
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Communications in Transportation Research 2 (2022) 100068

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Full Length Article

# Understanding travel behavior adjustment under COVID-19

Wenbin Yao [a,b], Jinqiang Yu [c], Ying Yang [d], Nuo Chen [a,b], Sheng Jin [a,b,e,*], Youwei Hu [a], Congcong Bai [a,b]

[a] Institute of Intelligent Transportation Systems, Zhejiang University, Hangzhou, China
[b] Center for Balance Architecture, Zhejiang University, Hangzhou, China
[c] Alibaba Cloud Computing Co. Ltd., Hangzhou, China
[d] School of Behavioural and Health Sciences, Australian Catholic University, Sydney, Australia
[e] Pengcheng Laboratory, Shenzhen, China

ARTICLE INFO

*Keywords:*
COVID-19
Travel pattern
Travel behavior adjustment
Prefix-span algorithm
Random forest

ABSTRACT

The outbreak and spreading of the COVID-19 pandemic have had a significant impact on transportation system. By analyzing the impact of the pandemic on the transportation system, the impact of the pandemic on the social economy can be reflected to a certain extent, and the effect of anti-pandemic policy implementation can also be evaluated. In addition, the analysis results are expected to provide support for policy optimization. Currently, most of the relevant studies analyze the impact of the pandemic on the overall transportation system from the macro perspective, while few studies quantitatively analyze the impact of the pandemic on individual spatio-temporal travel behavior. Based on the license plate recognition (LPR) data, this paper analyzes the spatiotemporal travel patterns of travelers in each stage of the pandemic progress, quantifies the change of travelers' spatiotemporal behaviors, and analyzes the adjustment of travelers' behaviors under the influence of the pandemic. There are three different behavior adjustment strategies under the influence of the pandemic, and the behavior adjustment is related to the individual's past travel habits. The paper quantitatively assesses the impact of the COVID-19 pandemic on individual travel behavior. And the method proposed in this paper can be used to quantitatively assess the impact of any long-term emergency on individual micro travel behavior.

## 1. Introduction

The novel coronavirus pneumonia pandemic (COVID-19) has been spreading rapidly worldwide (Sun et al., 2020; World Health Organization, 2022). Aiming to fight the pandemic, many countries have introduced a series of policies and measures to slow down the spread of the COVID-19 virus. These policies and measures refer to various fields, including foreign trade, transportation, and social life. The outbreak and spreading of the COVID-19 virus have had a large impact on the entire mankind, and this impact has been reflected in various aspects, such as foreign trade, economy, and transportation. The implementation of policies, normal development of society and economy, company operation, and personal travel all require a convenient transportation system as a foundation. However, the transportation system has undergone tremendous changes due to the pandemic and anti-pandemic measures. For instance, in China, on January 23, 2020, the government released a provision informing the citizens of Wuhan that urban buses, subways, ferries, and long-distance passenger transportation would be suspended, and the airport and railway station in Wuhan would be temporarily closed. In addition, citizens of Wuhan were requested not to leave the city without special reasons. Later, a series of anti-pandemic measures affecting the travel of residents across China, such as the travel ban, were promulgated in sequence. In addition to the impact of these policies and measures on the transportation system, in fact, residents' concerns about the spread of the pandemic have also had a great impact on the transportation system. For instance, in some of the cities across China, due to the fear of the pandemic, the number of motor vehicle trips during peak hours has increased compared to that before the outbreak of the pandemic according to our previous study. Hence, the pandemic has had a profound impact on the transportation system, and changes introduced in the transportation system have further affected other aspects of society, including environmental system (Wang and Su, 2020), aviation (Iacus et al., 2020), agriculture (Ilesanmi et al., 2021), foreign trade (Wang et al., 2021a), and political relation (Yang, 2021). Meanwhile, these fields have reciprocally affected the transportation system since they are mutually connected, thus affecting each other. Therefore, it has still been very challenging to assess the impact of the pandemic on

* Corresponding author. Institute of Intelligent Transportation Systems, Zhejiang University, Hangzhou, China.
*E-mail address:* jinsheng@zju.edu.cn (S. Jin).

https://doi.org/10.1016/j.commtr.2022.100068
Received 29 March 2022; Received in revised form 18 May 2022; Accepted 18 May 2022
Available online 24 May 2022

**Nomenclature**

| | |
|---|---|
| LPR | license plate recognition |
| POI | point of interest |
| *FD* | the frequency density |
| *CR* | the category ratio |
| **a** | trip chain |
| $P_i$ | the spatiotemporal information on a traveler obtained at the *i*-th detected point |
| *N* | the travel days of the traveler |
| **A** | sequence database, $\mathbf{A} = \{\mathbf{a^1}, \mathbf{a^2}, \ldots, \mathbf{a^N}\}$ |
| **α** | frequent travel pattern |
| *minsup* | minimum support value |
| **F** | frequent travel pattern set, $\mathbf{F} = \{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^R\}$ |
| *LCS* | the longest common substring, which is defined as the longest common part of two travel patterns |
| $s_{ij}$ | the similarity of two frequent travel pattern sets |
| $rug_i$ | the regularity of travel behavior of a traveler *i* |
| $s^{ij}$ | the similarity of the frequent travel pattern sets between two stages |
| *k* | number of initial clusters |
| *s*(*i*) | the Silhouette coefficient of the point $p_i$ |
| *Q* | the hourly detection frequency (veh/h) |
| $N_{VT}$ | The number of vehicles in transit (veh) |
| $A_{TI}$ | the average travel intensity |
| $X_1 - X_{29}$ | independent variables |
| *Y* | dependent variable |

society in general. For instance, the impacts of the pandemic on the urban transportation system and environmental system have not been profoundly studied yet (Jiang et al., 2020). With the accumulation of massive data in various fields and the development of big data technology, various industries have used multi-source big data to evaluate the impact of COVID-19 on society from different perspectives. Among these studies, the research on environmental system has been the most abundant (Wang and Su, 2020; Sicard et al., 2020; Kerimray et al., 2020; Wang et al., 2020; Edwards et al., 2021; Calderon-Tellez and Herrera, 2021).

There are mutual effects and correlations between the transportation system and various other industries, so it is difficult to analyze and evaluate the transportation system comprehensively, especially the impact of the pandemic on micro-travel behavior. In addition, the amount of traffic data is massive, and generally, only governments and large companies own the data, while these data are difficult for individuals to obtain. This further limits the researchers to quantitatively analyze of the impact of the pandemic on the transportation system. Due to the limitations of the above reasons, there are few studies on the impact of the pandemic on the transportation system, especially the impact of the pandemic on micro travel behavior. However, it is very important and meaningful to analyze the impact of the pandemic on the transportation system and people's travel behavior. Firstly, transportation system can reflect the status of urban vitality and social and economic development, whereas the analysis of travel behavior adjustments under COVID-19 and the corresponding influencing factors can help to deepen the understanding of how the pandemic has affected the daily life of residents, thus helping to analyze the far-reaching influence of the pandemic on society. Secondly, analyzing the impact of the pandemic on the transportation system can evaluate the effect of policy implementation to a certain extent, because many anti-pandemic policies slow down the spread of the pandemic by limiting travel. Third, analyzing the impact of the pandemic on the transportation system and individual behavior can provide support for policy formulation and optimization.

As far as we know, currently, there is little research on the change of individual travel pattern affected by the pandemic with the help of traffic big data of full sample motor vehicles, and we hope to propose a new method to quantitatively evaluate the change of each individual's spatiotemporal travel pattern in each stage of pandemic development. Besides, the travel behavior adjustment pattern of all motor vehicles under the influence of the pandemic can also be obtained.

Aiming at the current limitation in research of the influence of the pandemic on transportation, based on the license plate recognition (LPR) data of Yiwu City, China, this paper analyzes frequent travel patterns at various stages of the pandemic and quantitatively describes the change of travel behaviors in each of the stages. On this basis, the impact of the pandemic on the trip chain of travelers is analyzed from both temporal and spatial perspectives. Moreover, the spatiotemporal travel behavior of travelers before the pandemic is quantitatively described, and the travel behavior adjustments of different categories of travel groups under COVID-19 are analyzed. This paper proposes a new method to quantitatively evaluate the change of individual travel pattern under the influence of the pandemic, and this method can be used to quantitatively analyze the adjustment of individual travel behavior under any kind of long-term service disruption. With the help of the proposed method, taking Yiwu as an example, we successfully found the individual behavior adjustment pattern at each stage of the pandemic and the corresponding influencing factors.

The remaining part of this manuscript is organized as follows. Section 2 reviews the change of transportation system and travel behavior under the COVID-19. Section 3 introduces the data used in this study. Section 4 presents frequent travel patterns at each stage and quantitatively analyzes the change of travel behavior at each stage of the pandemic and travel behavior adjustments of travelers under the COVID-19. Section 5 discusses travel behavior adjustment and the influencing factors of travel behavior adjustments of different groups. Section 6 concludes the paper and presents future prospects.

## 2. Literature review

### 2.1. Impact of COVID-19 on transportation system from the macroscopic perspective

Most of the recent studies have analyzed the impact of the pandemic on the entire urban transportation system from the macroscopic perspective using open statistics. Xu et al. (2020) analyzed the impact of the pandemic on the socio-economic system from two aspects, the impact of the pandemic on population migration and the impact of the pandemic on urban traffic. They concluded that the relevant government departments should focus more on densely populated and economically developed provinces and cities when applying the anti-pandemic and work resumption policies. Wang et al. (2021b) conducted an origin and destination (OD) cluster analysis based on Baidu migration data, mainly analyzing the relationship between the migration of the Guangdong–Hong Kong–Macao Greater Bay Area and pandemic progress and concluded that the degree of pandemic progress in the Guangdong–Hong Kong–Macao Greater Bay Area was positively correlated with the degree of migration from Wuhan to the Greater Bay Area. In addition, they also analyzed the relationship between the migration level of cities inside and outside the Greater Bay Area and the spread rate of the pandemic. Some studies have modeled and simulated population migration between cities in China to characterize the relationship between the pandemic and population migration, thus providing necessary support to policy-making (Chen et al., 2020a). This kind of research mainly focuses on the urban transportation system using different data to extract various indicators to

observe the operating features of the urban transportation system, such as traffic flow, vehicle number in transit, travel intensity, and travel speed, at designated observation points. These studies have reported a similar conclusion, which is as follows. Affected by the pandemic and the corresponding measures, the operation of the urban transportation system has been hindered to a certain extent. Since the pandemic has been put under control, the urban transportation system gradually has recovered (Patra et al., 2021; Marinello et al., 2021; Harantová et al., 2020), but there is a significant positive correlation between the operating intensity of the urban transportation system and the spread of COVID-19 (Lee et al., 2020). The impact of the pandemic on travel distribution differs among cities. Aloi et al. (2020) reported that in Spain, due to the pandemic, travel decreased sharply; namely, private cars were relatively less affected, but public transportation was greatly affected. Also, it was pointed out that the decrease in the number of trips in the afternoon was more significant than in the morning and noon. While Muley et al. (2021) concluded that in the State of Qatar, although the travel volume reduced by about 30% after the promulgation of travel restrictions, the distribution of travel throughout the day did not change significantly.

Public transportation is a fundamental part of the transportation system, and it has often been severely affected by the pandemic. Using multi-source big data such as vehicle GPS data, bus station data, and mobile phone signaling data, Orro et al. (2020) analyzed changes in the traffic volume, station usage, main destinations' traffic flow, bus frequency, operation time, and reliability of Coruna's bus network in the post-COVID-19 era. In the post-COVID-19 era, the conventional transportation system and shared bicycle systems have recovered faster than the public transportation system. At present, there have been a number of studies that constructed models of a relationship between the transportation system and the spread of the pandemic, which simulate the process of issuing the traffic control measures and quantitatively evaluate the effect of the measures on the further spreading of the pandemic (Klise et al., 2021; O'Sullivan et al., 2020; Munshi et al., 2020; Anzai et al., 2020; Chang et al., 2021). For instance, Chang et al. (2021) established a resident travel-SEIR model using the SafeGraph data, US Census data, and the information on the number of confirmed cases and deaths published in the New York Times. This model can simulate the progress of the pandemic under different pandemic-prevention and pandemic-control policies and can explore the differences in the pandemic incidence in people belonging to different social strata and races, as well as causes of these differences. In addition, certain studies have analyzed the impact of the pandemic on flight (Hou et al., 2021) and traffic accidents (Vandoros, 2021; Zhang et al., 2021).

The above-mentioned studies mostly analyze the overall or partial transportation system status of countries, considering the inter- and intra-city transportation under the influence of the pandemic from a macro perspective. Generally, the impact of the pandemic on the transportation system can be quantitatively analyzed by examining changes in certain transportation indicators in different periods of the pandemic progress, such as traffic flow, a total number of travelers, travel intensity, and network speed. The relationship between the pandemic spread and the transportation system can be explored by a simulation method. However, there have been fewer studies on the impact of the pandemic on individual travel behavior and travel pattern.

### 2.2. Impact of COVID-19 on travel behavior

A number of studies have aimed to analyze and model the change of individual travel behavior and travel pattern under the influence of the pandemic. Jiao et al. (2021) analyzed the impact of the pandemic on the travel patterns of residents in Houston, Texas using the autoregressive distributed lag model and have reported that the week travel patterns before the pandemic outbreak had a significant impact on that of the next week. In addition, several factors affecting the number of walking trips have also been analyzed. Zheng et al. (2020) used the GPS trajectory data of Shenzhen taxis to analyze the impact of the pandemic on the taxi industry in Shenzhen. The considered time period was divided into four stages according to the promulgation of policies and the development of the pandemic. The changes in the taxi industry were analyzed during the pandemic progress from the perspectives of supply and demand. It has been concluded that the taxi industry in Shenzhen has decayed due to the pandemic. In response to the pandemic, taxi drivers have adjusted their operational behavior. Aiming to stimulate the taxi industry, the government had provided certain subsidies to drivers, but the subsidy method was slightly overcorrected, causing the phenomenon of oversupply in the market. Li et al. (2021a, 2021b) analyzed the changes of micro-mobility usage before and during the pandemic. The patterns of docked bike, docked e-bike, and dockeless e-bike were analyzed and compared with each other. The study revealed the similarities and differences between the three micro-mobility patterns and the changes of the three micro-mobility patterns before and during the pandemic.

These studies have addressed the problem of a lack of analysis of the change of individual travel behavior and travel patterns under the influence of the pandemic to a certain extent. However, in general, there have been insufficient studies of this topic, and the existing studies have described the travel behavior and travel patterns mainly using descriptive statistics, regression analysis or other methods at an aggregated level, but does not describe the changes of each individual's spatiotemporal travel behavior. In addition, the current research mainly focuses on a specific travel mode or research based on questionnaire data, but does not analyze the full sample of motor vehicle travelers in the road network.

## 3. Data description

### 3.1. License plate recognition (LPR) data

The data used in this study is structured LPR data obtained by LPR detectors installed on road intersections in Yiwu, Zhejiang Province, China. LPR data are widely used in traffic flow analysis (Wang et al., 2016; Ma et al., 2017a; 2017b; Luo et al., 2019a, 2019b; Shen et al., 2020; Yao et al., 2021; Ma et al., 2018). The data cover the period from December 2019 till June 2020, spanning the entire period from the outbreak of the COVID-19 to the post-COVID-19 era. A total of 5,510 LPR detectors have been deployed in the city, basically covering the entire Yiwu, and the specific distribution of detectors is shown in Fig. 1a. The average number of detected records per day during the research period is 8,596,466. The LPR data include the information on vehicle license plate numbers, the longitude and latitude of detectors, and the time when vehicles passed by the detectors. It fully characterizes the spatiotemporal information of vehicles traveling through Yiwu City and can be used to analyze the travel characteristics of vehicles in Yiwu City.

### 3.2. Point of interest data

The point of interest (POI) data denote the data type that has been paid great attention in the development of geographic information systems in recent years (Yuan et al., 2013). Many Internet map companies have established POI acquisition interfaces, such as Google, OpenStreetMap, and AMAP owned by Alibaba. Many service points on an electronic map, including restaurants, shopping spots, service centers, and other types of service facilities, have been collected by the map company and have been available to travelers in the form of POI data. In recent years, POI data have been widely used, for instance, to reflect the type of urban land usage and traveling purpose (Gao et al., 2021b; Liu et al., 2018). This study uses the Yiwu POI data provided by AMAP to analyze the influencing factors of travel behavior adjustments of different groups of travelers caused by the pandemic. The collected POI data include the name, latitude and longitude, category, and other relevant information of each POI point in Yiwu. A total of 54,320 POIs were collected, including catering services, shopping spots, life services, sports and leisure services, medical care services, scenic spots, commercial and

residential areas, government agencies and social organizations, education and cultural services, companies and enterprises.

This study uses POI data to reflect the type of land usage around a certain LPR detector. The method proposed by Yao et al. (2022) is used to analyze the land use types, and the method is improved combined with the method proposed by Gao et al. (2021b). The method mainly includes the following three steps:

1) The category of POI is reclassified to better reflect the land use type.
2) Calculate the frequency density (*FD*) and category ratio (*CR*) of the reclassified POI data, *FD* and *CR* are defined by Eqs. (1) and (2), respectively.
3) Use four ***CRs*** as a feature vector ($\boldsymbol{CR}_1$, $\boldsymbol{CR}_2$, $\boldsymbol{CR}_3$, $\boldsymbol{CR}_4$) and employ the *k*-means clustering algorithm to determine the land usage type within the buffer area of each LPR detector in Yiwu.

$$FD_i = \frac{n_i}{N_i} \tag{1}$$

$$\boldsymbol{CR}_i = \frac{FD_i}{\sum_i FD_i} \tag{2}$$

In Eqs. (1) and (2), $FD_i$ denotes the frequency density of POIs of class $i$ in the buffer, $n_i$ is the number of POIs of the class $i$ in the buffer, $N_i$ is the number of POIs of the class $i$ in the whole Yiwu city, and $CR_i$ is the $CR$ of POIs of the class $i$.

The distribution of POI of each category after reclassification is shown in Fig. 1b. The land usage types within the buffer area of the LPR detectors in Yiwu City are divided into three types. The first type mainly includes POIs of the commercial area and companies. The second type has a balanced proportion of various types of POIs, which are mostly distributed in the central district of Yiwu City. The third type mainly includes POIs of the public service type, which are mostly distributed in the suburbs of Yiwu City. The corresponding distributions are shown in Fig. 1c. On this basis, the Shannon entropy index within the buffer area around each LPR detector is calculated with the help of the method proposed by Gao et al. (2021a). The Shannon entropy index can be used to measure the land use mixture.

### 3.3. COVID-19 stages

Since the COVID-19 outbreak, Zhejiang Government has actively responded to the call of the country and put forward a variety of response policies. For instance, on January 23, the Wuhan Government announced the closure of the city, and Zhejiang Government also announced that it had entered a first-level response state (Chen et al., 2020b; Mei, 2020). On February 15th, the operation of taxis and online car-hailing has recovered in Yiwu city, Zhejiang Province. On February 21, urban traffic operations resumed, and certain important business centers in the city opened. In addition, the resumption of work and production also started. On March 2, the emergency response status of Zhejiang Province dropped to the second level. On April 13, Zhejiang Government announced that middle and high school students in various regions of the province would return to school (Mei, 2020). Combined with the evolution of pandemic response policies, the pandemic progress can be divided into three stages (Zheng et al., 2020), and the change of travel behavior in these three stages is analyzed in this study. The first stage is the control stage before the COVID-19 outbreak (hereinafter referred to as the first stage), and this stage is from December 16, 2019, to December 27, 2019. The second stage is the initial stage of resumption of work and production (hereinafter referred to as the second stage), and it includes the period from February 10, 2020, to February 21, 2020. The third stage is the post-COVID-19 era (hereinafter referred to as the third stage), and it lasts from June 8 to June 19, 2020. This study analyzes and models the travel behavior based on the workdays' LPR data of the three stages.

## 4. Travel behavior adjustment under COVID-19

Since December 2019, the COVID-19 pandemic has been experiencing rapid growth in China first and then has been gradually controlled. In this process, the control measures announced by the Chinese government have played an important role. Affected by the pandemic and the corresponding control measures, travel behavior of motor vehicles has changed to a certain extent. This section will present the changes in motor vehicle's travel behavior at different stages of the COVID-19 pandemic.

### 4.1. Frequent travel pattern mining

Travelers typically travel many times a day, and by linking multiple trips, a trip chain of travelers can be formed. By observing the trip chains of travelers over a certain period of time, frequent travel patterns of travelers can be determined. A frequent travel pattern represents a generalization of a traveler's regular travels, which can reflect the traveler's travel habits in a period of time.

This study uses the LPR data as a data source; however, the LPR data cannot show where exactly a traveler is at each moment but only

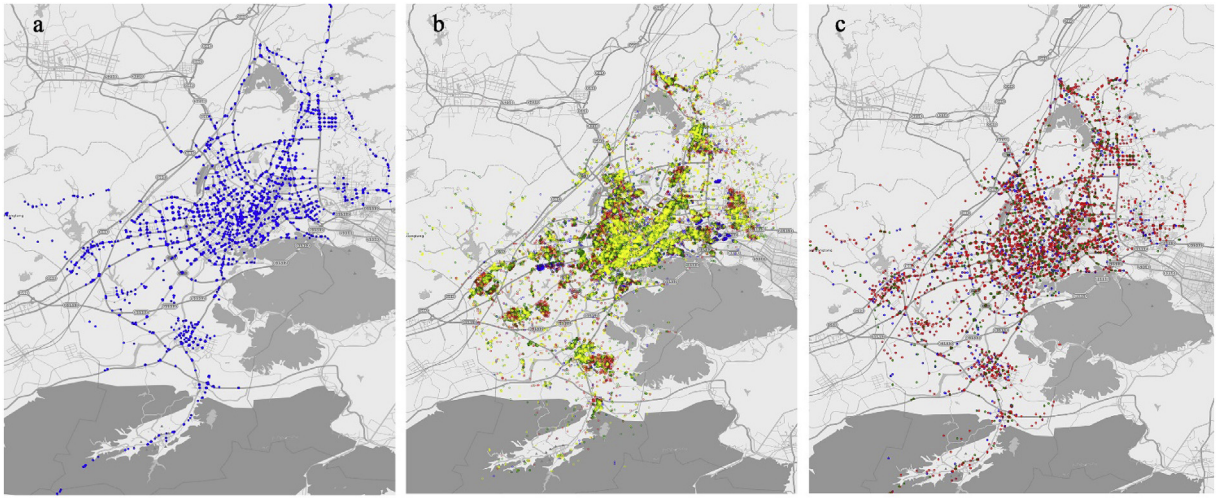



**Fig. 1.** Distributions of LPR detectors and POIs. (a) Distribution of LPR detectors in Yiwu City. (b) Distributions of various categories of POI data; blue denotes residential area and life services, red stands for commercial area and companies, green represents entertainment area, and yellow indicates public services. (c) Distributions of different land usage types; blue, red, and green indicate the first, second, and third types, respectively.

provides the spatiotemporal point information captured by the LPR detectors in part of the traveler's trip chain. Therefore, the trip chain is defined as **a**, $\mathbf{a} = \{P_1, P_2, P_3, \ldots, P_K\}$, where $P_i$ is the information on a traveler obtained at the $i$-th detected point. The specific form of $P_i$ is $[t_i, x_i, y_i]$, where $t_i$, $x_i$, and $y_i$ denote the time, longitude, and latitude of a traveler's $i$-th detected point, respectively; $P_i$ is defined as an **item**, **a** is defined as a **sequence**, and several items without sequence constitute an **item set** (Saraf et al., 2015). For commuters, $P_1$ is usually near home, as well as $P_K$, and they are likely to be located at the same intersection.

This study aims to mine frequent travel patterns from the travelers' daily trip chains. Frequent travel patterns constitute substrings of a trip chain. To describe frequent travel patterns more clearly and lay a foundation for subsequent frequent travel pattern mining, the following terms are defined (Shou and Di, 2018).

**Definition 4.1.** **(Subsequence and super sequence).** If for each item in a sequence **a'**, there is the same item in sequence **a**, and these items are in the same order in two sequences, then **a'** is called the **subsequence** of **a**, and **a** is called the **super sequence** of **a'**. Since the specific form of an item $P_i$ is $[t_i, x_i, y_i]$, if two items have the same time, longitude, and latitude, and then they are regarded as the same item. It is easy to prove that even the same traveler can hardly have the same item on different days, which can lead to poor results in the subsequent frequent pattern mining. Therefore, in this study, Eq. (3) is used to judge whether items $P_i$ and $P_j$ are the same.

$$\begin{cases} P_i = P_j, \left|\text{hour}(t_i) - \text{hour}(t_j)\right| \le t_t \& \text{dis}(P_i, P_j) \le t_d \\ P_i \ne P_j, \left|\text{hour}(t_i) - \text{hour}(t_j)\right| > t_t | \text{dis}(P_i, P_j)) t_d \end{cases} \tag{3}$$

In Eq. (3), hour($t$) refers to the hour of the corresponding time, excluding minutes and seconds; for instance, *hour*(09:25:36) = 09; dis($P_i$,$P_j$) refers to the Euclidian distance between $P_i$ and $P_j$ calculated by longitude and latitude; $t_t$ and $t_d$ are artificially set threshold parameters, and in this study, $t_t = 1$ h and $t_d = 1$ km according to relevant research (Chen et al., 2017; Mihai et al., 2015; Wilhelmsen et al., 2017). These two threshold parameters can be reset according to the actual situation.

**Definition 4.2.** **(Sequence database and support).** For each traveler, his/her daily trip chain is excavated. The daily sequence **a** constitutes the **sequence database A**, and $\mathbf{A} = \{\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^N\}$, where $N$ represents the travel days of the traveler. **Support** is used to measure the frequency of sequence **a**, which is defined as a ratio of the number of days of the occurrence of sequence **a** that is a subsequence in the traveler's sequence database to the total number of sequences in the sequence database.

**Definition 4.3.** **(Frequent travel pattern). Frequent travel pattern α** is defined as a sequence with the support greater than or equal to a specific minimum support value (*minsup*) in the sequence database. The length $l$ of a frequent travel pattern **α** is defined as the number of items. A traveler can have several frequent travel patterns, which are denoted by $\mathbf{F} = \{\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^R\}$.

In the following, specific examples of the above-defined terms are given, and frequent travel patterns of a traveler are extracted.

For a traveler who travels every day on weekdays, it is assumed that he/she has a sequence database $\mathbf{A} = \{\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^5\}$, where $\mathbf{a}^1 = \{P_1, P_2, P_3\}$, $\mathbf{a}^2 = \{P_1, P_2\}$, $\mathbf{a}^3 = \{P_1, P_4\}$, $\mathbf{a}^4 = \{P_1, P_2\}$, and $\mathbf{a}^5 = \{P_1, P_2\}$. For convenience, points detected on different weekdays are expressed by $P_i$; it should be noted that the same $P_i$ of travelers does not indicate that the travelers are detected at the same time and location. As long as Eq. (3) is satisfied, even though they are not detected at the same time and location, it is marked as the same detected point. For *minsup* = 0.6, the travel pattern and the corresponding support are shown in Fig. 2. Starting from the root and walking down through any path, arranging the items in order forms a travel pattern. By connecting items with a support value greater than or equal to *minsup*, a frequent travel pattern (the path where red items are located in Fig. 2) is obtained. In this example, the frequent travel pattern set **F** is $\{\{P_1\}, \{P_2\}, \{P_1, P_2\}\}$. With the help of the frequent travel pattern of a traveler, it can be inferred that the traveler is likely to be a commuter with a short commute distance. Further, $P_1$ denotes a detected point on the way from home to work, and $P_2$ represents a detected point on the way back home from work. For $\mathbf{a}^1$, on Monday, a traveler may go to shopping malls or other places after work and then returns home. With the help of frequent travel patterns of travelers, reasonable inferences about their travel behavior can be constituted.

First, the detected point information of a traveler is arranged in chronological order to obtain the traveler trip chain, which represents a sequence. After that, the daily trip chain in the first stage is mined to form a sequence database, which is then used to mine frequent travel patterns by the Prefix-Span algorithm (Han et al., 2001). The enumeration method can be employed to obtain frequent travel patterns from the sequence database, but this simple and rough method often consumes too much time and thus is not applicable to large data. The Prefix-Span algorithm has been designed to reduce the complexity of frequent pattern mining. See the **Appendix** for details of how to use Prefix-Span algorithm to mine frequent travel pattern.

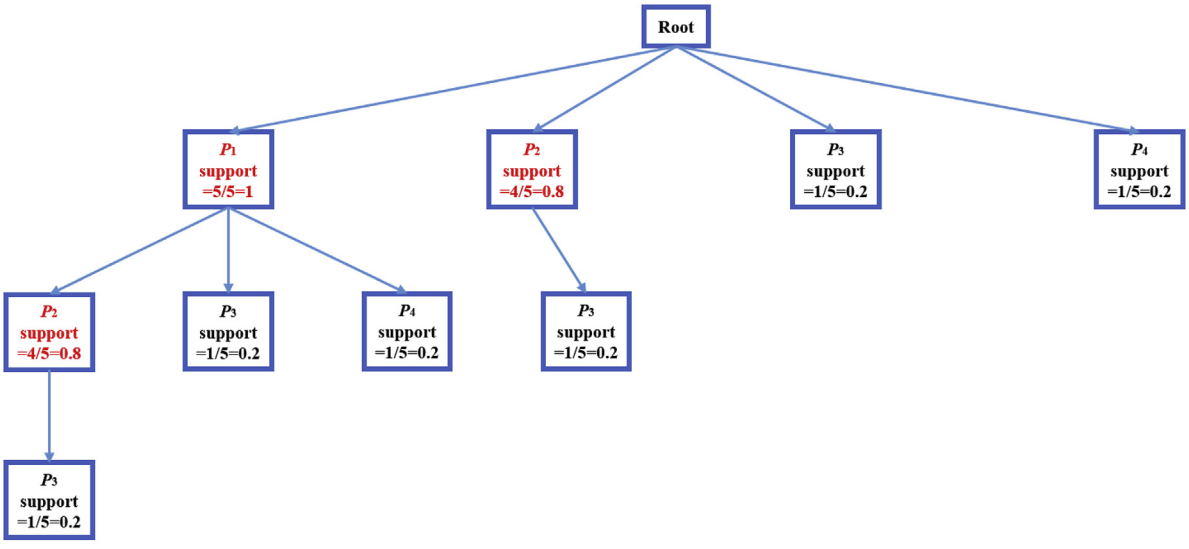



**Fig. 2.** Travel pattern tree.

Although the Prefix-Span algorithm can greatly reduce the complexity of the algorithm that mines frequent travel patterns compared with the enumeration method, in some cases, the results cannot be obtained in an acceptable time. In this study, there are 10-day LPR data in each stage, and *minsup* is set to 0.6 (Shou and Di, 2018; Saraf et al., 2015). Considering a typical commuter, it indicates that the commuter drives from home to work at the morning peak and returns home at the evening peak; the LPR detectors are densely arranged along the way. The commuter travels every day on weekdays, and the travel trajectory on each day is the same; that is, the spatiotemporal information of the corresponding detected points on each day is the same. The commuter is detected eight times from home to the company in the morning peak and eight times from the company to home in the evening peak; $\mathbf{A} = \{\mathbf{a^1}, \mathbf{a^2}, \ldots, \mathbf{a^{10}}\}$, $\mathbf{a^1} = \mathbf{a^2} = \ldots = \mathbf{a^{10}}$, where $\mathbf{a^1} = \{P_1, P_2, \ldots, P_{16}\}$. In this case, the total number of frequent travel patterns of the commuter can be calculated as $C_{16}^1 + C_{16}^2 + \cdots + C_{16}^{16} = 65,535$. It is worth noting that in this case, it has been assumed that the spatiotemporal detected points constituting the trip chain of every day are the same; that is, a commuter pass through the same location at the same time every day. However, a commuter is more likely to pass through a similar place at a similar time, which will make the detected points be judged as the same point by Eq. (3). In this case, the number of frequent travel patterns will be a certain number times that of 65,535. For travelers with a longer trip chain than that mentioned above, the number of frequent travel patterns increases explosively with the increase in the detected point number.

To solve the problem that the Prefix-Span algorithm cannot obtain results within an acceptable time for travelers with a too-long trip chain, sequence compression is used for such travelers. The following criterion is adopted to judge whether a trip chain is too long; if the trip chain meets the following criterion, the sequence compression method will be used.

1) The trip chain length of a traveler is longer than or equal to 10 in more than or equal to (*minsup* × 10) days.
2) If the first criterion is not met, then it is still needed to judge whether the average daily trip chain length is longer than or equal to 10.

The sequence compression algorithm is given in **Algorithm 1**.

**Algorithm 1**. Sequence compression.

**Input:** a sequence database $\mathbf{A}=\{\mathbf{a^1},\mathbf{a^2},\ldots,\mathbf{a^5}\}$

**Output:** a sequence database $\mathbf{A'}=\{\mathbf{a^{1'}},\mathbf{a^{2'}},\ldots,\mathbf{a^{5'}}\}$

**Algorithm steps:**

1. For each sequence $\mathbf{a^i}$ in $\mathbf{A}$, compress items in $\mathbf{a^i}$; if several items are detected in the same hour, then keep only the one with the largest support and delete others.

2. After Step 1, judge whether it is still needed to perform sequence compression; if so, for $\mathbf{a^i}$ whose length is longer than or equal to 10, keep the first five items and the last four items; otherwise, return result $\mathbf{A'}$.

Based on the sequence compression algorithm and the Prefix-Span algorithm, frequent travel patterns of the three pandemic stages are obtained. The flowchart of the frequent travel pattern mining is shown in Fig. 3.

### 4.2. Clustering feature extraction

In Section 4.1, frequent travel patterns at various stages have been extracted. Frequent travel patterns reflect the regular travel behavior of travelers at a particular stage. Affected by the pandemic, the travel behaviors of travelers can change to a certain extent. Aiming to measure the change in travel behavior in different stages of the pandemic, it is necessary to define the similarity of frequent travel patterns. This study adopts the similarity measurement method of frequent travel patterns proposed by Shou and Di (2018), and a few concepts need to be introduced before defining the similarity of frequent travel patterns.

**Definition 4.4**. (*LCS*). The longest common substring (*LCS*) is defined as the longest common part of two travel patterns. For instance, for travel patterns $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_2}$, where $\boldsymbol{\alpha_1} = \{P_1, P_2, P_3, P_4, P_5\}$ and $\boldsymbol{\alpha_2} = \{P_1, P_2, P_6, P_7, P_8, P_5\}$, the longest common substring is $LCS(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}) = \{P_1, P_2, P_5\}$. The ratio of $LCS(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2})$ to travel pattern $\boldsymbol{\alpha_1}$ is defined as a ratio of the length of $LCS(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2})$ to the length of travel pattern $\boldsymbol{\alpha_1}$, which is denoted by $Ratio(LCS(\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}), \boldsymbol{\alpha_1})$.

Based on the above definition, the similarity $s_{ij}$ (Shou and Di, 2018) of two frequent travel pattern sets, $\mathbf{F_i} = \{\boldsymbol{\alpha}_i^r, r = 1, 2, \ldots, R\}$ and $\mathbf{F_j} = \{\boldsymbol{\alpha}_j^q, q = 1, 2, \ldots, Q\}$, can be calculated by Eq. (4). In this study, frequent travel patterns with a length of one are deleted from the frequent travel pattern set because they contain little information, and the frequent travel patterns with a length longer than one are retained. As for frequent travel patterns with a length of two, they might represent the travel behavior of a typical commuter, with a frequent item in the morning peak and a frequent item evening peak.

$$s_{ij} = \frac{\frac{\sum_{r=1}^{R} w\left(\alpha_i^r, \alpha_j^{qr}\right) \times s_p\left(\alpha_i^r, \alpha_j^{qr}\right)}{\sum_{r=1}^{R} w\left(\alpha_i^r, \alpha_j^{qr}\right)} + \frac{\sum_{q=1}^{Q} w\left(\alpha_j^q, \alpha_i^{rq}\right) \times s_p\left(\alpha_j^q, \alpha_i^{rq}\right)}{\sum_{q=1}^{Q} w\left(\alpha_j^q, \alpha_i^{rq}\right)}}{2} \tag{4}$$

$$s_p\left(\alpha_i, \alpha_j\right) = \frac{len(\alpha_i) + Ratio\left(LCS\left(\alpha_i, \alpha_j\right), \alpha_i\right) + len\left(\alpha_j\right) + Ratio\left(LCS\left(\alpha_i, \alpha_j\right), \alpha_j\right)}{len(\alpha_i) + len\left(\alpha_j\right)} \tag{5}$$

In Eq. (4), $q_r = \underset{1 \le q \le Q}{\arg\max}(s_p(\alpha_i^r, \alpha_j^q))$, $r_q = \underset{1 \le r \le R}{\arg\max}(s_p(\alpha_i^r, \alpha_j^q))$, and $w(\alpha_i^r, \alpha_j^q) = \sqrt{\text{support}(\alpha_i^r) \times \text{support}(\alpha_j^q)}$.

The described similarity measurement procedure of frequent travel pattern sets can be used to measure the similarity of frequent travel pattern sets between travelers in different stages of the pandemic, i.e., to explore the change in the regular travel behaviors of travelers in different stages of the pandemic.

In addition to measuring the change in travel behavior in each stage, it is also desirable to analyze its regularity quantitatively in each stage. The regularity of travel behavior of a traveler $i$ can be calculated by Eq. (6) (Shou and Di, 2018).

$$rug_i = \frac{\sum_{r=1}^{N} \sum_{q=1}^{N} w\left(\alpha_i^r, \alpha_i^q\right) \times s_p\left(\alpha_i^r, \alpha_i^q\right)}{\sum_{r=1}^{N} \sum_{q=1}^{N} w\left(\alpha_i^r, \alpha_i^q\right)} \tag{6}$$

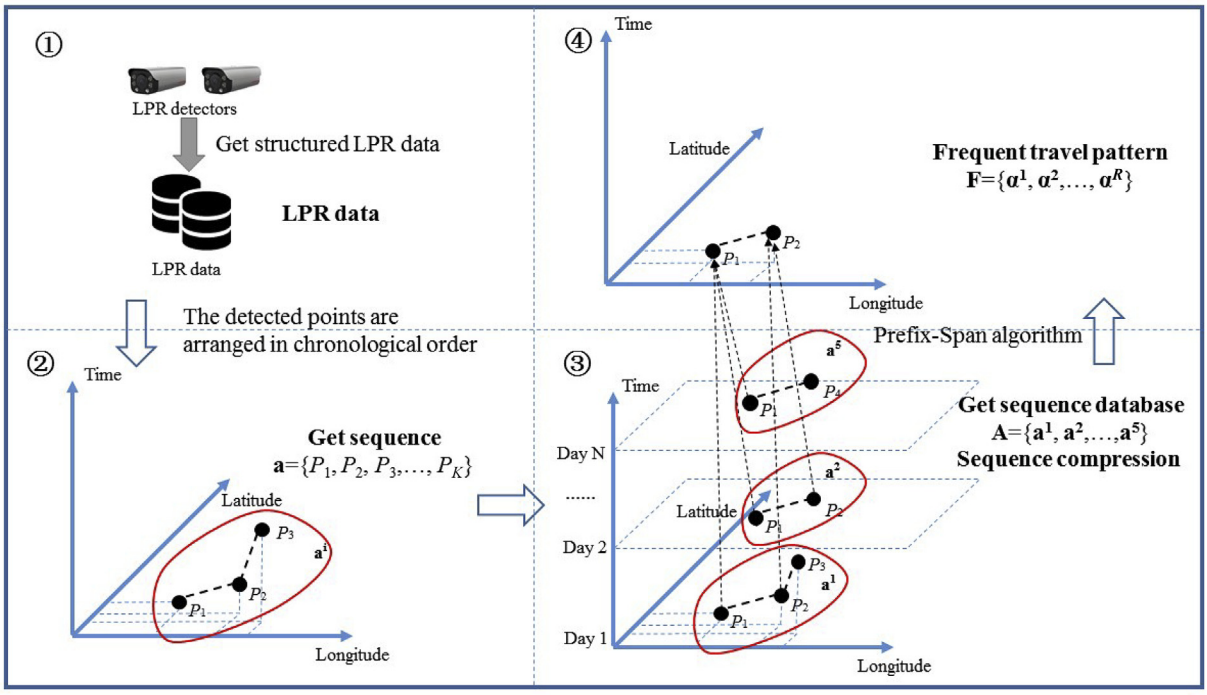



**Fig. 3.** Frequent travel pattern mining flowchart.

where $rug_i$ denotes the regularity of the travel behavior of a traveler $i$, and $w(\alpha_i^r, \alpha_i^q) = 1, \forall r, q \in \{1, 2, ..., N\}$.

After obtaining the similarity of two frequent travel pattern sets and the travel behavior regularity of travelers, five clustering features, including the travel behavior regularity values $rug^1$, $rug^2$, and $rug^3$ in the first, second, and third stages, respectively, similarity $s^{12}$ of frequent travel pattern sets between the first and second stages, and the similarity $s^{13}$ of frequent travel pattern sets between the first and third stage, are calculated.

### 4.3. Cluster analysis

To analyze possible behavior adjustments of travelers due to the pandemic, cluster analysis is performed. Taking the first stage as a control group, vehicles with frequent travel patterns in the first stage are extracted, and a total of 245,743 vehicles are considered. Due to a large number of data samples, to obtain results in an acceptable time, the *k*-means algorithm (MacQueen, 1967) is used to cluster travelers into several different categories. The *k*-means clustering is a method of vector quantization, which divides *n* observations into *k* clusters such that each observation belongs to the cluster with the nearest mean. In practice, the *k*-means algorithm is very fast, and currently, it is one of the fastest clustering algorithms (Pedregosa et al., 2011). However, to successfully implement the *k*-means algorithm, it is necessary to specify the number of clusters *k* in advance. In this study, the number of clusters varies from two to 18, and the Silhouette coefficient (Rousseeuw, 1987) is used to evaluate the effect of the clustering algorithm. It is necessary to calculate the mean intra-cluster distance ($a(i)$) and the mean nearest-cluster distance ($b(i)$) of a point $p_i$ first, and then the Silhouette coefficient of the point $p_i$ can be calculated by Eq. (7).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{7}$$

Finally, the Silhouette coefficients of all samples are averaged to obtain the Silhouette coefficient of the whole dataset using the clustering algorithm. The range of the Silhouette coefficient is [–1, 1]; the closer the value of the Silhouette coefficient is to one, the better the clustering effect is; if the value of the Silhouette is less than zero, it indicates that the clustering effect is not good, and many points are classified incorrectly. The clustering performance change with the number of clusters is shown in Fig. 4.

As shown in Fig. 4, when the number of clusters is three, the clustering performance is the best, and the Silhouette coefficient reaches 0.573. The average values of the five features of each cluster are given in Table 1.

Therefore, travelers are clustered into three clusters, accounting for 48.70%, 41.97%, and 9.32% of the total number of travelers, respectively. The similarity of frequent travel patterns of vehicles between the first and second stages and between the first and third stages of Cluster 0 is low. This result indicates that travelers from Cluster 0 did not recover their frequent travel patterns after the pandemic, even in the post-COVID-19 era. In Cluster 1, the similarity of frequent travel patterns between the first and second stages is also low, but the similarity of frequent travel patterns between the first and third stages is high; thus, travelers have recovered the pre-pandemic frequent travel patterns in the post-COVID-19 era. The similarity of frequent travel patterns between the first and second stages and between the first and third stages of Cluster 2 is high, indicating that travelers from this cluster have recovered their pre-pandemic frequent travel patterns at the initial stage of resumption of work and production.

## 5. Results and discussions

### 5.1. Travel behavior adjustment

The behavior adjustments made by travelers under the COVID-19 pandemic have been analyzed using the clustering method, and three different clusters have been obtained. This section conducts a more in-depth behavior analysis of travel behaviors of travelers in each cluster, exploring the behavior adjustments of travelers of each cluster at each pandemic stage.

First, changes in the average length of the frequent travel patterns in each cluster at each stage are obtained. Since each traveler has a frequent travel pattern set, the length of the longest frequent travel pattern in the frequent travel pattern set is used to calculate the average length. The

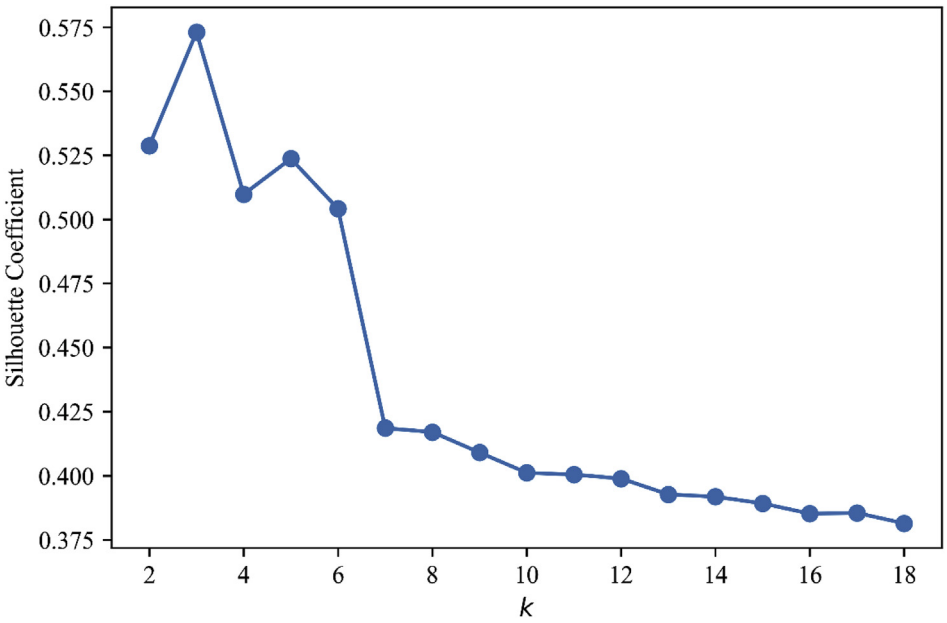



**Fig. 4.** Clustering performance change with the number of clusters.

**Table 1**
Average values of the five features of all clusters.

| | $rug^1$ | $rug^2$ | $rug^3$ | $s^{12}$ | $s^{13}$ |
|---|---|---|---|---|---|
| Cluster 0 | 0.624 | 0.505 | 0.413 | 0.002 | 0.001 |
| Cluster 1 | 0.683 | 0.420 | 0.642 | 0.001 | 0.536 |
| Cluster 2 | 0.689 | 0.756 | 0.637 | 0.471 | 0.371 |

results are presented in Fig. 5.

Next, the distributions of the first and last detected times of frequent travel patterns of each cluster are obtained. The extraction algorithm of the first and last detected times of the frequent travel patterns of each cluster is as follows:

1) Find the longest frequent travel pattern subset from a frequent travel pattern set.
2) From the longest frequent travel pattern subset, find the frequent travel pattern with the earliest time to be detected for the first time.
3) Extract the first and last detected times from the frequent travel pattern obtained in Step 2.
4) Obtain the distribution of the first and last detected times of travelers.

After extracting the first and last detected times by the above-given method, draw the cumulative probability distribution curves of the first and last detected times of each stage for each cluster. The obtained distribution curves of the three clusters are shown in Figs. 6a–6c.

According to Figs. 5 and 6a–6c, the average lengths of frequent travel patterns of vehicles in the second and third stages of Cluster 0 are almost zero, indicating that most travelers in these clusters have no frequent travel patterns in the second and third stages. From the distribution of the first and last detected times of Cluster 0, it can be concluded that the first detected time in the second and third stages of this cluster is delayed compared to that in the first stage, and the last detected time in the second and third stages is earlier than that in the first stage. Also, the distributions of the second and third stages are similar, and the time to go out which means the duration between the first travel and the last travel in a day is shorter in the second and third stages than in the first stage.

The average length of frequent travel patterns of vehicles in Cluster 1 is almost zero in the second stage, and it increases to 2.898 in the third

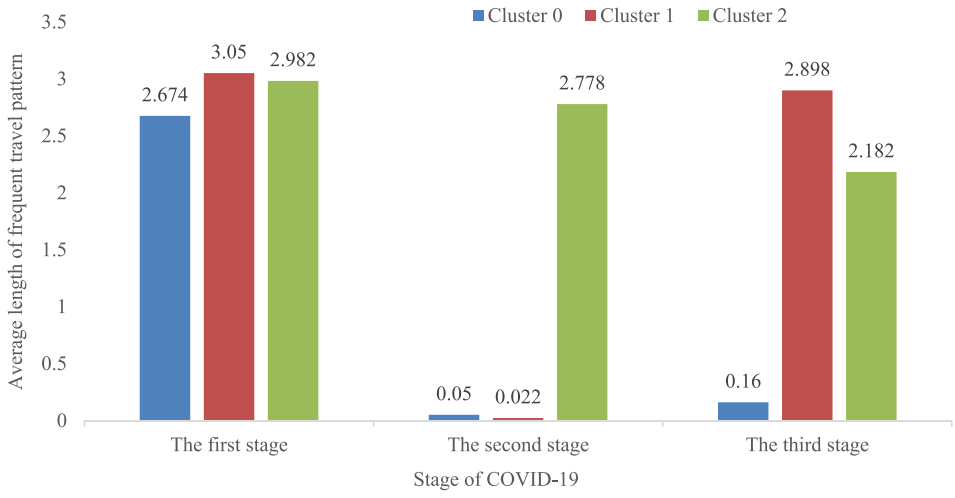



**Fig. 5.** Average lengths of the frequent travel patterns of each cluster at each stage.

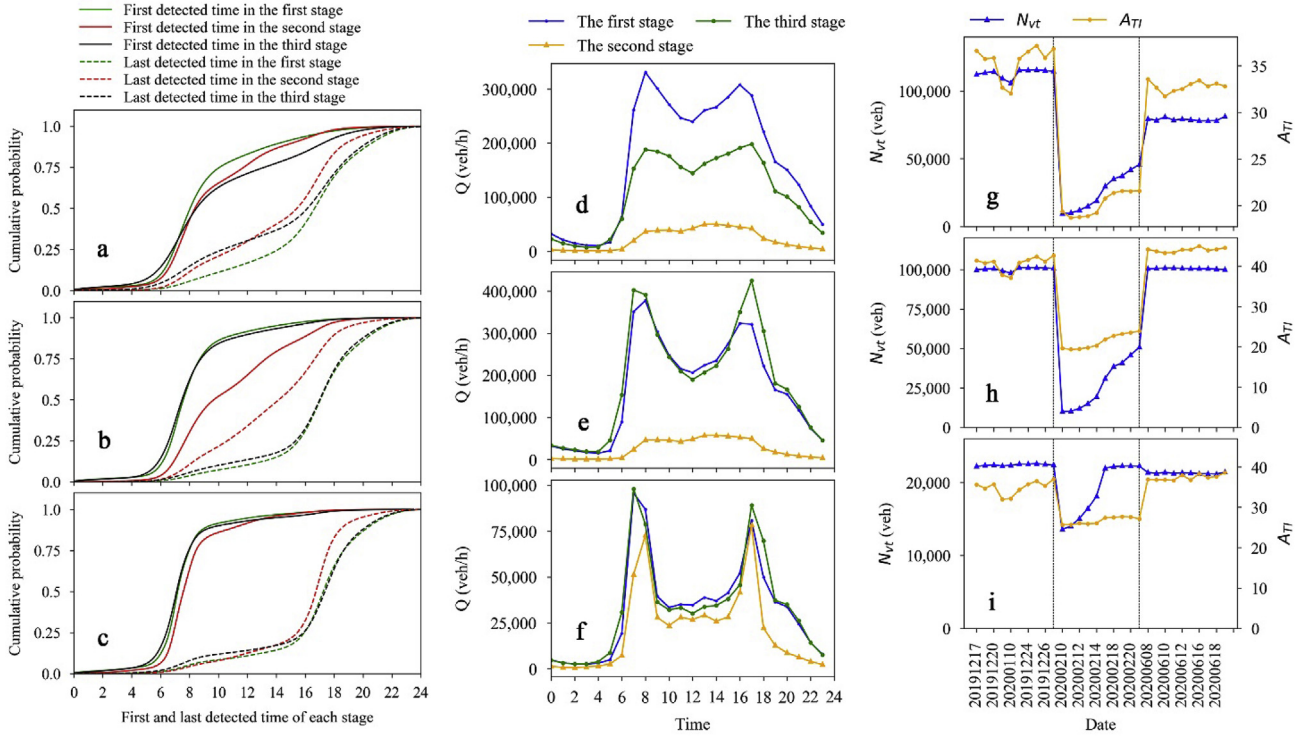



**Fig. 6.** Travel behavior adjustment of each cluster at each stage. (a–c) Cumulative distribution curves of the first and last detected times of the three clusters for each stage. (a), (b), (c) are the cumulative distribution for cluster 0, cluster 1, cluster 2, respectively. (d–f) The hourly detection frequencies of the three clusters in the three pandemic stages. (d), (e), (f) are the hourly detection frequencies for cluster 0, cluster 1, cluster 2, respectively. (g–i) The number of vehicles in transit and the average travel intensity of the three clusters on each day. (g), (h), (i) are the number of vehicles in transit and the average travel intensity for cluster 0, cluster 1, cluster 2, respectively.

stage, which is similar to the value in the first stage. For vehicles in this cluster, the first- and last-detected time distributions in the first stage are almost the same as those in the third stage. In the second stage, the first detected time is delayed, and the last detected time is earlier compared to those in the first stage. About 80% of vehicles in the first and third stages travel before 9:00, and about 80% of the vehicles finish their trips before 19:00, indicating the commuting travel patterns.

As for vehicles in Cluster 2, their frequent travel patterns in three stages are similar. The average length of frequent travel patterns and the first- and last-detected time distributions show similar trends. In all three stages, about 90% of vehicles start traveling before 9:00, and about 90% of vehicles finish their trips before 19:00, indicating the commuting travel patterns. Compared with the vehicles in Cluster 1, these vehicles start traveling earlier and end their trips earlier as well.

The average hourly detection frequency of vehicles in each cluster in each pandemic stage is calculated. The hourly detection frequency ($Q$) of vehicles in a cluster is calculated by Eq. (8).

$$Q = \frac{\sum_i D_i}{\triangle T} \tag{8}$$

where $D_i$ is the detection frequency of vehicles in the cluster detected by the $i$-th detector within period $\Delta T$.

After obtaining the hourly detection frequency of vehicles in each stage of each cluster, the average values of the hourly detection frequency of the three clusters for each of the stages are calculated, and the results are shown in Figs. 6d–6f.

As shown in Figs. 6d–6f, travelers in Cluster 0 have obvious morning and evening peak travel patterns in the first stage; in the second stage, the daily travel frequency is very low; whereas, in the third stage, the morning and evening peak travel patterns are recovered, but the travel frequency is still lower than that in the first stage. As for Cluster 1, the hourly detection frequency distributions in the first and third stages are almost the same; the travel frequency at the evening peak in the third stage is even greater than that in the first stage, while the second stage's travel frequency throughout the day is small. Lastly, for vehicles in Cluster 2, the hourly detection frequency distributions in the three stages are very similar, showing obvious morning and evening peak travel patterns.

Aiming to analyze the travel behavior of each cluster further, the number of vehicles in transit and the average travel intensity are defined. The number of vehicles in transit ($N_{VT}$) of a cluster represents the number of vehicles with different license plate numbers detected by the LPR detectors on the road network during the day. Meanwhile, the average travel intensity ($A_{TI}$) of a cluster is calculated by Eq. (9).

$$A_{TI} = \frac{\sum_i D_{di}}{N_{VT}} \tag{9}$$

where $D_{di}$ is the total detected times of vehicles in a cluster detected by the $i$-th detector during the day.

The number of vehicles in transit and the average travel intensity of each cluster are calculated on each day, and the results are shown in Figs. 6g–6i.

According to Figs. 6g–6i, the number of vehicles in transit and the average travel intensity of vehicles of Cluster 0 show similar trends; namely, they first decrease in the second stage and then increase in the third stage but do not return to the level in the first stage. The average values of $N_{VT}$ and $A_{TI}$ of Cluster 0 in the first stage are 113,221 and 35.5, respectively, while they respectively change to 79,109, and 32.8 in the third stage. Thus, compared with the first stage, in the third stage, the indicators $N_{VT}$ and $A_{TI}$ decrease by 30.1% and 7.6%, respectively.

The $N_{VT}$ and $A_{TI}$ of vehicles in Cluster 1 also show a similar trend. They decay significantly in the second stage but completely return to the level of the first stage in the third stage. In the third stage, $A_{TI}$ even slightly exceeds its value in the first stage.

For Cluster 2, $N_{VT}$ in the first stage is similar to that in the third stage, and this indicator shows a rising trend in the second stage, indicating that some

of the vehicles in this cluster have resumed their frequent travel patterns during the second stage. Further, $A_{TI}$ of vehicles in Cluster 2 is similar in the first and third stages, while it is significantly lower in the second stage compared with the values in the first and third stages. This result indicates that although $N_{VT}$ has returned to the level before the pandemic in the second stage, there is still the decay in $A_{TI}$. This reduced travel intensity has been mostly caused by the cancellation of flexible travel.

### 5.2. Influencing factors of travel behavior adjustment

Faced with the pandemic, travelers have adjusted their travel behaviors differently. These adjustments have been characterized by either homogeneity or heterogeneity. Heterogeneity indicates that, under the pandemic, specific reactions of travelers have not been the same. Heterogeneity originates from the individual differences between travelers, which are defined by differences in travelers' personal characteristics such as behavioral habits and external characteristics such as occupations. In contrast, homogeneity indicates that reactions of travelers in a cluster on the pandemic have been similar regarding certain indicators.

As explained above, after cluster analysis, travelers have been divided into three clusters, corresponding to three different categories of behavior adjusting patterns under the COVID-19 pandemic. According to the results, the behavior adjusting patterns of travelers in the same cluster under the COVID-19 pandemic have been homogenous to a certain extent. To analyze the affecting factors of the behavior pattern adjusting of different clusters in response to the pandemic, several features of travel behavior are extracted, and the random forest model is used for the analysis.

#### 5.2.1. Travel behaviors of different groups

Aiming to analyze the affecting factors of the travel behavior pattern adjusting of different groups in response to the pandemic, the travel behaviors of travelers before the pandemic are analyzed. The spatiotemporal travel behavior characteristics of the travelers are extracted. To describe the travelers' travel behaviors from the temporal perspective, a day is divided into six time periods: before dawn (00:00–4:00), early morning (4:00–6:00), morning (6:00–12:00), afternoon (12:00–17:00), evening (17:00–21:00), and night (21:00–24:00). For each traveler, the average and standard deviation of his/her daily detection frequency in each time period of the first stage, as well as the number of traveling days in each time period of the first stage, are calculated.

The origin and destination of the travelers' first trips are analyzed from the spatial perspective. For each traveler, the origin and destination of his first trip are extracted. The origin of the first trip is typically the traveler's residence location, and the destination commonly indicates where the traveler works. Therefore, it is expected that the land usage types of the origin and destination of a traveler's first trip can reflect the types of land usage of the traveler's residence and workplace, respectively. The distance between the origin of the first trip and the center of Yiwu City and the distance between the destination of the first trip and the center of Yiwu City are extracted. In addition, the Shannon entropy index mentioned in Section 3.2 is also used as a feature.

The features of the travelers' spatiotemporal travel characteristics are extracted by the above-presented method, and $rug^1$, which has been defined in Section 4.2, is also used as a feature. A total of 29 features are used to describe the spatiotemporal travel behaviors of travelers before the pandemic. The dependent variable is the label of three clusters obtained by the clustering method, which are marked as 0, 1, 2, respectively. The labels and meanings of the independent and dependent variables are given in Table 2.

Due to the large number of independent variables ($X_i$), a feature selection is performed first; the recursive feature elimination with cross-validation (RFECV) method is used as a feature selection method. The feature ranking with recursive feature elimination and cross-validated selection of the optimal number of features are performed (Pedregosa et al., 2011). In the five-fold cross-validation, accuracy is used for performance measurement. The number of samples in the dataset with different labels is different, and there is a problem of sample imbalance. Therefore, the Borderline-SMOTE (Han et al., 2005) algorithm is used to oversample the dataset, making the ratio of labels 0, 1, and 2 be 1:1:1. The final independent variables selected by the RFECV method are $X_1$, $X_4$–$X_7$, $X_8$, $X_{10}$–$X_{13}$, $X_{16}$–$X_{19}$, $X_{20}$, $X_{21}$, $X_{25}$, and $X_{26}$. The selected 18 independent variables are used to analyze the behavior adjustments of different groups in response to the pandemic further.

**Table 2**
The labels and meanings of the independent and dependent variable.

| Label | $X_1$ | $X_2$–$X_7$ | $X_8$–$X_{13}$ |
|---|---|---|---|
| Meaning | $rug^1$ | The number of traveling days in the six time periods of the first stage | The average daily detection frequency in the six time periods of the first stage |
| Label | $X_{14}$–$X_{19}$ | $X_{20}$ | $X_{21}$ |
| Meaning | The standard deviation of the daily detection frequency in the six time periods of the first stage | The distance between the first trip's starting point and the center of Yiwu | Shannon entropy index in the buffer zone of the first trip's starting point |
| Label | $X_{22}$–$X_{24}$ | $X_{25}$ | $X_{26}$ |
| Meaning | Performed one-hot encoding on the land usage category of the buffer zone of the first trip's ending point | The distance between the first trip's ending point and the center of Yiwu | Shannon entropy index in the buffer zone of the first trip's ending point |
| Label | $X_{27}$–$X_{29}$ | $Y$ | |
| Meaning | Performed one-hot encoding on the land usage category of the buffer zone of the first trip's ending point | The cluster label after clustering | |

#### 5.2.2. Analysis on travel behavior adjustments of different groups

The random forest method is used to analyze the affecting factors of travel behavior patterns adjusting under the COVID-19 pandemic. The features selected by the RFECV method are used as independent variables, and the cluster label is used as a dependent variable. To overcome the problem of sample imbalance, the Borderline-SMOTE algorithm is adopted for oversampling. After fitting the random forest model, the random forest model is used to rank the importance of the features, and the result is shown in Fig. 7.

The first six features with the highest feature importance are selected to draw partial dependence plots (Ma et al., 2021) for further analysis. The partial dependence plots of $X_4$, $X_6$, $X_5$, $X_1$, $X_{18}$, and $X_{16}$ and each of the clusters are shown in Fig. 8. The three graphs in the same column represent the changes in the relative probability that a vehicle is predicted to belong to Clusters 0, 1, and 2 as the target feature value changes. In the dependence plots, where the $y$-axis represents a change in the model prediction probability compared to the leftmost value, the coordinate values all start from zero, and the shaded blue areas represent the confidence intervals.

The dependence plots show that the more days a vehicle travels in the morning (6:00–12:00) and evening (17:00–21:00), the more likely it will be to resume its pre-pandemic frequent travel patterns as soon as possible. Travelers who have more traveling days in the morning and evening are more likely to have their usual commuting travel patterns, which is to go to work in the morning and return home from work in the evening. These travelers tend to resume their pre-pandemic frequent travel patterns as early as possible under COVID-19. This conclusion is consistent with the fact that the more regular a travel behavior of a vehicle in the first stage is, the more likely it is to resume the pre-pandemic frequent travel patterns as soon as possible.

However, more days a vehicle travels in the afternoon (12:00–17:00), the more likely it is to resume its pre-pandemic frequent travel patterns in the post-COVID-19 era or never recover its frequent travel pattern, and less likely it is to resume it in the second stage. Perhaps this is because travelers who like to travel in the afternoon are likely not commuters.

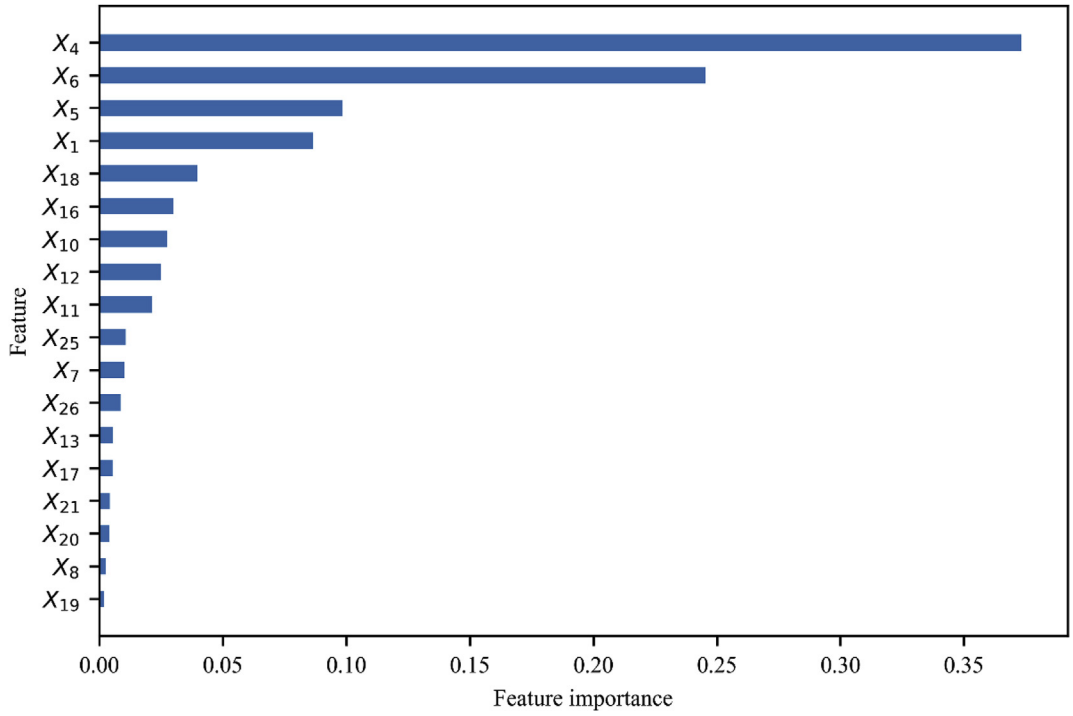



**Fig. 7.** Ranking of feature importance.

Besides, the spatial information on a traveler's origin and destination of the first trip and the types of the surrounding land usage also have a certain impact on the traveler's behavior adjustment, but the impact is smaller than those of the temporal-perspective features.

## 6. Conclusions and future prospects

Using the LPR data, this paper first extracts the travelers' spatiotemporal trip chains, then the frequent travel pattern set of each traveler based on the spatiotemporal trip chains, and finally, quantitatively describes the travel behavior change in each stage. The clustering method is used to analyze the behavior adjustments of travelers under the COVID-19 pandemic, and the spatiotemporal travel behaviors of travelers before the pandemic are studied. On this basis, the influence of spatiotemporal travel behavior before the pandemic on the behavior adjustment under the COVID-19 pandemic is analyzed. The main conclusions of this paper are as follows.

The travel behavior adjustments of travelers under the COVID-19 pandemic can be divided into three categories. The first category refers to travelers whose frequent travel patterns have greatly changed from the first stage under the COVID-19 pandemic; their trips reduced significantly in the second and third stages, and even in the post-COVID-19 era, their frequent travel patterns did not recover. The second category refers to travelers whose frequent travel patterns also have significantly changed compared to those in the first stage, but in the post-COVID-19 era, their frequent travel patterns have been recovered. The third category refers to travelers that resumed their frequent travel patterns in the initial stage of resumption of work and production. The proportions of the first, second, and third categories are 48.70%, 41.97%, and 9.32%, respectively.

The first category of vehicles (i.e., travelers) almost did not have frequent travel patterns in the second and third stages. For a traveler of this category, the time to go out which means the duration between the first travel and the last travel in a day shortens in the second and third stages. The first-category travelers have obvious morning and evening peak travel patterns in the first stage, while in the second stage, their daily travel frequency is very low, and morning and evening peak travel patterns are not obvious. In the third stage, the morning and evening peak travel patterns are recovered, but the travel frequency is still relatively low. Compared with the first stage, the number of vehicles in transit and the average travel intensity in the third stage decreased by 30.1% and 7.6%, respectively.

For the second-category vehicles, the travel pattern in the third stage is the same as that in the first stage, and the travel intensity in the third stage even exceeds that in the first stage. About 80% of the vehicles in the first and third stages start traveling before 9:00, and about 80% of the vehicles finish their travel before 19:00, indicating the commuting travel patterns.

For the third-category vehicles, their frequent travel patterns in the three stages are similar. The average length of frequent travel patterns and the first and last detected time distributions also show similar trends. In all three stages, about 90% of the vehicles start traveling before 9:00, and about 90% of the vehicles finish their trips before 19:00, indicating the commuting travel patterns. Compared with the second-category vehicles, the third-category vehicles travel earlier, and their trips end earlier. Some of the vehicles of this category resume their frequent travel patterns before the second stage, while the others recover the frequent travel patterns during the second stage. However, although the number of vehicles in transit return to the level before the pandemic in the second stage, there is still a gap in the average travel intensity in the second stage. The reduced travel intensity is most likely due to the cancellation of flexible travel.

The behavior adjusting patterns have a certain relationship with the pre-pandemic travel behavior of travelers. A traveler who has more traveling days in the morning (6:00–12:00) and evening (17:00–21:00) and a stronger travel regularity is more likely to resume his/her pre-pandemic frequent travel patterns as soon as possible. A traveler who always travels in the afternoon (12:00–17:00) may not recover his/her pre-pandemic frequent travel patterns or recover frequent travel patterns until post-COVID-19 era. In addition, the spatial information on a traveler's origin and destination and types of the surrounding land usage also have certain impacts on the traveler's travel behavior adjustment, but their impacts are smaller than those of the temporal-perspective features.

This paper uses the LPR data as a basis to analyze the spatiotemporal behavior adjustment and its influencing factors under the COVID-19 pandemic profoundly. However, due to the data dimension limitation, the range of travelers analyzed in this study is relatively small, and there is no in-depth analysis and discussion for travelers who travel without using motor vehicles. In addition, the analysis of influencing factors is not comprehensive enough, and the trip's purpose, traveler's occupation, and

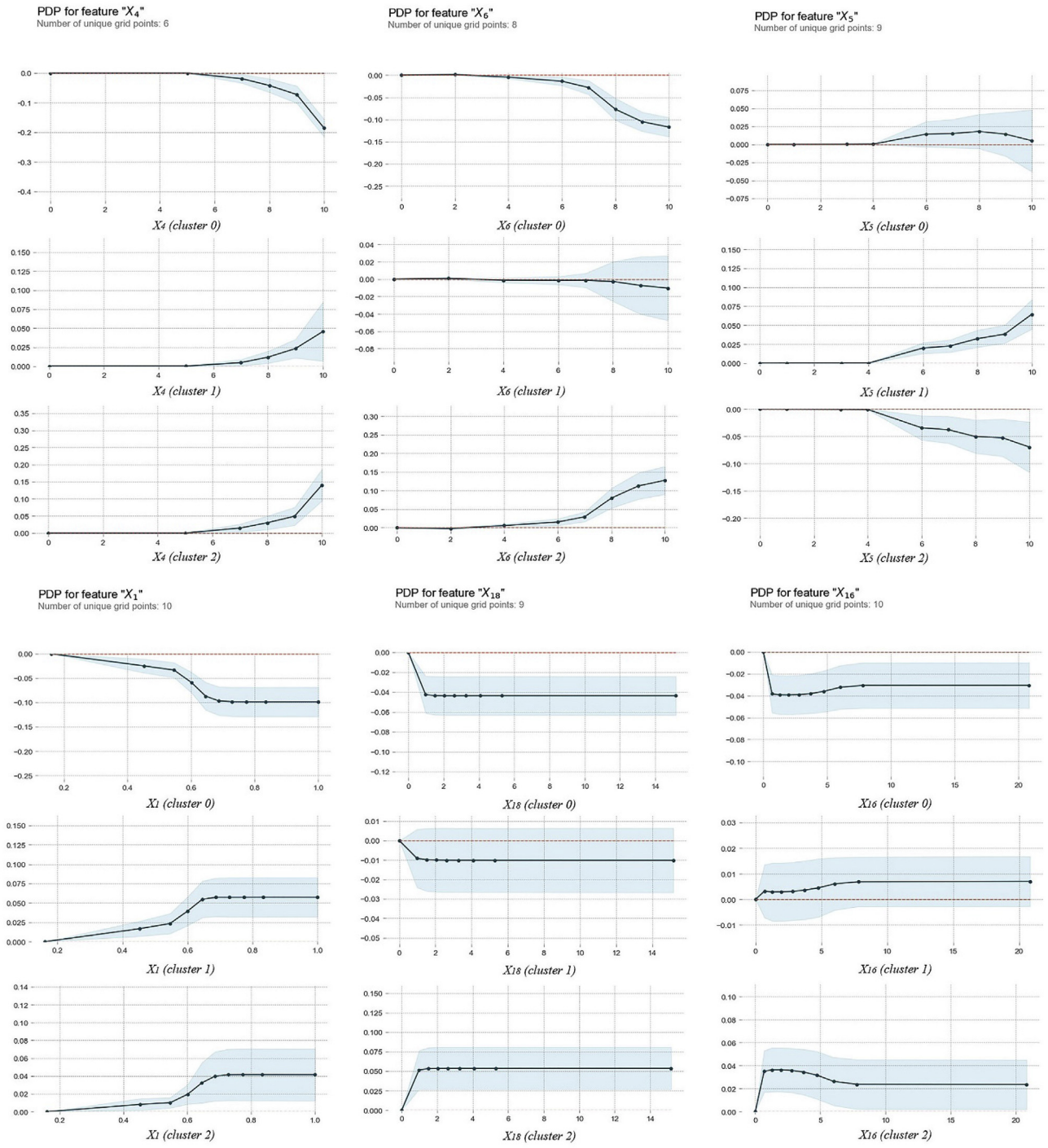



**Fig. 8.** Partial dependence plots between the top-six features and the clusters.

other factors are not considered. In the future research, multi-source data from mobile phone signaling, smart card, and bicycle sharing will be used to analyze the characteristics of urban residents' travels in all modes, and urban planning information and questionnaire survey data will be combined to expand the data dimension to conduct a more in-depth analysis of travel behaviors and their influencing factors, which can help to understand the impact of COVID-19 on transportation system and travel behavior more deeply.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This study was supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2022C01042), the National Natural Science Foundation of China (Grant No. 92046011), Center for Balance Architecture Zhejiang University, and Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies.

## Appendix

There are certain important concepts in the Prefix-Span algorithm that need to be clarified before we introduce the Prefix-Span algorithm. The first is prefix; for two sequences, $\mathbf{a} = \{P_1, P_2, P_3, \ldots, P_K\}$ and $\mathbf{a'} = \{P_1', P_2', P_3', \ldots, P_M'\}$ $(M \leq K)$, if $i \leq M$ and $P_i' = P_i$, then **a'** is a prefix of **a**, correspondingly, $\{P_{M+1}, P_{M+2}, \ldots, P_K\}$ is the postfix of **a** about **a'**. For **a**, **b**, and **a'** that denote the subsequences of **a**, if **b** is the longest subsequence of **a** that has prefix **a'**, then **b** is called the projection of **a** with prefix **a'**. For a sequence database **A** and a travel pattern $\boldsymbol{\alpha}$, $\mathbf{A}|_{\boldsymbol{\alpha}}$ is defined as a set consisting of all the postfix sequences of prefix $\boldsymbol{\alpha}$ in sequence database **A**.

The pseudo-code of the Prefix-Span algorithm is given as Algorithm A1.

**Algorithm A1**. Prefix-Span algorithm.

**Input:** a sequence database **A**, a pre-defined minimum support threshold *minsup*

**Output:** a frequent travel pattern set **F**

**Method:** Call Prefix-Span({}, 0, **A**)

**Subroutine:** Prefix-Span($\boldsymbol{\alpha}$, $l$, $\mathbf{A}|_{\alpha}$)

**Parameters:** $\alpha$—frequent travel pattern; $l$—length of frequent travel pattern; $\mathbf{A}|_{\alpha}$—$\boldsymbol{\alpha}$-projected database, if $\boldsymbol{\alpha} \neq \{\}$; otherwise, the sequence database **A**.

**Algorithm steps:**

1. Scan $\mathbf{A}|_{\alpha}$ once; find frequent items whose support values are larger than *minsup*; append them to $\boldsymbol{\alpha}$ to form a frequent travel pattern $\boldsymbol{\alpha}$'; output $\boldsymbol{\alpha}$'.

2. For each $\boldsymbol{\alpha}$', construct $\boldsymbol{\alpha}$'-projected database $\mathbf{A}|_{\alpha'}$ and call Prefix-Span($\boldsymbol{\alpha}$', $l$+1, $\mathbf{A}|_{\alpha'}$)

*Replication and data sharing*

The license plate recognition data used in this research was provided by Yiwu City Brain, and the point of interest data used in this research was crawled from the application programming interface (API) of AMAP. The code and data sample of this research can be found at https://github.com/RobinYaoWenbin/Understanding-travel-behavior-adjustment-under-COVID-19.

## References


Aloi, A., Alonso, B., Benavente, J., et al., 2020. Effects of the COVID-19 lockdown on urban mobility: empirical evidence from the city of Santander (Spain). Sustainability 12 (9), 3870.

Anzai, A., Kobayashi, T., Linton, N., et al., 2020. Assessing the impact of reduced travel on exportation dynamics of novel coronavirus infection (COVID-19). J. Clin. Med. 9 (2), 601.

Calderon-Tellez, J.A., Herrera, M.M., 2021. Appraising the impact of air transport on the environment: lessons from the COVID-19 pandemic. Transp. Res. Interdiscip. Perspect. 10, 100351.

Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. Nature 589 (7840), 82–87.

Chen, H., He, J., Song, W., Wang, L., Wang, J., Chen, Y., 2020a. Modeling and interpreting the COVID-19 intervention strategy of China: a human mobility view. PLoS One 15 (11), e0242761.

Chen, H., Yang, C., Xu, X., 2017. Clustering vehicle temporal and spatial travel behavior using license plate recognition data. J. Adv. Transport., 1738085

Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T., 2020b. COVID-19 control in China during mass population movements at New Year. Lancet 395 (10226), 764–766.

Edwards, L., Rutter, G., Iverson, L., Wilson, L., Chadha, T.S., Wilkinson, P., Milojevic, A., 2021. Personal exposure monitoring of PM2. 5 among US diplomats in Kathmandu during the COVID-19 lockdown, March to June 2020. Sci. Total Environ. 772, 144836.

Gao, F., Li, S., Tan, Z., Wu, Z., Zhang, X., Huang, G., Huang, Z., 2021a. Understanding the modifiable areal unit problem in dockless bike sharing usage and exploring the interactive effects of built environment factors. Int. J. Geogr. Inf. Sci. 35 (9), 1905–1925.

Gao, K., Yang, Y., Li, A., Qu, X., 2021b. Spatial heterogeneity in distance decay of using bike sharing: an empirical large-scale analysis in Shanghai. Transport. Res. Transport Environ. 94, 102814.

Han, H., Wang, W.Y., Mao, B.H., 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: International Conference on Intelligent Computing. Springer, Berlin, Heidelberg, pp. 878–887.

Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M., 2001. Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth. In: Proceedings of the 17th International Conference on Data Engineering. IEEE Washington, , DC, USA, pp. 215–224.

Harantová, V., Hájnik, A., Kalašová, A., 2020. Comparison of the flow rate and speed of vehicles on a representative road section before and after the implementation of measures in connection with COVID-19. Sustainability 12 (17), 7216.

Hou, M., Wang, K., Yang, H., 2021. Hub airport slot Re-allocation and subsidy policy to speed up air traffic recovery amid COVID-19 pandemic—case on the Chinese airline market. J. Air Transport. Manag. 93, 102047.

Iacus, S.M., Natale, F., Santamaria, C., Spyratos, S., Vespe, M., 2020. Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. Saf. Sci., 104791

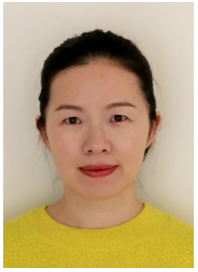
Ilesanmi, F.F., Ilesanmi, O.S., Afolabi, A.A., 2021. The effects of the COVID-19 pandemic on food losses in the agricultural value chains in Africa: the Nigerian case study. Pub. Health Pract. 2, 100087.

Jiang, P., Fu, X., Van Fan, Y., Klemeš, J.J., Chen, P., Ma, S., Zhang, W., 2020. Spatial-temporal potential exposure risk analytics and urban sustainability impacts related to COVID-19 mitigation: a perspective from car mobility behaviour. J. Clean. Prod. 279, 123673.

Jiao, J., Bhat, M., Azimian, A., 2021. Measuring travel behavior in Houston, Texas with mobility data during the 2020 COVID-19 outbreak. Transport. Lett. 1–12.
Kerimray, A., Baimatova, N., Ibragimova, O.P., Bukenov, B., Kenessov, B., Plotitsyn, P., Karaca, F., 2020. Assessing air quality changes in large cities during COVID-19 lockdowns: the impacts of traffic-free urban conditions in Almaty, Kazakhstan. Sci. Total Environ., 139179
Klise, K., Beyeler, W., Finley, P., Makvandi, M., 2021. Analysis of mobility data to build contact networks for COVID-19. PLoS One 16 (4), e0249726.
Lee, H., Park, S.J., Lee, G.R., Kim, J.E., Lee, J.H., Jung, Y., Nam, E.W., 2020. The relationship between trends in COVID-19 prevalence and traffic levels in South Korea. Int. J. Infect. Dis. 96, 399–407.
Li, A., Gao, K., Zhao, P., Qu, X., Axhausen, K.W., 2021a. High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data. J. Clean. Prod. 296, 126423.
Li, A., Zhao, P., He, H., Ali, M., Axhausen, K.W., 2021b. How did micro-mobility change in response to covid-19 pandemic? a case study based on spatial-temporal-semantic analytics. Comput. Environ. Urban Syst. 90, 101703.
Liu, Y., Zhan, Z., Zhu, D., Chai, Y., Ma, X., Lun Wu, L., 2018. Incorporating multi-source big geo-data to sense spatial heterogeneity patterns in an urban space. Geomatics Inf. Sci. Wuhan Univ. 43 (3), 327–335.
Luo, X., Jin, S., Gong, Y., Wang, D., Ma, D., 2019a. Queue length estimation for signalized intersections using License Plate Recognition data. IEEE Intell. Transport. Syst. Mag. 11 (3), 209–220.
Luo, X., Wang, D., Ma, D., Jin, S., 2019b. Grouped travel time estimation in signalized arterials using point-to-point detectors. Transport. Res. Part B 130, 130–151.
MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.
Marinello, S., Lolli, F., Gamberini, R., 2021. The impact of the COVID-19 emergency on local vehicular traffic and its consequences for the environment: the case of the city of reggio emilia (Italy). Sustainability 13 (1), 118.
Ma, D., Luo, X., Jin, S., Guo, W., Wang, D., 2018. Estimating maximum queue length for traffic lane groups using travel times from video-imaging data. IEEE Intell. Transport. Syst. Mag. 10 (3), 123–134.
Ma, D., Luo, X., Li, W., Jin, S., Guo, W., Wang, D., 2017a. Traffic demand estimation for lane groups at signal-controlled intersections using travel times from video-imaging detectors. IET Intell. Transp. Syst. 11 (4), 222–229.
Ma, D., Luo, X., Wang, D., Jin, S., Wang, F., Guo, W., 2017b. Lane-based saturation degree estimation at signalized intersections using travel time data. IEEE Intell. Transport. Syst. Mag. 9 (3), 136–148.
Ma, S., Li, S., Zhang, J., 2021. Diverse and nonlinear influences of built environment factors on covid-19 spread across townships in China at its initial stage. Sci. Rep. 11 (1), 12415.
Mei, C., 2020. Policy style, consistency and the effectiveness of the policy mix in China's fight against COVID-19. Pol. Soc. 39 (3), 309–325.
Muley, D., Ghanim, M.S., Mohammad, A., Kharbeche, M., 2021. Quantifying the impact of COVID–19 preventive measures on traffic in the State of Qatar. Transport Pol. 103, 45–59.
Mihai, F.C., Ursu, A., Ichim, P., Chelaru, D.A., 2015. Determining Rural Areas Vulnerable to Illegal Dumping Using GIS Techniques. Case study: Neamt county, Romania. arXiv preprint arXiv:1507.00627.
Munshi, J., Roy, I., Balasubramanian, G., 2020. Spatiotemporal Dynamics in Demography-Sensitive Disease Transmission: COVID-19 Spread in NY as a Case Study arXiv preprint arXiv:2005.01001.
Orro, A., Novales, M., Monteagudo, Á., Pérez-López, J.B., Bugarín, M.R., 2020. Impact on city bus transit services of the COVID–19 lockdown and return to the new Normal: the case of A Coruña (Spain). Sustainability 12 (17), 7206.
O'Sullivan, D., Gahegan, M., Exeter, D.J., Adams, B., 2020. Spatially explicit models for exploring COVID-19 lockdown strategies. Trans. GIS 24 (4), 967–1000.
Patra, S.S., Chilukuri, B.R., Vanajakshi, L., 2021. Analysis of road traffic pattern changes due to activity restrictions during COVID-19 pandemic in Chennai. Transport. Lett. 13, 1–9.
Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.
Saraf, P., Sedamkar, R.R., Rathi, S., 2015. Prefixspan algorithm for finding sequential pattern with various constraints. Int. J. Appl. Info. Syst. 9 (3), 2249-0868.
Shen, X., Zhou, Y., Jin, S., Wang, D., 2020. Spatiotemporal influence of land use and household properties on automobile travel demand. Transport. Res. Part D 84, 102359.
Shou, Z., Di, X., 2018. Similarity analysis of frequent sequential activity pattern mining. Transport. Res. C Emerg. Technol. 96, 122–143.
Sicard, P., De Marco, A., Agathokleous, E., et al., 2020. Amplified ozone pollution in cities during the COVID-19 lockdown. Sci. Total Environ. 735, 139542.
Sun, J., He, W.T., Wang, et al., 2020. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol. Med.
Vandoros, S., 2021. Covid-19, Lockdowns and Motor Vehicle Collisions: Empirical Evidence from Greece. MedRxiv, 2020-12.
Wang, D., Fu, F., Luo, X., Jin, S., Ma, D., 2016. Travel time estimation method for urban road based on traffic stream directions. Transportmetrica 12 (6), 479–503.
Wang, C., Huang, X., Hu, X., Zhao, L., Liu, C., Ghadimi, P., 2021a. Trade characteristics, competition patterns and COVID-19 related shock propagation in the global solar photovoltaic cell trade. Appl. Energy 290, 116744.
Wang, F., Tan, Z., Yu, Z., Yao, S., Guo, C., 2021b. Transmission and control pressure analysis of the COVID-19 pandemic situation using multisource spatio-temporal big data. PLoS One 16 (3), e0249145.
Wang, Q., Su, M., 2020. A preliminary assessment of the impact of COVID-19 on environment – a case study of China. Sci. Total Environ. 728, 138915.
Wang, Y., Yuan, Y., Wang, Q., Liu, C., Zhi, Q., Cao, J., 2020. Changes in air quality related to the control of coronavirus in China: implications for traffic and industrial emissions. Sci. Total Environ. 731, 139133.
Wilhelmsen, C.K., Skalleberg, K., Raanaas, R.K., Tveite, H., Aamodt, G., 2017. Associations between green area in school neighbourhoods and overweight and obesity among Norwegian adolescents. Prevent. Med. Rep. 7, 99–105.
World Health Organization, 2022. Weekly Epidemiological Update on COVID-19 - 8 March 2022. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—8-march-2022.
Xu, X., Wang, S., Dong, J., Shen, Z., Xu, S., 2020. An analysis of the domestic resumption of social production and life under the COVID-19 pandemic. PLoS One 15 (7), e0236387.
Yang, D.L., 2021. The COVID-19 pandemic and the estrangement of US-China relations. Asian Perspect. 45 (1), 7–31.
Yao, W., Zhang, M., Jin, S., Ma, D., 2021. Understanding vehicles commuting pattern based on license plate recognition data. Transport. Res. Part C 128, 103142.
Yao, W., Chen, C., Su, H., Chen, N., Jin, S., Bai, C., 2022. Analysis of key commuting routes based on spatiotemporal trip chain. J. Adv. Transport., 6044540
Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M., 2013. Time-aware point-of-interest recommendation. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 363–372.
Zhang, J., Feng, B., Wu, Y., Xu, P., Ke, R., Dong, N., 2021. The effect of human mobility and control measures on traffic safety during COVID-19 pandemic. PLoS One 16 (3), e0243263.
Zheng, H., Zhang, K., Nie, M., 2020. The Fall and Rise of the Taxi Industry in the COVID-19 Pandemic: A Case Study. Available at SSRN 3674241.

**Wenbin Yao** is currently a Ph.D. candidate in the Institute of Intelligent Transportation Systems, College of Civil Engineering and Architecture, Zhejiang University. His main research interests are spatiotemporal big data analysis, travel behavior analysis, and traffic flow theory.

**Jinqiang Yu** received both the B.Eng. and Ph.D. degrees in electrical engineering from National University of Singapore, in 2008 and 2015, respectively. He is currently a Senior Algorithm Engineer with Alibaba Cloud Computing. His research is focused on intelligent transportation systems, traffic light control, autonomous and connected vehicle, etc.



**Ying Yang** is currently a Lecturer in psychology from Australian Catholic University. She received her Bachelor's degree from Renmin University of China and Ph.D. degree from Beijing Normal University. Her research is focused on cross-cultural psychology, behavioral decision theory, and traffic psychology. She has great interest and passion in understanding how people from different cultures perceive, think, and behave. She is an awardee of the prestigious Australian Government Endeavour Research Fellowship in 2016, and earned ACU Women researcher award in 2020. Her studies were published at *Journal of Cross-Cultural Psychology*, *Personality and Individual Differences*, *Transportation Research Part A*, *Part D*, and *Part F*.

**Nuo Chen** is currently a M.S. candidate in the Institute of Intelligent Transportation Systems, College of Civil Engineering and Architecture, Zhejiang University. She focuses on traffic flow forecasting and big data mining.

**Sheng Jin** is currently a Professor in Institute of Intelligent Transportation Systems, Zhejiang University. He received both the B.Eng. and Ph.D. degrees in transportationengineering from Jilin University, in 2005 and 2010 respectively. His research is focused on traffic flow theory, traffic signal control, shared mobility, and big data and machine learning for transport systems. He published over 80 articles in journals of transportation engineering including *IEEE Trans. Intelligent Transportation Systems*, *Transportation Research Part A*, *Part B*, *Part C*, and *Part D*.

**Youwei Hu** is currently a M.S. candidate in the Institute of Intelligent Transportation Systems, College of Civil Engineering and Architecture, Zhejiang University. Her research interests include traffic planning and travel behavior analysis.

**Congcong Bai** is currently a Ph.D. candidate in the Institute of Intelligent Transportation Systems, College of Civil Engineering and Architecture, Zhejiang University. He received his B.S. degree in Zhejiang University, China. He focuses on the transportation systems modeling, transportation big data analysis, and driver behavior analysis.