INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 17, Issue: 1, Month: February, Year: 2022
Article Number: 4673, https://doi.org/10.15837/ijccc.2022.1.4673

CCC Publications

Travel preference of bicycle-sharing users: A multi-granularity sequential pattern mining approach

Y. Zhou, M.D. Zhang, G. Kou, Y.M. Li

Yu Zhou
College of Economics and Management
Inner Mongolia University, China
No. 235, College Road, Hohhot, 010021, China
zhouyu738@126.com

Mengdie Zhang
College of Economics and Management
Inner Mongolia University, China
No. 235, College Road, Hohhot, 010021, China
imuzmd@163.com

Gang Kou*
Business School of Chengdu University, China
No. 2025, Chengluo Avenue, Chengdu, 610106, China
*Corresponding author: 1067026647@qq.com

Yiming Li
College of Economics and Management
Inner Mongolia University, China
No. 235, College Road, Hohhot, 010021, China
liyiming_bit@foxmail.com

Abstract

Public bicycles are an indispensable part of green public transportation and are also a convenient and economical manner for the general public. In operation management, it is very important and imperative to understand the user demand and pattern of the public bicycle system. This paper took the public bicycle system in Hohhot as the research object, collected nearly 4 years of operating data, and studied the travel preferences of users in the public bicycle system in view of multiple granularities. Specifically, the data of car rental users at three time-granularities were obtained through data extraction technology. Finally, frequent pattern mining was performed on car rental data based on different time granularities and mapped to the user's riding preference, and then the riding modes of different car rental users founded on different time granularities were determined. Finally, this article gave different management opinions based on the different riding preferences of public bicycle users in Hohhot.

Keywords: public bicycle system, user riding preference, frequent pattern, sequential pattern, multi-granularity.

1 Introduction

On September 22, 2020, General Secretary Xi Jinping officially stated in the general debate of the 75th United Nations General Assembly that China's carbon dioxide emissions will strive to reach the peak by 2030 and achieve carbon neutrality by 2060. In the 2021 government work report, "better performance in carbon peaking and carbon neutrality" is listed as one of the key tasks in 2021 as well. To promote low-carbon development, green manufacturing must be steadily improved in the industrial field; energy-saving standards must be gradually heightened in the construction field. Meanwhile, green and low-carbon transportation methods must be accelerated in the field of green transportation [5, 12]. For green transportation, it is imperative to place energy conservation in the first place, implement a comprehensive conservation strategy, and advocate a simplified and moderate, green and low-carbon lifestyle [19]. Therefore, green and low-carbon travel will surely become an indispensable way for urban transportation to attain the national strategic call of "carbon peak and carbon neutrality". Green and low-carbon travel will also continue to be a trending topic again.

Compared with the rapid development of urban motorization, the supply of urban transportation is far from sufficient. Road construction and management cannot meet the rapid growth of motor vehicles, which has brought a series of urban traffic problems such as congestion, accidents and environmental pollution. Urban public transportation is an economical, fast and efficient environment-friendly transportation mode, which includes rail transit, conventional buses, taxis, public bicycles and some others [31]. Urban public transportation can bring some advantages including low per capita resource occupation, high operation efficiency and low environmental pollution. As a consequence, prioritizing the development of public transportation and increasing the share of public transportation trips can effectively alleviate urban traffic problems. As a mode of public transportation, public bicycles precisely have the advantages of zero pollution, zero emissions and rather strong accessibility. Meanwhile, it can provide strong support for residents to have the daily commute, leisure, entertainment, shopping and other activities near the community, provide more alternatives for citizen's travelling [3]. On the other hand, it can also be used as a connection method for other public transportation modes such as rail transit and conventional buses, to make up for their inadequate accessibility and opportunity to provide "door-to-door" full-course transportation services. "One kilometer" and "last kilometer" issues can be solved effectively. In addition to this, the literature [15, 17, 21, 22] found that public bicycles also can reduce greenhouse gas emissions, which is exactly in line with the national energy-saving emission reduction and sustainable development strategy.

When local governments vigorously promote the use of public bicycles for travel, some docked public bicycle-sharing system (PBS for short) has been built. But some inevitable problems have gradually been exposed along with the continuous expansion of its scale. For instance, the user requirements are random and uncontrollable on a regular basis. Especially during peak hours, the demand for bicycles will be quite different. For example, the demand for returning bicycles at stations near the business district on weekends is often greater than that in working areas. Some stations often appear completely empty. This situation is rather intensely intractable, which causes users to fall into the dilemma of "difficulty in renting and returning a bicycle", which cannot meet the user requirements during peak hours. Based on this, public bicycle managers and a large number of scholars optimize to solve these problems by studying public bicycle scheduling, feature analysis and demand forecasting.

The PBS records the time and location of the user's renting and returning bicycle. With the continuous advancement of data mining technology and visualization technology, data analysis of renting and returning bicycle behavior has attracted extensive attention from the academic community. Most of the research work focuses on analyzing the user's trip time and space characteristics, bicycle scheduling optimization, demand forecast and some other analyses based on the user's bicycle renting and returning records [30]. Following, the literature will be reviewed from the time granularity perspective, namely month-based long cycle, week-based medium cycle and hour-based short cycle.

In the month-based long cycle, Chen et al. [4] explored social activities in the city through public bicycle renting and returning dat and inferred the location and duration of relevant social activities based on series associated data. Kumar [16] aimed to identify important factors of selecting a public bicycle for tripping in the long-term granularity. Fournier et al. [6] constructed a model for predicting bicycle trips under the influence of different seasons, and found that the model conforms to the form of a simple sine function. Zhang et al. [37] explored the annual changes in the use of public bicycles and the impact of the PBS expansion. Finally, they proposed that the factor of the season should be considered when exploring the number change of public bicycle users. Liu et al. [20] proposed a new method of mining virtual sites founded on a multi-scale geographic perspective, which not only facilitates people to use shared bicycles for green travel but also can make shared bicycles more visible and enhance citizen's green travel awareness.

Regarding the medium cycle of weekdays and weekends, Kaspi et al. [10] proposed a Bayesian model to detect the probability and number of unavailable bicycles in PBS stations. Zhang et al. [38] presented the statistical analysis of the renting and returning conditions of different stations on weekends and weekdays, and the PBS station characteristics were summed up.

For the short cycle of time points during a day, Zhang et al. [36] provided the PBS user's trip characteristics analysis at different time periods during a day and found that PBS user's travel purposes are mainly concentrated in shopping, entertainment, commuting, and school-based the five-month data from Zhongshan City, China. Zhou et al. [41] studied the Spatio-temporal patterns of public bicycles based on the Chicago data in 2013-2014. And they also compared the flow patterns of public bicycles between the morning peak and evening peak hours on weekdays, along with the public bicycle trips of tourists and residents on weekends. Meanwhile, the distribution of active hotspots during working hours through a visual analysis was also exhibited. In some other research, for example, literature [28] and [29], the flow of bicycles in the morning peak and evening peak was shown by heat map. And each station of PBS can be clustered based on the statistical characteristics and temporal-spatial patterns of users' renting and returning behavior. Borgnat et al. [2] and Kaltenbrunner et al. [9] applied time series analysis to predict the user demand in hours based on the PBS in Lyon and Barcelona respectively. They first modeled the user travel patterns and the typical usage patterns of bicycles in PBS, then the time series model was used to predict the user demand. Li et al. [17] found that the user riding time in the morning peak hour on weekdays was shorter than that in the evening peak hour, and the riding time in the urban area was higher than that in the suburban area. Vogel et al. [32] trained a model to predict the user demand based on the data of the Vienna PBS. They constructed the time series model of the hourly renting volume with two parts, which contained a relatively fixed part and a given fluctuations part.

Mateo-Babiano et al. [23] conducted their research on medium and short operation cycles of PBS. Through the analysis of the peak hours during the day and demand patterns during a week, they found that travel speed and travel frequency were highest in the morning and evening peak hours. Meanwhile, the three-peak mode of a weekday was more obvious than a weekend. Randriamanamihaga et al. [27] combined medium and short operation cycles of PBS to explore the time change of Dutch demand for bicycles. Finally, they found that rainfall has a greater impact on weekend entertainment needs, but has no significant impact on workday commuting needs. Xing et al. [35] combined PBS medium and short operation data with points of interest to study the travel mode and purpose of shared bicycles based on the k-means++ clustering method. The results manifested that the place of departure and destination on weekdays can be divided into five typical areas: dining, transportation, house, company, and shopping. But to the best of our knowledge, there are very few studies on the frequent patterns of PBS users considering three time-granularity simultaneously, namely month-based long cycle, week-based medium cycle and hour-based short cycle. Therefore, it is imperative to mine the frequent patterns of public bicycle users based on multiple time granularity. The analysis results can not only provide a basis for user travel forecasts and a reference for reasonable operation scheduling but also provide a basis for how to release user needs. Therefore, this paper will conduct a thorough analysis of the frequent patterns of public bicycle users and explore the long-term cycling habits of different users and recent trends inhabits as well as cycling preferences. Finally, the effectiveness of the proposed analysis method will be validated by the real cycling data of public bicycle users in Hohhot as an instance.

2 Data description and problem definition

2.1 Data and background

As the urban traffic congestion becomes more and more serious, the problem of the deterioration of the traffic environment will emerge. The residents' environmental awareness and daily travel mode have also been greatly changed. In order to promote the construction of the urban bicycle transportation network system, the PBS of Hohhot was officially put into practice on October 1, 2013, and it is still in operation today. Judging from the regional distribution, the distribution of PBS has covered the city's 4 main urban areas. As an emerging mode of transportation, public bicycles have their convenient and economic advantages contrasted to other modes of transportation in short and medium-distance travel and public transportation connections. The emergence of the PBS has promoted the optimization of the traffic travel structure in Hohhot. The travel sharing rates of bicycle travel, chronic traffic and green traffic have been increased to a certain extent, and the travel and carbon emissions of private cars have been reduced [34]. This paper collects the cycling data of the public bicycle system in Hohhot City from October 1, 2013 to August 31, 2017 for the past four years to conduct research on the frequent patterns of PBS users. During these four years, the public bicycle system has put into function a total of 9,437 bicycles, with a total of 317 stations, more than 130,000 registered users, and more than 36.34 million cycling data.

If you want to obtain information about the user's travel characteristics, you must first organize the data format according to the analysis requirements. For a large amount of IC card swiping data, due to data collection, transmission, and collection errors, where data errors occur, and the wrong data needs to be eliminated. Error data includes data with incomplete borrowing and returning records, data with negative riding time, and data with the same station being borrowed and returned within 2 minutes, etc. In addition, for effective trips, the distance of each trip is more than 500 meters, the average speed of bicycles is 10km/h, and the calculated travel time is greater than 180 seconds. The upper limit of travel time is 2 hours, which is 7200 seconds. Exclude the travel records with travel time less than 180 seconds and greater than 7,200 seconds. The data structure after removal is shown in Table 1.

Table 1: Data structure after cleaning

User ID	Bicycle ID	Trip origin	Trip destination	Trip start time	Trip end time
3001570	132454	4003-7	1004-15	2013/10/31 20:38	2013/10/31 20:52
3005246	132478	1005-27	1004-15	2013/10/31 20:38	2013/10/31 20:59
...	...	...	...	...	...
3000221	130065	2013-3	4028-12	2013/10/31 20:27	2013/10/31 20:47
3005116	130101	4028-6	1030-35	2013/10/31 20:27	2013/10/31 20:46
...	...	...	...	...	...

In order to describe the riding habits and riding patterns of different PBS users accurately, it is fundamental to use Python software to perform statistics on the cleaned data at different time granularities. According to the provided data set, it can be found that each user does not use a bicycle every day within a month, so there will be a situation where the data is empty on a certain day. For this part of the vacant data, it is requisite to fill in the vacant part. The value 0 means that the renting times in the current month (day/time period) are zero. In the month-based long cycle, the statistical data is shown in Table 2.

Table 2: Extracted data under month-based long cycle

User ID	January 2014	February 2014	...	July 2017	August 2017
3001570	10	15	...	27	23
3005246	13	17	...	21	26
...	...	...	...	...	...
3000221	15	23	...	30	37
3005116	11	42	...	27	46
...	...	...	...	...	...

In order to analyze the law of the change of use demand with the month, a trend chart of the change in the frequency of bicycle rental numbers over four years is drawn. As shown in Figure 1, the frequency of public bicycles rental has a periodicity, which shows an inverted "U" shape and the cycle is one year. Looking at the overall trend, the frequency of public bicycle rental has increased year by year. And the increase in the frequency of public bicycles rental in 2017 has increased significantly, which indicates that users' demand for public bicycles is increasing year by year.

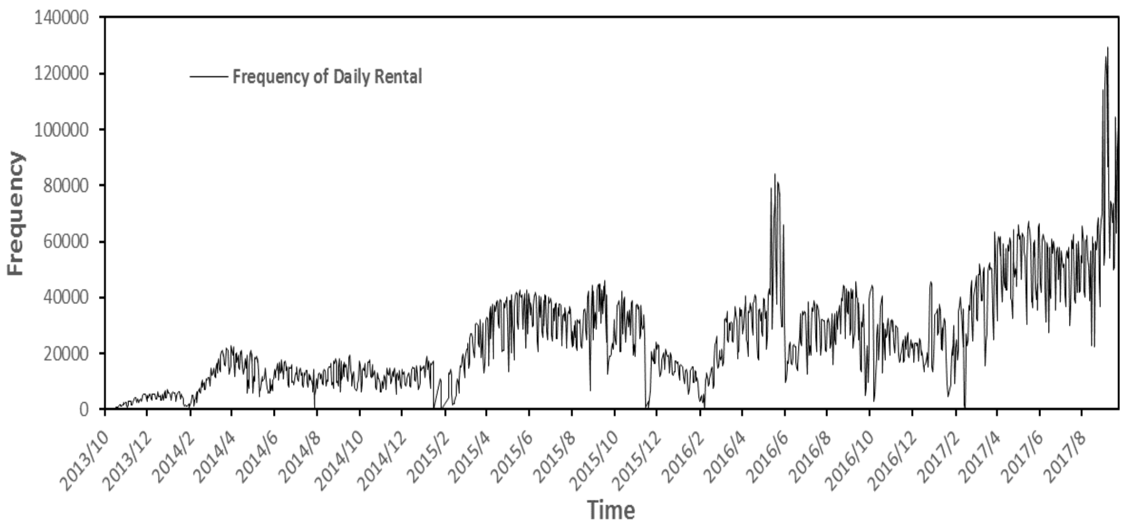

Figure 1: Travel frequency of users of the public bicycle system

The data extracted from the medium cycle of weekdays and weekends are shown in Table 3. Figure 2 exhibits the PBS user needs for 24 hours during a day from Monday to Sunday in 2016. It can be witnessed from the figure that the user needs from workdays are similar, and the distribution pattern shows obvious peaks in the morning, middle and evening. This may be caused by the purpose of commuting. It can also be seen that the demand on workdays is much greater than that on weekends.

Table 3: Extracted data under month-based long cycle

User ID	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
3001423	10	15	18	27	23	2	0
3001556	26	17	22	11	2	0	0
...	...	...	...	...	...	...	...
3004557	1	0	2	3	52	61	40
...	...	...	...	...	...	...	...

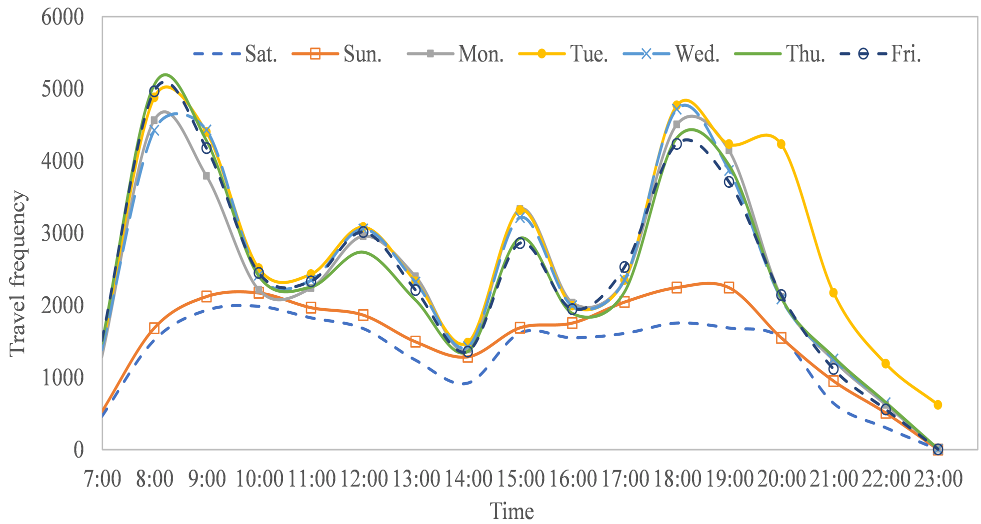

Figure 2: Travel frequency of users of the public bicycle system

For the short cycle, the day is divided into six time periods, namely 6:00-9:00, 9:00-11:00, 11:00-14:00, 14:00-16:00, 16:00-19:00, 19:00-23:00. Table 4 displays the data extracted in six time periods.

Table 4: Extracted data under month-based long cycle

User ID	**6:00-9:00**	**9:00-11:00**	**11:00-14:00**	**14:00- 16:00**	**16:00-19:00**	**19:00-23:00**
3001423	45	0	0	2	23	15
3001556	17	1	41	0	19	0
...	...	...	...	...	...	...
3004557	5	50	32	36	35	1
...	...	...	...	...	...	...

As mentioned earlier, the PBS user demand demonstrates three obvious peaks in the morning, middle and evening. It is evident from Figure 2 that the morning peak hour is around 7:30-9:00, the noon peak hour is around 11:00-13:30, and the evening peak hour is about 17:00-19:00. To sum up the above mentions, the PBS user demand pattern can be depicted under multi-time granularity. As a consequence, a multi-granularity sequential pattern-based user preference analysis for the PBS system will be presented in the follow-up research. With the analysis results, management opinions should be provided based on the different riding preferences, which will not only help improve the user's travel experience but also improve the economic benefits of the PBS.

2.2 Problem definition

It is very important to study the travel modes at different time-granularities for analyzing the PBS user travel preferences and improving their travel experience. This study assumes that the effect of the business promotion activities of the PBS on the user demand can be negligible. And the effect of all external factors, such as the influence of the environment, similarly can be ignored. The effect of these factors on PBS user travel preferences may be insignificant on the one hand, on the other hand, the associated data is very hard even fail to collect. In order to obtain the riding data under the three time-granularities, it is necessary to clean the original rough data $\{T_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., m\}$. Where T_{ij} represents the j-th riding time of the i-th user, such as $\{T_{11} = 20 : 38\ 2013/10/31\}$ represents the first riding time of user 3001570 (user ID). After extracting, loading and transforming, the data under different time granularity should be obtained, which can be defined as a variable as $\{S^k_{il}, i = 1, 2, ..., n; l = 1, 2, ..., p_k; k = 1, 2, 3\}$. Where i represents the user, $k = 1, 2, 3$ marks the time granularities namely month-based long cycle, week-based medium cycle and hour-based short cycle, while l is the observation points under the given time granularity k. For example, user 3001570 has ridden a public bicycle a total of 362 times in the past 5 months. And he has almost zero rides on weekends. On workdays, he mostly rides around 8:00 am and 17:00 pm, and occasionally rides at noon. It is reasonable to assume that the user preference under different time granularity exists and is very significant. So the i-th user riding preference (URP for short) under different time granularity can be represented as $\text{URP}_i = \{M^1_x, M^2_y, M^3_z\}$, where M^1_x, M^2_y, M^3_z is defined as the user riding patterns under k time granularity. Following, the frequent pattern mining will be performed on mining the riding patterns of PBS users at different time granularities.

3 Analysis of URP under multi-granularities

In this section, the proposed analysis method of URP will be introduced in detail. And the basic spectral clustering and prefixSpan algorithm are also described in the proposed analysis method.

3.1 Proposed URP analysis framework

The research framework of this paper is manifested in Figure 3. In order to obtain data in different formats, the cleaned data set $\{T_{ij}, i = 1, 2, ..., n; j = 1, 2, ..., m\}$ is first analyzed and processed online to form a three-dimensional data cube [7, 25]. Then the data cube set is sliced and drilled to obtain the three-time granularity data set $\{S^k_{il}, i = 1, 2, ..., n; l = 1, 2, ..., p_k; k = 1, 2, 3\}$, which indicate the rental number of the PBS users under the time granularities namely month-based long cycle, week-based medium cycle and hour-based short cycle.

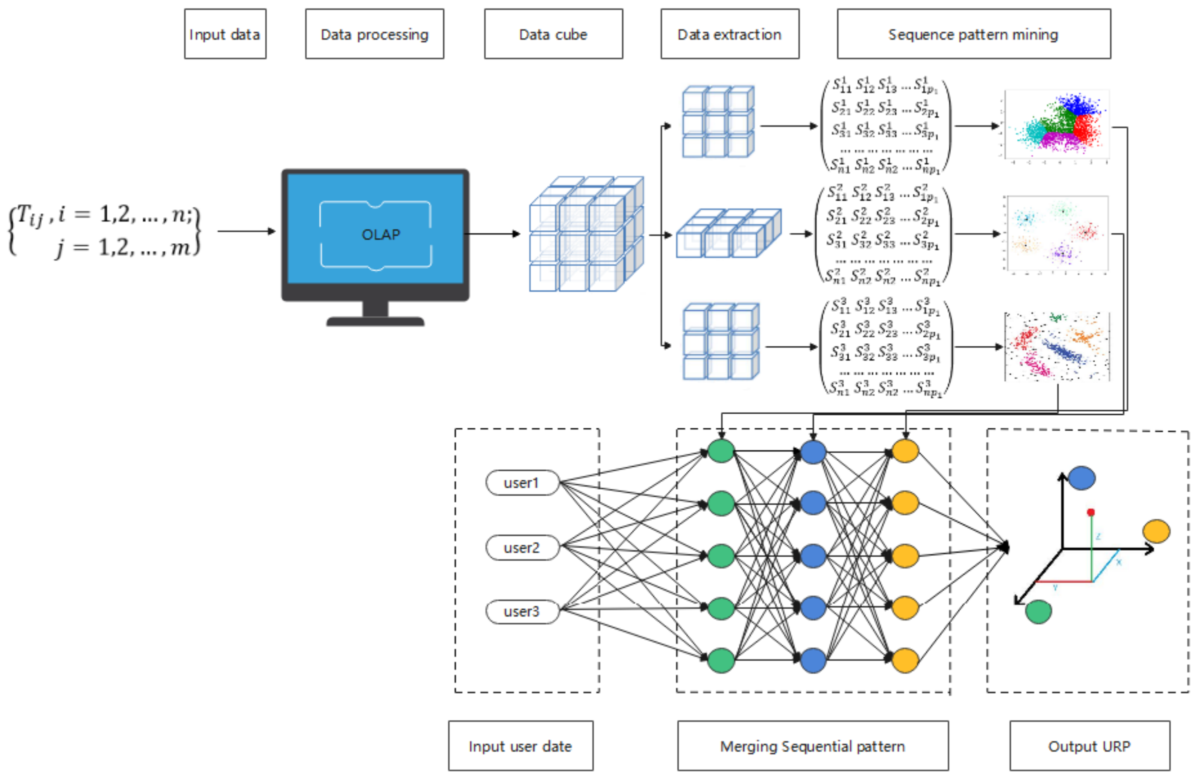

Figure 3: Proposed multi-granularity sequential pattern-based URP analysis framework

Based the three-time granularity data set $\{S^k_{il}, i = 1, 2, ..., n; l = 1, 2, ..., p_k; k = 1, 2, 3\}$, frequent pattern mining are used to find the user's riding preferences on three-time granularity data set [1, 18, 33]. Finally, the user's riding preferences should be mapped. In the existing research, there are many applications of data mining, blockchain and machine learning [11, 39, 40]. In this paper, the spectral clustering and PrefixSpan algorithm will be drawn into the proposed multi-granularity sequential pattern-based URP analysis framework, which will be introduced as follows.

3.2 Sequence pattern mining

The method of spectral clustering is based on the optimal partitioning theory of classical spectral graphs in graph theory physics [8, 13, 26]. In its essence, it transforms the problem of clustering and segmentation of the data set into the problem of optimal graph partitioning. The spectral clustering algorithm can quickly find clusters of any shape and obtain its convergence to the global optimal partition of the data set. The algorithm of spectral clustering uses the data set samples as vertices, and the similarity of the data set samples as the weights of all connected edges between the vertices. An undirected weighted graph will be obtained, which is used to represent the data set samples and their similarity to distinguish the clusters of the data sets. So, the problem of class segmentation is transformed into the problem of segmentation of directed weighted graphs.

The spectral clustering algorithm can be described as Table 5, where the month-based long cycle data set is taken as an example.

Table 5: Standard spectral clustering algorithm

Input: sample matrix $\{S^1_{il}, i = 1, 2, ..., n; l = 1, 2, ..., p_1\}$ and the number of clusters c
Output: clusters M^1_1, M^1_2, ..., M^1_x
Step 1. construct the sample similarity matrix Q
Step 2. construct adjacency matrix W and degree matrix D based on sample similarity matrix Q
Step 3. calculate the Laplacian matrix $L = D - W$ and the normalized Laplacian matrix $D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$
Step 4. calculate the smallest x' eigenvalues and corresponding eigenvectors f
Step 5. cluster eigenvectors f using k-means algorithm
Step 6. output user riding preference clusters M^1_1, M^1_2, ..., M^1_x

In addition to the spectral clustering algorithm, another clustering algorithm named the PrefixSpan algorithm will also be applied. The goal of the PrefixSpan algorithm is to discover frequent sequences that meet a predetermined threshold [5, 24]. The frequent sequence can be obtained through searching the corresponding projection database. Similar to the Apriori algorithm, it starts by mining patterns from sequence data with a prefix length 1. The frequent sequence of prefix length 1 will be obtained by searching the corresponding projection database. Then recursively mines of the frequent sequence will be carried based on the prefix length plus 1, which will be terminated until no longer prefix can be mined [14, 24]. The spectral clustering algorithm steps are presented in Table 6, where the hour-based short cycle data set is taken as an example.

Table 6: Standard PrefixSpan algorithm

Input: sequence data set $\{S^3_{il}, i = 1, 2, ..., n; l = 1, 2, ..., p_3\}$ and support threshold α
Output: all frequent sequence sets M^3_1, M^3_2, ..., M^3_z that meet the support threshold α
Step 1. find out all prefixes of length 1 and the corresponding projection database
Step 2. count the found prefixes and delete the items corresponding to the prefixes whose support is lower than the threshold α in the data set. All frequent one-item sequences are obtained.
Step 3. regarding each projection as a sequence of data, find out all frequent sequence patterns of length 1 in it. Then merge the frequent one-item sequence in the projection with the prefix to form a new prefix and increase the frequent sequence pattern.
Step 4. mining the frequent sequential patterns on those databases recursively.
Step 5. output frequent sequential patterns M^3_1, M^3_2, ..., M^3_z

After obtaining the riding data of PBS rental data at different time granularities, the PrefixSpan and spectral clustering algorithm are function to mine the URP under three time granularities. For example, the month-based long cycle, week-based medium cycle and hour-based short cycle riding preference is represented by $\{M^1_x\}, \{M^2_y\}, \{M^3_z\}$ respectively. For a given PBS user i, based on whose riding data, the URP under different time granularity can be mapping as $\mathrm{URP}_i = \{M^1_x, M^2_y, M^3_z\}$.

4 URP analysis of PBS users in Hohhot

This section will analyze the user riding preferences of PBS in Hohhot based on the collected riding data from October 1, 2013 to August 31, 2017. During these four years, the public bicycle system has put into function a total of 9,437 bicycles, with a total of 317 stations, more than 130,000 registered users, and more than 36.34 million cycling data.

4.1 Analysis results of URP

For the URP under the month-based long cycle, the spectral clustering algorithm is adopted to model the PBS rental data. The optimal clusters of URP under the month-based long cycle are five categories $\{M_1^1, M_2^1, M_3^1, M_4^1, M_5^1\}$, which is shown as (a) to (e) in Figure 4. The average curve of the five categories is displayed in Figure 5.

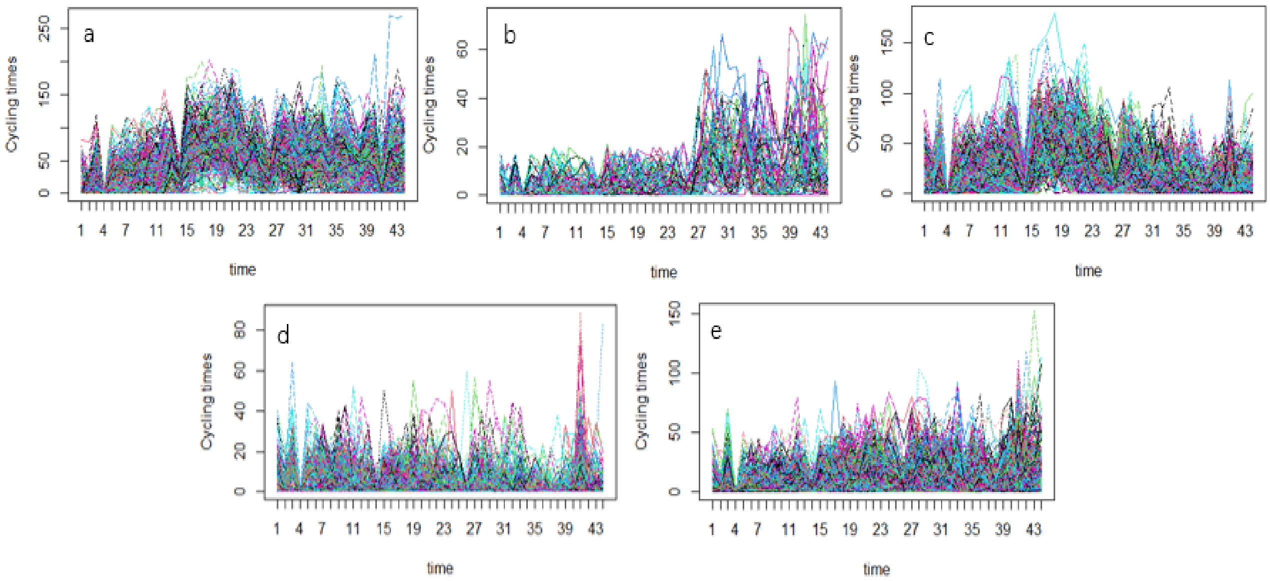

Figure 4: Optimal clusters of URP under the month-based long cycle

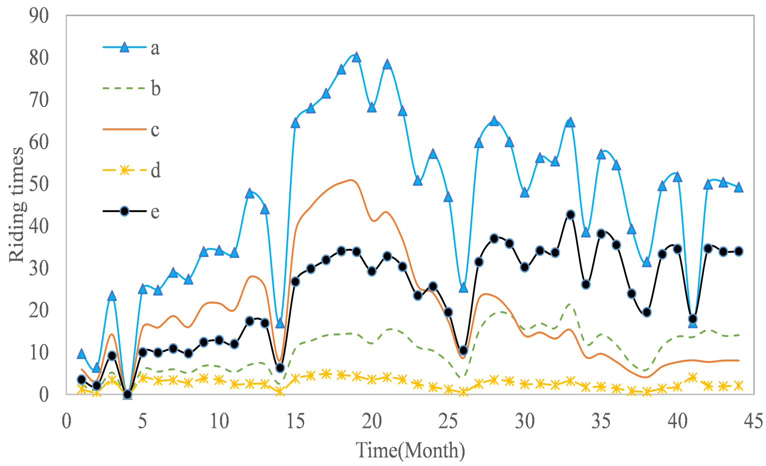

Figure 5: Average curve of the five categories under the month-based long cycle

The first URP category of Hohhot PBS has a relatively high demand than other URP categories. The PBS user demand pattern of the first category displayed a trend of rising first and then falling and remained generally stable. On the whole, PBS user demand exhibits a downward trend in winter and an upward trend in summer. The PBS user demand pattern of the second URP category has shown an upward trend year by year, but the overall demand is relatively low. Judging from the trend of the mean curve, the PBS user demand of this category reached a peak in 2016 and declined slightly in 2017, but will continue to maintain the current demand level for a period of time in the future. While the PBS user demand of the third type reached its peak in 2015 and drop suddenly. The mean curve also shows that this type of user's demand is very low and they have reduced their riding number greatly even reduce to zero. The lowest user demand is the fourth category with an average of one riding per week. This type of user can be regarded as silent users, whose demand may continue to be low in the future. However, the fifth category shows a continuation in the upward trend in PBS demand in the whole four years. The bicycle rental number of PBS users has always kept a certain frequency of 20 to 30 times a month since 2015. And the high travel demand will also trend to stable seemingly.

One of the goals of PBS is to release the pressure of commuting, so the URP under the week-based medium cycle must be presented. For the URP under the week-based medium cycle, the PrefixSpan algorithm is adopted to model the PBS rental data. The optimal clusters of URP from Monday to Sunday are eight categories $\{M_1^2, M_2^2, \ldots, M_7^2, M_8^2\}$, which is shown as (a) to (h) in Figure 6. The average curve of the eight categories is displayed in Figure 7. In the figure, the horizontal axis represents seven days from Monday to Sunday, and the vertical axis represents the daily car rental demand of car rental users.

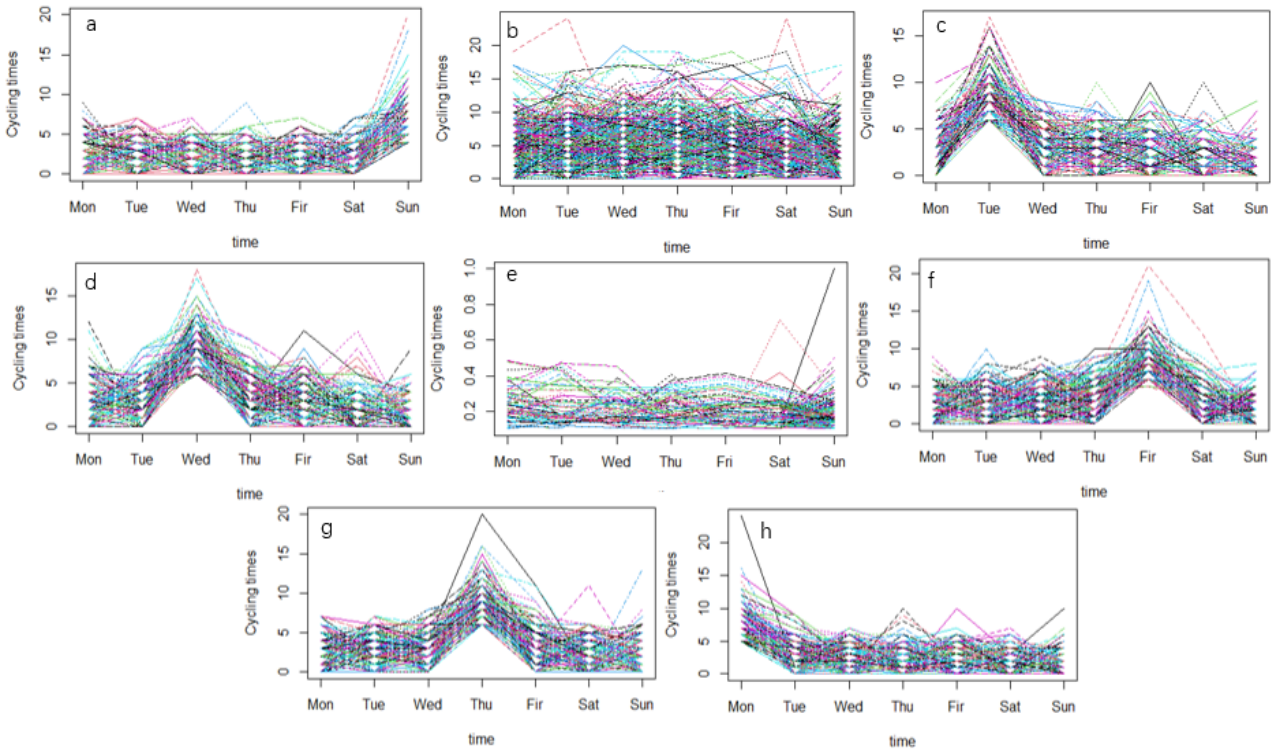

Figure 6: Optimal clusters of URP under the week-based medium cycle

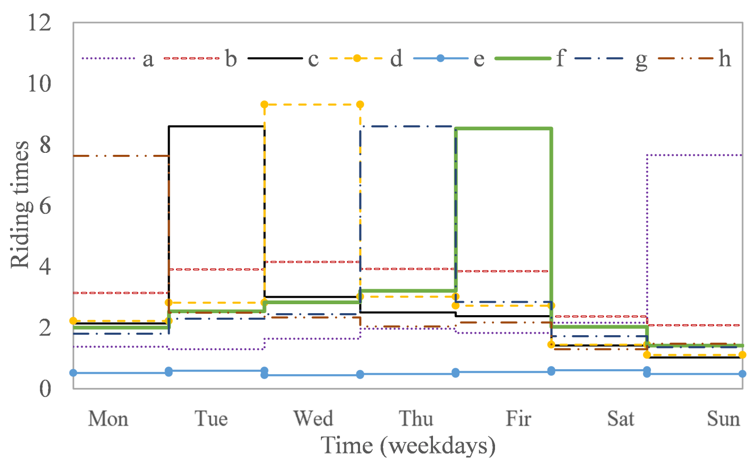

Figure 7: Average curve of the eight clusters under the week-based medium cycle

In Figure 6 and 7, the eight preference patterns of the PBS user can be further summarized into three patterns. The demand of the first pattern is medium, and the riding demand on one day during the week will be significantly higher than on other days. For example, the riding demand on Sunday is higher than other days as shown in Figure 6-(a). The other two patterns are shown in Figure 6-(b) and (e). The riding demand in these two patterns changes relatively smoothly during the weekdays, but the patterns should be differentiated by high and low riding frequency.

In the exploration of PBS user riding demand during a day, there is an obvious three peaks pattern namely morning, middle and evening peaks. In order to distinguish the demand pattern of PBS user in different periods, a day is divided into 6 periods namely 6:00-9:00, 9:00-11:00, 11:00-14:00, 14:16:00 , 16:00-19:00, and 19:00-23:00. For the URP under the hour-based short cycle, the spectral clustering algorithm is adopted to model the PBS rental data. The optimal clusters of URP under hour-based short cycle are five categories $\{M_1^3, M_2^3, M_3^3, M_4^3, M_5^3\}$, which is exhibited as (a) to (e) in Figure 8. The average curve of the five categories is displayed in Figure 9.

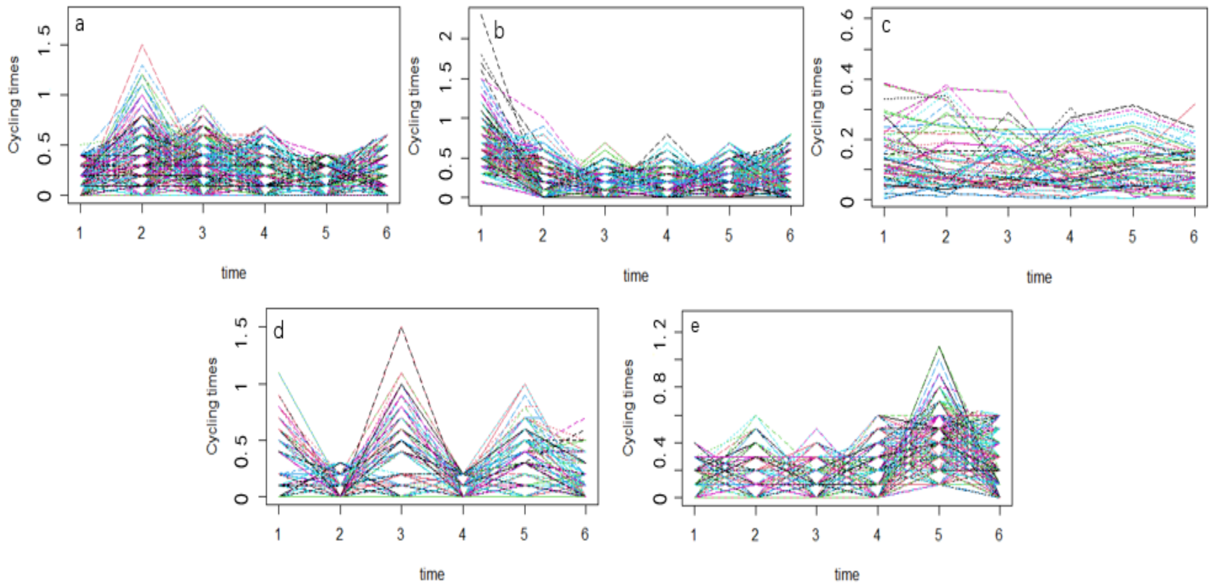

Figure 8: Optimal clusters of URP under the hour-based short cycle

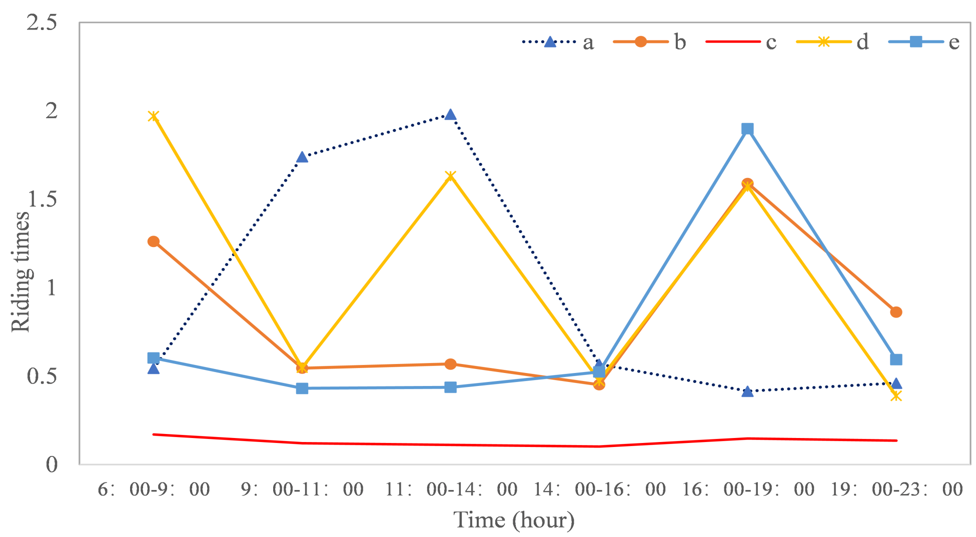

Figure 9: Average curve of the eight clusters under the hour-based short cycle

In Figure 8 and 9, the five riding demand patterns of the PBS user can be further summarized into three patterns. The first pattern can be generalized as commuting preference as displayed in 9-(b), (d) and (e). The PBS users with commuting preference have higher riding demand on three commuting peaks than other periods. All the riding demand on three commuting peaks in (b) category are significant, while which are partially significant (d) and (e) categories. The second pattern can be generalized as high demand on the morning as displayed in 9-(a). The demand in the third category is low riding frequency throughout the day.

4.2 Mapping the riding preference of Hohhot PBS users

Under the three time granularities, the riding preference of Hohhot PBS users are divided into $\{M_5^1\} \cdot \{M_8^2\} \cdot \{M_5^3\}$. So a variety of different riding preferences can be mapped, which is shown in Figure 10. For any PBS user i, based on whose riding data, the URP under different time granularity can be mapped as $\text{URP}_i = \{M_1^1, M_2^2, M_4^3\}$.

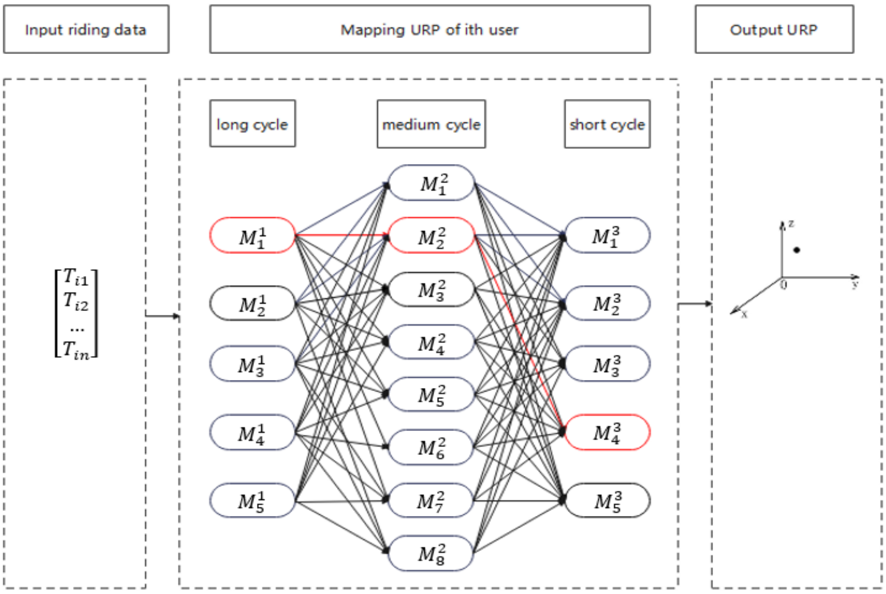

Figure 10: Merging the multi-granularity sequence pattern

5 Conclusions and management recommendations

Public bicycles are an indispensable part of green public transportation and are also a convenient and economical manner for the general public. Nonetheless, in the operation of public bicycles, in addition to planning the public bicycle system itself, it is also imperative to understand the travel requirements and travel modes of users. This paper utilized the operating data of the public bicycle system in Hohhot as research, combined with descriptive statistical analysis, spectral clustering and PrefixSpan algorithm to analyze the travel needs of car rental users, and mined the frequent patterns of users from three different time granularities. The main research progress is as follows.

First, the descriptive statistical analysis is functioned to analyze the use of the public bicycle system in Hohhot from the characteristics of car rental users and the characteristics of changes in the needs of car rental users over time. Second, the method of spectral clustering is used to analyze the characteristics of Hohhot City four consecutive times. The cluster analysis of sticky users riding for four years exhibits that different types of rental users have different needs for public bicycles, and according to the demand trend, the riding needs of each type of user in the future can be judged. Last but not least, for improving its accuracy, analyze the riding habits of car rental users, and use the PrefixSpan algorithm to mine the riding patterns of users, finding that different car rental users have different riding patterns in short cycles and fine-grained ones. Besides, combining the general characteristics of various models under different standards and the similarities and differences between the various models, a strategy on operation management was proposed.

Based on the analysis results, the following management recommendations are shown to be given out. In accordance with the various riding pattern formed by different time granularity and user, the user preference should be considered when improving the PBS scheduling and service optimization. In the first place, the analysis shows that different workdays can affect the bicycle rental number, so the manager should carry out real-time dynamic updating of the schedule based on the above-mentioned research results.

In the second place, the proportion of active users in Hohhot PBS users is about 3%. This phenomenon is also can be concluded from the URP analysis results, which display that there is a certain number of users who basically do not use public bicycles when traveling. The average rental number of public bicycles is extremely low, almost can be negligible, which can be regarded. Unfortunately, inactive users account for a considerable proportion of the total number of users. Public bicycle system operators can conduct a deepened exploration for such users, such as directional interview surveys, adopt appropriate preferential strategies to attract such users to increase their riding demand.

The last but the most important is improving the service quality. In terms of service quality, the public bicycle system should improve the overall service quality of the system in terms of improving the overall comfort of riding, ensuring comprehensive maintenance and providing timely services to customers. So as to enhance the user experience and increase the quantities of active users.

Funding

This study was supported by the National Natural Science Foundation of China (No. 71961025 and 72104114), the Natural Science Foundation of Inner Mongolia Autonomous Region (No. 2019MS07020), program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (No. NJYT-20-B08), and the Graduate Innovation Fund of Inner Mongolia University. The authors would like to thank the editors and anonymous reviewers for their constructive suggestions for improving the quality of this study.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Abdelaal, A.; Al-Shayeji, M. and Allaho, M.(2021). Customized frequent patterns mining algorithms for enhanced Top-Rank-K frequent pattern mining, *Expert Systems with Applications*, 169, 114530, 2021.

[2] Borgnat, P.; Fleury, E.; Robardet, C. and Scherrer, A.(2009). Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program, *Proceedings of European Conference on Complex Systems*, 1–6, 2009.

[3] Cerutti, P.; Martins, R.; Macke, J. and Sarate, J.(2019). "Green, but not as green as that": An analysis of a Brazilian bike-sharing system, *Journal of Cleaner Production*, 217, 185–193, 2019.

[4] Chen, L.; Ma, X.; Nguyen, T.; Pan, G. and Jakubowicz, J.(2016). Understanding bike trip patterns leveraging bike sharing system open data, *Frontiers of Computer Science*, 77, 1–11, 2016.

[5] Fisch-Romito, V.; Guivarch, C. (2019). Transportation infrastructures in a low carbon world: An evaluation of investment needs and their determinants, *Transportation Research Part D: Transport & Environment*, 72, 203–219, 2019.

[6] Fournier, N.; Christofa, E. and Knodler, M.(2017). A sinusoidal model for seasonal bicycle demand estimation, *Transportation Research Part D: Transport & Environment*, 50(1), 154–169, 2017.

[7] Gallinucci, E.; Golfarelli, M.; Rizzi, S.; Abello, A. and Romero, O.(2018). Interactive multidimensional modeling of linked data for exploratory OLAP, *Information Systems*, 77, 86–104, 2018.

[8] Ge, Y.; Lu, H. and Peng, P.(2021). Mixed-order spectral clustering for complex networks, *Pattern Recognition*, 117, 107964, 2021.

[9] Kaltenbrunner, A.; Meza, R.; Grivolla, J.; Codina, J. and Banchs, R.(2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system, *Pervasive and Mobile Computing*, 6, 455–466, 2010.

[10] Kaspi, M.; Raviv, T. and Tzur, M.(2016). Detection of unusable bicycles in bike-sharing systems, *Omega*, 65, 10–16, 2016.

[11] Keramati, A.; Ghaneei, H. and Mirmohammadi, S.(2016). Developing a prediction model for customer churn from electronic banking services using data mining, *Financial Innovation*, 2(1), 1–13, 2016.

[12] Khan, M.K.; Khan, M.I. and Rehan, M.(2020). The relationship between energy consumption, economic growth and carbon dioxide emissions in pakistan, *Financial Innovation*, 6(1), 1–13, 2020.

[13] Kou, G.; Lu, Y.; Peng, Y. and Shi, Y.(2012). Evaluation of classification algorithms using MCDM and rank correlation, *International Journal of Information Technology & Decision Making*, 11(1), 197–225, 2012.

[14] Kou, G.; Xiao, H.; Cao, M. and Lee, L.(2021). Optimal computing budget allocation for the vector evaluated genetic algorithm in multi-objective simulation optimization, *Automatica*, 129, 109599, 2021.

[15] Kou, Z.; Wang, X.; Chiu, S. and Cai, H.(2020). Quantifying greenhouse gas emissions reduction from bike share systems: a model considering real-world trips and transportation mode choice patterns, *Resources Conservation and Recycling*, 153, 104534, 2020.

[16] Kumar, D.(2021). Meteorological barriers to bike rental demands: a case of Washington D.C. using NCA approach, *Case Studies on Transport Policy*, 9(2), 830–841, 2021.

[17] Li, A.; Gao, K.; Zhao, P.; Qu, X. and Axhausen, K.W.(2021). High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data, *Journal of Cleaner Production*, 296, 126423, 2021.

[18] Li, T.; Kou, G. and Peng, Y.(2020). Improving malicious URLs detection via feature engineering: Linear and nonlinear space transformation methods, *Information Systems*, 91, 101494, 2020.

[19] Liu, T.; Wang, Y.; Li, H. and Qi, Y.(2021). China's low-carbon governance at community level: A case study in Min'an community, Beijing, *Journal of Cleaner Production*, 311, 127530, 2021.

[20] Liu, Y. and Tian, L. (2020). A graded cluster system to mine virtual stations in free-floating bike-sharing system on multi-scale geographic view, *Journal of Cleaner Production*, 281, 124692, 2020.

[21] Luo, H.; Kou, Z.; Zhao, F. and Cai, H.(2019). Comparative life cycle assessment of station-based and dock-less bike sharing systems, *Resources Conservation and Recycling*, 146, 180–189, 2019.

[22] Luo, H.; Zhao, F.; Chen, W. and Cai, H.(2020). Optimizing bike sharing systems from the life cycle greenhouse gas emissions perspective, *Transportation Research Part C: Emerging Technologies*, 117, 102705, 2020.

[23] Mateo-Babiano, I.; Bean, R.; Corcoran, J. and Pojani, B.(2016). How does our natural and built environment affect the use of bicycle sharing?, *Transportation Research Part A: Policy and Practice*, 94, 295–307, 2016.

[24] Niyazmand, T. and Izadi, I.(2019). Pattern mining in alarm flood sequences using a modified PrefixSpan algorithm, *ISA Transactions*, 90, 287–293, 2019.

[25] Panda, A. and Nanda, S.(2017). Short-term and long-term interconnectedness of stock returns in western europe and the global market, *Financial Innovation*, 7(1), 1–36, 2021.

[26] Peng, Y.; Kou, G.; Shi, Y. and Chen, Z.(2008). A descriptive framework for the field of data mining and knowledge discovery, *International Journal of Information Technology & Decision Making*, 7(4), 639-682, 2008.

[27] Randriamanamihaga, A.; Come, E.; Oukhellou, L. and Govaert, G.(2014). Clustering the Vélib' dynamic origin/destination flows using a family of poisson mixture models, *Neurocomputing*, 141, 124–138, 2014.

[28] Reiss, S. and Bogenberger, K.(2015). GPS-data analysis of munich's free-floating bike sharing system and application of an operator-based relocation strategy, *Proceedings of IEEE International Conference on Intelligent Transportation Systems*, 584–589, 2015.

[29] Schimohr, K. and Scheiner, J.(2021). Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption, *Journal of Transport Geography*, 92, 103017, 2021.

[30] Si, H.; Shi, J.; Wu, G.; Chen, J. and Zhao, X.(2019). Mapping the bike sharing research published from 2010 to 2018: a scientometric review, *Journal of Cleaner Production*, 213, 415–427, 2019.

[31] Soza-Parra, J.; Raveau, S. and Muoz, J.(2021). Travel preferences of public transport users under uneven headways, *Transportation Research Part A: Policy and Practice*, 147, 61–75, 2021.

[32] Vogel, P. and Mattfeld, D.(2011). Strategic and operational planning of bike-sharing systems by data mining: a case study, *Proceedings of International Conference on Computational Logistics*, 127–141, 2011.

[33] Wang, H.; Kou, G. and Peng, Y.(2021). Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending, *Journal of the Operational Research Society*, 72(4), 923-934, 2021.

[34] Wang, Z.; Xue, M.; Zhao, Y. and Zhang, B.(2020). Trade-off between environmental benefits and time costs for public bicycles: An empirical analysis using streaming data in China, *Science of The Total Environment*, 715, 136847, 2020.

[35] Xing, Y.; Wang, K. and Lu, J.(2020). Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai China, *Journal of Transport Geography*, 87, 102787, 2020.

[36] Zhang, Y.; Brussel, M.; Thomas, T. and Van Maarseveen, M.(2018). Mining bike-sharing travel behavior data: an investigation into trip chains and transition activities, *Computers Environment and Urban Systems*, 69, 39–50, 2018.

[37] Zhang, Y.; Thomas, T.; Brussel, M. and Maarseveen, M.(2016). Expanding bicycle-sharing systems: lessons learnt from an analysis of usage, *PLoS ONE*, 11(12), e0168604, 2016.

[38] Zhang, Y.; Wen, H.; Qiu, F.; Wang, Z. and Abbas, H.(2019). iBike: intelligent public bicycle services assisted by data analytics, *Future Generation Computer Systems*, 95, 187–197, 2019.

[39] Zhao, J.; Fan, S. and Yan, J.(2016). Overview of business innovations and research opportunities in blockchain and introduction to the special issue, *Financial Innovation*, 2(1), 1–7, 2016.

[40] Zhong, X. and Enke, D.(2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms, *Financial Innovation*. 5(1), 1–20, 2019.

[41] Zhou, X. and Chen, Y.(2015). Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago, *Plos ONE*, 10(10): e0137922, 2015.

Cite this paper as:

Zhou Y.; Zhang M.D.; Kou G.; Li Y.M. (2022). A multi-granularity sequential pattern-based user preference analysis for public bicycle system, *International Journal of Computers Communications & Control*, 17(1), 4673, 2022.
https://doi.org/10.15837/ijccc.2022.1.4673