram of path vector.

Path vector	Subroad segments	Road segments
P_1	$s_{2r1}, s_{3r1}, s_{1r7}, s_{2r7}$	r_1, r_7
P_2	$s_{2r1}, s_{3r1}, s_{1r7}$	r_1, r_7
P_3	$s_{2r1}, s_{3r1}, s_{1r2}, s_{2r2}, s_{1r3}$	r_1, r_2, r_3
P_4	$s_{2r1}, s_{3r1}, s_{1r2}, s_{2r2}$	r_1, r_2
P_5	$s_{2r1}, s_{3r1}, s_{1r2}$	r_1, r_2

expansion will not participate in this expansion. Until no outgoing subroad segment can be expanded downward, the clustering in this direction ends.

The above steps will be carried out repeatedly until no paths starting from the core original subroad segment with the number of trajectories contained higher than the predetermined threshold value can be found.

3.5. Frequent Pattern Mining and Merging.
The Apriori algorithm is used to obtain frequent patterns from multiday path clusters. Frequent patterns refer to patterns that frequently appear in data sets. In the scenario of this paper, frequent pattern refers to the frequent riding routes of travelers using shared bicycles for a certain period, which constitutes the potential service area of a flexible bus. However, some of the frequent patterns overlap with each other in terms of route segments. The overlapped frequent patterns with the same temporal attribute will be merged into one pattern. Flexible bus service is proposed to operate on these merged path patterns.

As presented in Figure 6, after obtaining the path clusters, the frequent pattern paths within a time slice can be obtained using association rules. To find frequent paths over the entire time period, a spatial-temporal association rules mining algorithm (i.e., Apriori algorithm) is further used by modelling each subroad segment as an item and each path as a transaction. An association rule is a group of items that often appear in transactions at the same time. The relationship between transaction data is usually explicit, while the spatial objects often have implicit relationships. Thus, to apply the Apriori algorithm for detecting frequent pattern paths in road networks, the concepts related to the association rule must be redefined.

3.5.1. Transaction.
Each path detected from different time slices is materialized as a transaction. Each transaction used for the Apriori algorithm consists of a group of items (connected subroad segments) in the database. The transaction of a path P is defined as follows:

$$Transaction(p) = (routeCode, p_1, p_2, \ldots, p_i), \quad (1)$$

where $p_1, p_2, \ldots, p_i$ are a group of items corresponding to the subroad segments in path P.

3.5.2. Support.
Given the transaction database, the Apriori algorithm generates frequent groups of items utilizing

support measure. A path is regarded as frequent if its support is larger than the predetermined threshold. Given a set of transactions MP $= (P_1, P_2, \ldots, P_m)$, the support of an association rule P_j is defined as follows:

$$\sup port(P_j) = \frac{\left| P_j \in MP \cap P_j \subset P_i \right|}{|MP|}. \quad (2)$$

The transaction-based association rules mining algorithm improved based on the basic algorithm [28] can be easily applied to find the frequent pattern paths in various time slices. Spatio-temporal correlation rules will be used as the merging criteria of frequent patterns results.

4. Application Analysis and Verification

4.1. Data

4.1.1. Data Description.
The proposed method is applied to the FFBS trip data of Tianjin, China. Tianjin is located on the East Bank of the Eurasian continent at midlatitude, with cold winter. Tianjin covers a total area of 1,996,6.45 square kilometers, including 16 districts. 10 central urban districts are selected as the research area in this study. Although the number of FFBS trips decreased significantly in winter, the FFBS operators in Tianjin do not suspend the bicycle-sharing service. There is a lack of flexible and convenient travel options for short-distance travel, except for FFBS in winter. The reason why this paper chooses Tianjin as a case study city is to be able to use FFBS trip data in cold weather to verify the planning results of flexible bus service.

The experimental dataset includes two parts, the application dataset for the data-driven planning algorithm and the verification experimental dataset. The application dataset contains bicycle-sharing trajectory data and order data collected from September 9 to September 15, 2019, involving 7,111,899 trip records, 5,321,603 on weekdays and 1,790,296 on weekends. The average time span between two successive locations in the trajectory dataset is about 3 seconds. Each trajectory contains the bicycle code, the route code, position type, latitude, longitude, position time, and update time. Each order contains bicycle code, route code, latitude and longitude of origin and destination, time of origin and destination, and riding distance. The verification experimental dataset is introduced in detail in Section 4.3.

4.1.2. Data Preprocessing.
The road network data used in this paper is downloaded on the OSM (Open Street Map) platform. The road segments that are not relevant to this paper are firstly deleted, and then all the subroad segments are divided into lengths between 80 meters and 160 meters using ArcGIS Pro software. Next, a fine-grained road network model is constructed considering the topological relationships. The road network of the 10 districts studied is finally interrupted into 1,066,63 subroad segments, with an average length of 104.9 meters.

The quality of FFBS trajectory data is relatively rough due to the influence of acquisition equipment failure, network delay, and other factors. In order to improve data
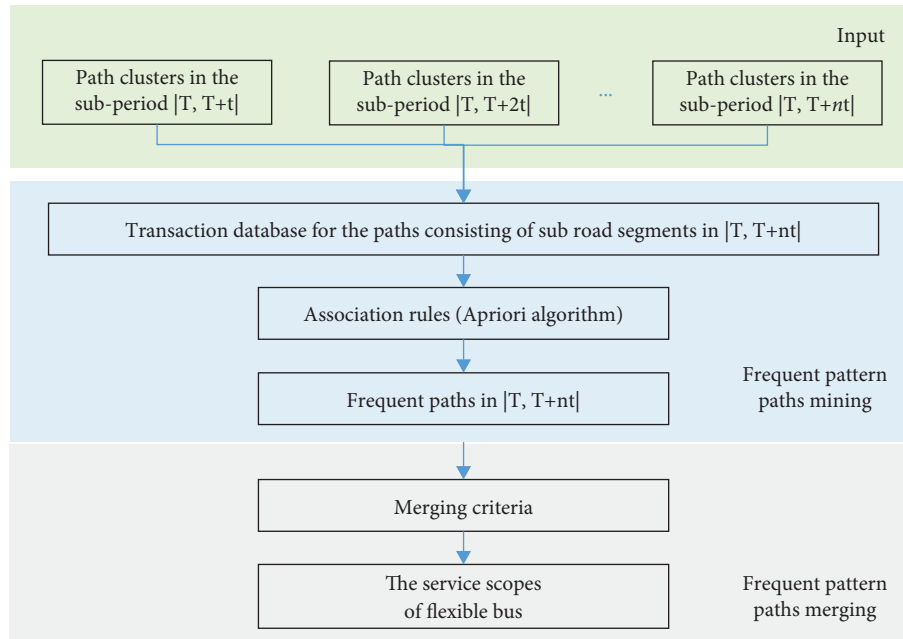
FIGURE 6: Process of frequent pattern paths mining and merging.

quality, this paper deals with data missing, data error, and data duplication. By analyzing the trip duration time, riding distance, and speed, errors and abnormal data are eliminated. In addition, considering the relatively low probability of passengers choosing a flexible bus for very short-distance travel, this paper selects about three subroad segments, i.e., 300 meters, as the screening threshold.

Then, the order data and trajectory data are matched according to their common field, which can supplement the missing information in some trajectory data. The trajectory data are mapped to the road network according to the algorithm in Section 3. For the order data without trajectory data, search the similar trips from matched results, the trajectory with the smallest spatial distance difference within the set threshold is selected as the missing trajectory fill, and if there are no matching results, the order data is deleted. The final processing results are shown in Table 3.

4.2. Result Analysis. In this study, the threshold α is set to 5, and the time slice is set to 30 minutes according to the postevaluation and the actual sensitivity analysis needs. The path clustering results of less than 30 trips in a single path clusters are further filtered. The path clustering results of four representative periods on September 9 are shown in Table 4. After path clustering, 37 path clusters are generated between 8:00 and 8:30 in the morning peak, which contained 1,235 trips in total, with an average of 24.96 subroad segments involved per cluster. Moreover, 112 path clusters are found between 18:00 and 18:30 in the evening peak, which contained 4,052 trips in total, with an average of 29.54 subroad segments involved per cluster. Compared with the results in the peak hours, the number of clusters is lower during the off-peak hours.

The path clustering results from 8:00 to 8:30 from September 9 to 15, 2019 are shown in Table 5. The number of path clusters on weekdays from the 9th (Monday) to the 13th (Friday) is much higher than that on the weekends of the 14th (Saturday) and the 15th (Sunday), and the number of trips involved in such path clusters is in different orders of magnitude. The result is consistent with the reality, that is, the period from 8:00 to 8:30 is the travel peak hours on weekdays and the off-peak period of travel on weekends. The number of subroad segments included in the path cluster reflects the length of the path cluster, and the results on weekdays are also significantly higher than those on weekends. Affected by weather, data quality, and other factors, the path clustering results of the morning peak on the weekdays still show great differences. The number of path clusters on September 11 and 12 is relatively large, 233 and 232, respectively, while there are only 37 on September 9.

The path clustering results from 18:00 to 18:30 on September 9 to 15 are shown in Table 6. The path clustering results of weekdays and weekends also show significant differences. On weekdays, the generation results of path clusters from Monday to Thursday are relatively stable, while the number of clusters on Friday is significantly reduced, which may be potentially related to the fact that many commuters go home directly after work from Monday to Thursday, and participate in some activities after work on Friday. The reduction of regular trips leads to a significant reduction in the number of path clusters.

Figure 7 shows the path clustering results from 8:00 to 8:30 on September 10 to 13, 2019. Different path clusters have different colors. The thickness of the subroad segment reflects the number of trips included in the paths starting from this segment. Obviously, through the comparative analysis

TABLE 3: Data processing results.

Date	Total number of orders	Orders within the research area	The number of trajectories matched
2019.09.09	1,105,779	953,256	623,781
2019.09.10	1,146,270	991,183	686,963
2019.09.11	1,145,758	990,931	701,346
2019.09.12	1,195,881	996,871	675,557
2019.09.13	727,915	602,072	416,978
2019.09.14	906,672	747,850	526,331
2019.09.15	883,624	727,301	498,822

TABLE 4: Path clustering results of representative periods on September 9, 2019.

Time period	Number of path clusters	Number of contained trips	Average number of subroad segments per path cluster
8:00–8:30	37	1,235	24.96
12:00–12:30	10	429	18.27
18:00–18:30	112	4,052	29.54
21:00–21:30	4	132	17.13

TABLE 5: Path clustering results from 8:00 to 8:30.

Date	Number of path clusters	Number of contained trips	Average number of subroad segments per path cluster
2019.09.09	37	1,235	24.96
2019.09.10	171	5,330	27.84
2019.09.11	233	7,894	29.41
2019.09.12	232	7,274	28.11
2019.09.13	49	1,497	23.87
2019.09.14	6	192	17.72
2019.09.15	10	303	19.31

TABLE 6: Path clustering results from 18:00 to 18:30.

Date	Number of path clusters	Number of contained trips	Average number of subroad segments per path cluster
2019.09.09	112	3,852	29.54
2019.09.10	115	4,189	32.37
2019.09.11	146	5,111	32.59
2019.09.12	118	3,915	28.02
2019.09.13	67	2,439	24.65
2019.09.14	13	483	27.95
2019.09.15	10	375	19.77

of the path clusters visualization on different weekdays, many path clusters have a strong spatial similarity.

The frequent patterns mining results from 8:00 to 8:30 on September 10 to 14 (weekdays) are shown in Figure 8. Different colors represent different frequent patterns. According to the visualization results, many frequent patterns have a spatio-temporal correlation. For example, frequent patterns 21 and 105 have overlaying relationship in space.

After merging the frequent patterns with spatio-temporal correlation, the flexible bus service area with clear spatio-temporal information is shown in Figure 9. Figure 9(a) shows the service area of 8:00 to 8:30 in the morning peak of weekdays. Figure 9(b) shows the service area of 18:00 to 18:30 in the evening peak of weekdays. Different colors represent different service scopes of the flexible bus. There are significant differences in the service areas in different periods on the same day, which is due to the different characteristics of travel demand in different periods.

4.3. Verification. A verification experiment is conducted and presented in this section. The FFBS trip data in cold weather on October 21, 23, and 31, 2019 are selected as the experimental verification data to verify the generated service area
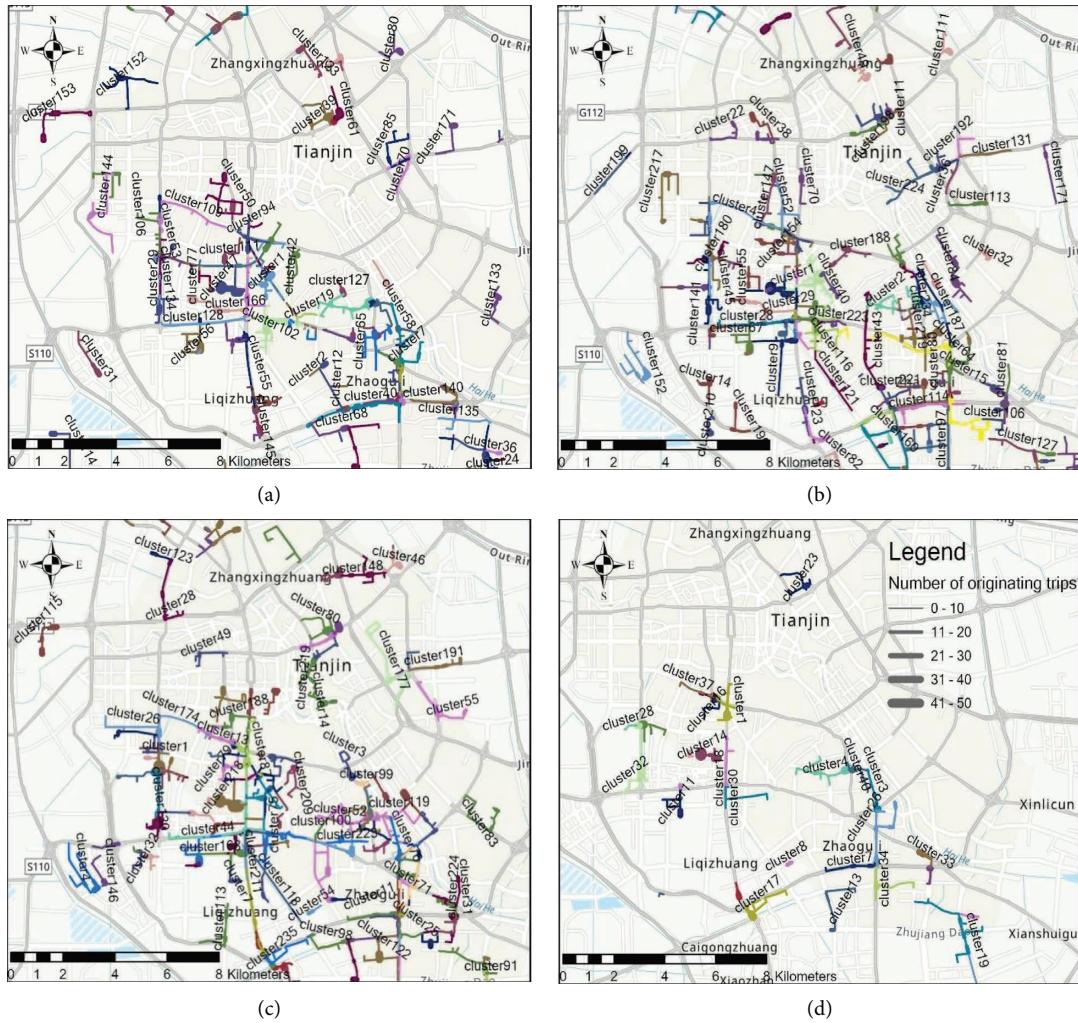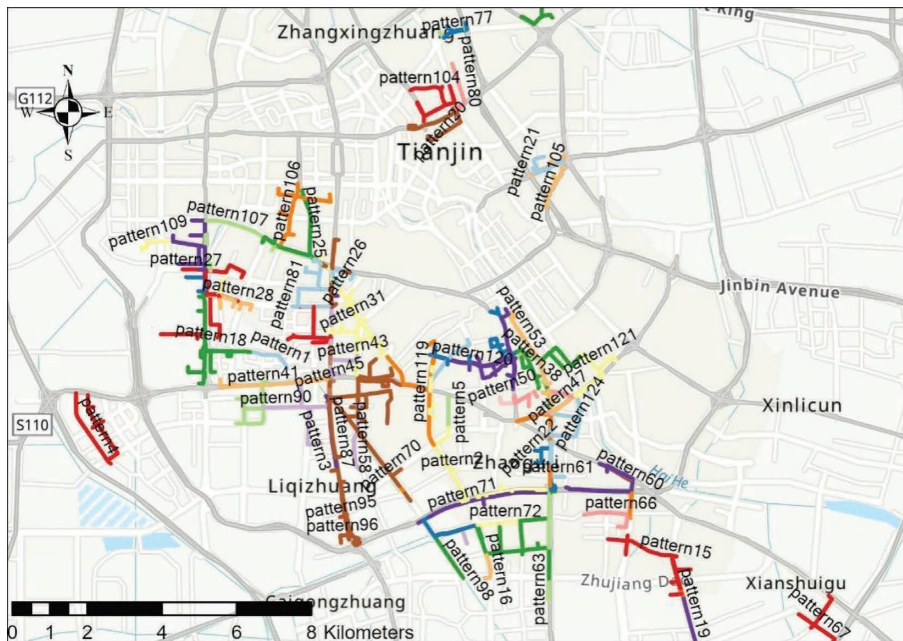
Figure 7: Path clustering results from 8:00 to 8:30 on September 10 to 13. (a) September 10th, (b) September 11th, (c) September 12th, and (d) September 13th.

Figure 8: Frequent patterns on September 10 to 14 (8:00 to 8:30).
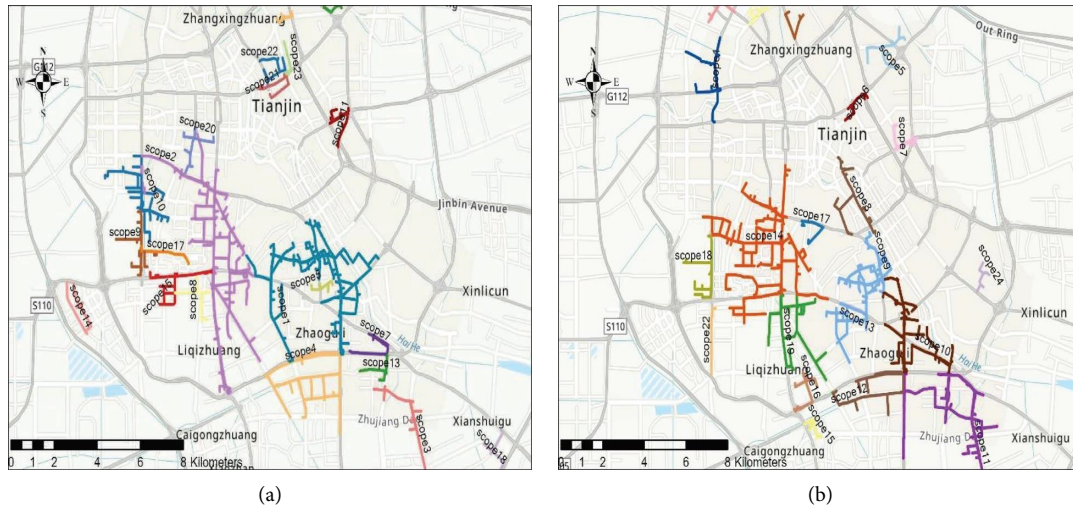
(a)

(b)

Figure 9: Service scopes of the flexible bus (a) from 8:00 to 8:30 and (b) from 18:00 to 18:30.

Table 7: Verification result of FFBS trips covered by the generated service scopes.

Date	Total number of FFBS trips during 8:00 to 8:30	Number of trips that can be covered	Number of covered trips in the same scope
2019.10.21	67,847	7,032	5,522
2019.10.23	70,221	6,840	5,451
2019.10.31	65,967	6,614	5,319

during the operation period from 8:00 to 8:30. If the origin and destination locations of one FFBS trip are both within the generated service scopes, it is considered that the trip can be replaced by flexible bus service.

As shown in Table 7, during the operation period 8:00 to 8:30, more than 6800 trips per day on average can be covered by the generated service area, accounting for about 10% of all trips in this period. In the verification experiment, the buffer radius of the generated service scopes is set to 40 meters. In other words, the trip can be seen substituted if the flexible bus can be reached within 40 meters. The buffer radius set in the experiment is small, which is maybe the reason why the flexible bus service just covers about 10% of all bicycle trips. Even if it only accounts for roughly 10% of the total, nearly seven thousand travel demands can be covered by flexible bus in half an hour period in one day is still considerable. Besides, most of the covered trips are within the same service scope. For example, on October 21, among the 7,032 trips that can be replaced by the flexible bus service, 5,522 trips can be serviced by only one service scope, accounting for 78.53%. In other words, at least 78.53% of the potential travel demand will certainly be satisfied by the flexible bus service. There are some trips that need to cross two different scopes, which may involve interchanges during the actual flexible bus operation.

Further, in order to verify the planning effect of the flexible bus service area, the ratio of the road distance involved in the service scopes to the daily road distance involved in the FFBS trajectories is calculated respectively. The

results on October 21, 23, and 31 are 2.26%, 2.29%, and 2.33%, respectively. As shown in Figure 10, although the service area proposed in this paper accounts for a relatively small road space, it can cover a relatively large proportion of travel demand during the operational period. It is worth noting that many bicycle-sharing trips are discrete and random [29]. As an intensive service, the flexible bus cannot respond to travel demands that are too discrete and occasional. Serving centralized and frequent pattern travel demands is more conducive to the practical application of flexible buses.

Besides, as an alternative transportation mode after the suspension of bicycle-sharing service in winter, flexible bus service may overlap with a shuttle bus in terms of functions. The shuttle bus is an efficient solution for last-mile transportation, mainly designed for the shuttle travel connecting the metro stations. Therefore, a further verification experiment is conducted to compare the difference in operation objectives between the flexible bus service and the shuttle bus service.

In the verification experiment, a buffer zone with a radius of 100 meters for each entrance and exit of the metro station is established, as shown in Figure 11. It is reasonable to assume that if one of the origin or destination locations of the bicycle-sharing trip is within the metro station buffer zone, the trip can be substituted by a shuttle bus service. The total number of FFBS trips that can be covered by flexible bus service during 8:00 to 8:30 in the three verification days is 20,486. As shown in Figure 11 and Table 8, only 25.80% of
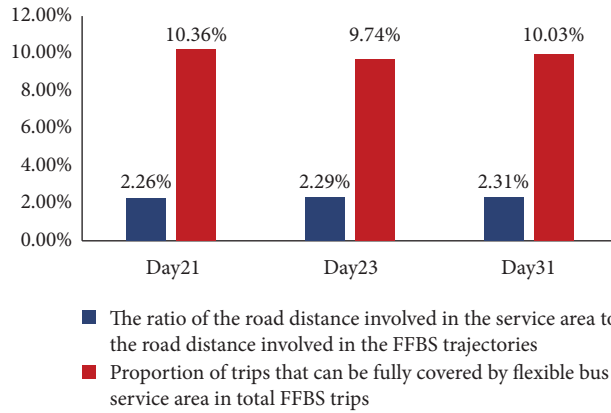
Figure 10: Planning effect of flexible bus service during 8:00 to 8:30.

- Origin location of flexible bus service covered trip
- Destination location of flexible bus service covered trip
- Buffer zone with a radius of 100 m at each entrance of metro station
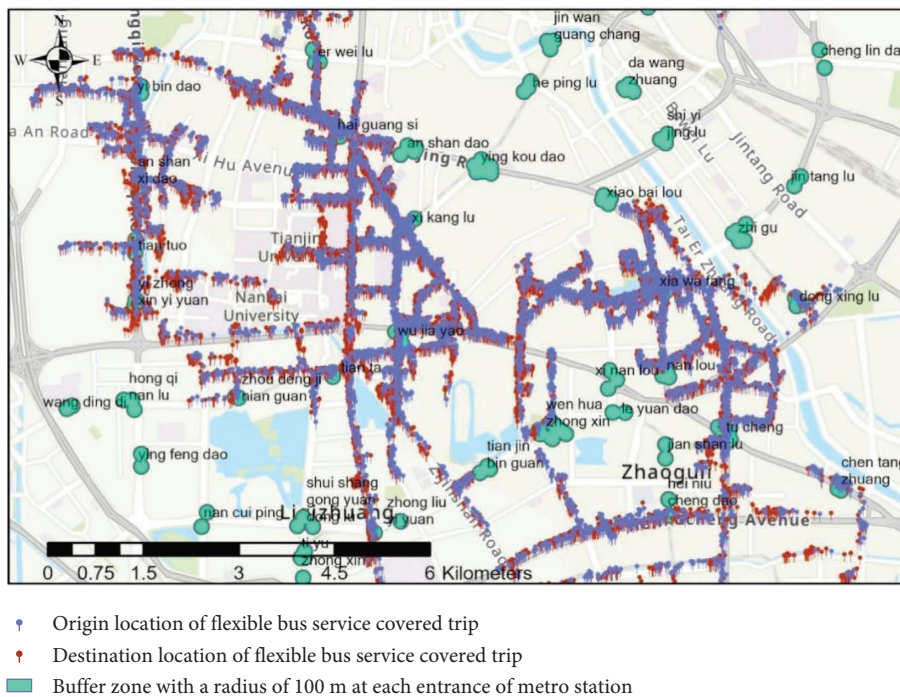
Figure 11: Visualize how many trips covered by flexible bus can fall into the service scope of shuttle bus.

Table 8: Verification results of FFBS trips covered by flexible bus that can fall into the service scope of shuttle bus during 8:00 to 8:30 in three verification days.

	Trip origin located within the buffer zone of metro station entrances	Trip destination located within the buffer zone of metro station entrances
Number of trips involved in three verification days	5,285	1,840
Proportion to the total number of flexible bus service covered trips	25.80%	8.98%

bicycle-sharing trips start in the metro station buffer zone, and 8.98% of trips end in the buffer zone, which means that there is a significant difference in target demand between the two transportation modes. The path clustering algorithm of PATHSCAN developed in this paper is based on the density-based clustering principle and considering topology between trajectories, is actually a data-driven ride-matching method, similar to carpooling routes generation, which fundamentally determines the difference between flexible bus service and shuttle bus service.

5. Conclusion and Future Work

Many FFBS operators in cold region cities suspend the bicycle-sharing services in winter, which leads to many inconveniences for people traveling short distances. The flexible bus is a suitable alternative transportation mode after the suspension of FFBS considering the travel characteristics of both short-distance and sustainable.

A data-driven method to determine the service area and operation time of flexible bus based on FFBS trip data is proposed. Different from previous studies, this paper considers flexible bus service planning from both spatial and temporal dimensions based on potential travel demand. The constructed data-driven planning method tracks personal path selection preference and travel demand aggregation degree. The core of the method is a spatio-temporal path clustering algorithm called PATHSCAN. In the defined time slice, the subroad segment with the largest number of bicycle-sharing trip starting points is selected as the core original subroad segment, which is the initial clustering center. The riding paths that start from the current core original subroad segment, and the number of trajectories involved larger than the threshold, will be extracted into the cluster. Moreover, then expand clusters from the next core subroad segments in the current path cluster. After that, a frequent pattern mining algorithm is applied to the multiday path clusters, and frequent pattern results with spatio-temporal correlation are merged into the final service area composed of multiple service scopes.

The data used for planning comes from the nonwinter season, while the actual travel in winter may change. For this reason, this study aims at mining only the high-frequency riding paths to generate planning results. The generated spatio-temporal planning results aims at to solve a large number of relatively clustered potential short-distance travel demands after the suspension of the bicycle-sharing service. These potential short-distance travel demands can be considered to have relatively high conversion probability. Although it is difficult to determine the demand transfer rates, compared with many other demand-responsive transit services that lack a potential demand basis in planning, the potential demands driven planning in this study can ensure the success of the actual operation of flexible bus services to a certain extent. Besides, similar to the actual planning and operation of other transportation modes, the initial planned operation area and time are not unchangeable, and can be adjusted according to specific actual demands and operation conditions after being put into operation. Besides, the

generated service area will vary with working days, weekends, and different periods of time to cope with the travel characteristics of large fluctuations.

The application analysis based on the actual FFBS trip data in Tianjin showed the detailed process of service planning. A verification experiment was further conducted, which proved that there are still a large number of trips within the generated service area in cold weather. Besides, although the generated service area accounts for a relatively small road space, it can cover a relatively large proportion of travel demand during the operational period. Another experiment result verified the significant difference between flexible bus service and shuttle bus service in operation objectives. However, the experimental data set used in this study is relatively limited, if a longer time dimension data set can be applied to the experiment, more convincing results will be obtained.

This paper studies the first stage of alternative flexible bus planning after bicycle-sharing suspension in cold regions in winter. Next, this study will continue to study how to operate and dispatch such a customized bus within the generated service area and operation time. How to serve the travel demand that needs to cross two different flexible bus service scopes will also be studied in the future. In addition, the combination of flexible bus services and other public transit modes in the city is also desirable to research.

Data Availability

The bicycle-sharing data used to support the findings of this study were supplied under license and so cannot be made freely available. For detailed questions related to data access, please contact the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Y. Tian, X. Zhang, B. Yang, J. Wang, and S. An, "An individual-based spatio-temporal travel demand mining method and its application in improving rebalancing for free-floating bike-sharing system," *Advanced Engineering Informatics*, vol. 50, pp. 101365–101416, 2021.

[2] W. El-Assi, M. Salah Mahmoud, and K. Nurul Habib, "Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto," *Transportation*, vol. 44, no. 3, pp. 589–613, 2017.

[3] J. Mattson and R. P. Godavarthy, "Bike share in Fargo, North Dakota: keys to success and factors affecting ridership," *Sustainable Cities and Society*, vol. 34, pp. 174–182, 2017.

[4] B. Kutela and E. Kidando, "Towards a better understanding of effectiveness of bike-share programs: exploring factors affecting bikes idle duration," *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 29, no. 1, pp. 33–46, 2017.

[5] H. Wang and A. Odoni, "Approximating the performance of a "last mile" transportation system," *Transportation Science*, vol. 50, no. 2, pp. 659–675, 2016.

[6] H. Wang, "Routing and scheduling for a last-mile transportation system," *Transportation Science*, vol. 53, no. 1, pp. 131–147, 2019.

[7] L. C. Tong, L. Zhou, J. Liu, and X. Zhou, "Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 451–475, 2017.

[8] G. Qiu, R. Song, S. He, W. Xu, and M. Jiang, "Clustering passenger trip data for the potential passenger investigation and line design of customized commuter bus," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3351–3360, 2019.

[9] Y. Lyu, C. Y. Chow, V. C. Lee, J. K. Ng, Y. Li, and J. Zeng, "CB-Planner: a bus line planning framework for customized bus systems," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 233–253, 2019.

[10] C. Park, J. Lee, and S. Y. Sohn, "Recommendation of feeder bus routes using neural network embedding-based optimization," *Transportation Research Part A: Policy and Practice*, vol. 126, pp. 329–341, 2019.

[11] X. Lu, J. Yu, X. Yang, S. Pan, and N. Zou, "Flexible feeder transit route design to enhance service accessibility in urban area," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 507–521, 2016.

[12] Y. Zheng, W. Li, and F. Qiu, "A slack arrival strategy to promote flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 442–455, 2018.

[13] S. Wang, J. Wang, L. Song, Z. Niu, and X. Hu, "Design of two-regional flexible-route bus systems considering interregional demands," *Journal of Advanced Transportation*, vol. 2022, Article ID 9372011, 2022.

[14] M. Edward Kim, J. Levy, and P. Schonfeld, "Optimal zone sizes and headways for flexible-route bus services," *Transportation Research Part B: Methodological*, vol. 130, pp. 67–81, 2019.

[15] M. E. Kim and A. Roche, "Optimal service zone and headways for flexible-route bus services for multiple periods," *Transportation Planning and Technology*, vol. 44, no. 2, pp. 194–207, 2021.

[16] L. Wang, S. C. Wirasinghe, L. Kattan, and S. Saidi, "Optimization of demand-responsive transit systems using zonal strategy," *International Journal on the Unity of the Sciences*, vol. 22, no. 3, pp. 366–381, 2018.

[17] S. M. Nourbakhsh and Y. Ouyang, "A structured flexible transit system for low demand areas," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 204–216, 2012.

[18] S. Pan, J. Yu, X. Yang, Y. Liu, and N. Zou, "Designing a flexible feeder transit system serving irregularly shaped and gated communities: determining service area and feeder route planning," *Journal of Urban Planning and Development*, vol. 141, no. 3, 2015.

[19] P. Shu, Y. Sun, B. Xie, S. X. Xu, and G. Xu, "Data-driven shuttle service design for sustainable last mile transportation," *Advanced Engineering Informatics*, vol. 49, pp. 101344–101416, 2021.

[20] S. Gaffney and P. Smyth, "Trajectory clustering with mixtures of regression models," in *Proceedings of the fifth ACM SIGKDD International conference on Knowledge discovery and data mining*, San Diego, CA, USA, August 1999.

[21] F. Liu, W. Bi, W. Hao, F. Gao, and J. Tang, "An improved fuzzy trajectory clustering method for exploring urban travel patterns," *Journal of Advanced Transportation*, vol. 2021, Article ID 6651718, 2021.

[22] H. Chen, C. Yang, and X. Xu, "Clustering vehicle temporal and spatial travel behavior using license plate recognition data," *Journal of Advanced Transportation*, vol. 2017, Article ID 1738085, 2017.

[23] G. Reyes, L. Lanzarini, W. Hasperué, and A. F. Bariviera, "Proposal for a pivot-based vehicle trajectory clustering method," *Transportation Research Record*, vol. 2676, no. 4, pp. 281–295, 2022.

[24] Z. Hong, Y. Chen, H. S. Mahmassani, and S. Xu, "Commuter ride-sharing using topology-based vehicle trajectory clustering: methodology, application and impact evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 573–590, 2017.

[25] Z. Hong, Y. Chen, and H. S. Mahmassani, "Recognizing network trip patterns using a spatio-temporal vehicle trajectory clustering algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2548–2557, 2018.

[26] C. Chen, Z. Ye, F. Hu, S. Gong, L. Sun, and Q. Yu, "Vehicle trajectory-clustering method based on road-network-sensitive features," *Journal of Intelligent and Fuzzy Systems*, vol. 41, no. 1, pp. 2357–2375, 2021.

[27] Y. Han, Y. Kim, J. Ku, Y. Jung, and J. Roh, "Map matching algorithm for real-time data processing of non-route GPS data in seoul," *KSCE Journal of Civil Engineering*, vol. 25, no. 9, pp. 3511–3522, 2021.

[28] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 855–866, 2019.

[29] B. Yang, Y. Tian, and J. Wang, "How to improve urban transportation planning in big data era? A practice in the study of traffic analysis zone delineation," *Transport Policy*, vol. 127, pp. 1–14, 2022.