


## Research Article

# Travel Patterns Analysis Using Tensor-Based Model from Large-Scale License Plate Recognition Data

Zhaoxin Liu,<sup>1</sup> Xiaolu Wang,<sup>2</sup> Yufeng Bi,<sup>3</sup> Jun Kong,<sup>1,4</sup> Run Xu,<sup>3</sup> Yuanpei Chen,<sup>1,4</sup> and Jinjun Tang<sup>2</sup> 

<sup>1</sup>Shandong Hi-speed Infrastructure Construction Co. LTD, Jinan 250000, China

<sup>2</sup>Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

<sup>3</sup>Shandong Provincial Communications Planning and Design Institute Group Co. LTD, Jinan 250000, China

<sup>4</sup>Shandong Hi-speed Jiwei Expressway Co. LTD, Jinan 250000, China

Correspondence should be addressed to Jinjun Tang; jinjuntang@csu.edu.cn

Received 13 April 2022; Accepted 8 August 2022; Published 31 August 2022

Academic Editor: Guohui Zhang

Copyright © 2022 Zhaoxin Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Travel patterns reflect the regularity of residents' mobility, and it is a crucial factor to evaluate the reasonability of urban spatial structure and connectivity of road networks. Therefore, exploring travel patterns is of practical significance for urban planning, traffic management, and improvement of the operational efficiency of the transportation system. In this study, we apply the tensor model to explore travel patterns under temporal and spatial dimensions based on the license plate recognition (LPR) data collected from the Changsha city, China. As travel patterns are influenced by many variables, a method framework based on the tensor model is proposed to explore the influence of variables on travel characteristics. Firstly, we apply clustering algorithms and the principal component analysis method to extract main feature variables, which can achieve the purpose of dimensionality reduction and eliminate the complex collinearity among variables. Then, the tensor decomposition and reconstruction algorithms are performed based on extracted feature variables to analyze their influence on travel patterns. The experiments demonstrate the advantages of the proposed method framework.

## 1. Introduction

The transportation system is the key part of cities, which affects the social activities of citizens and urban structure. How to enhance the efficiency of the transportation systems operation becomes the primary task in urban planning. Travel patterns reflect the regularity of residents' mobility, and it is a crucial factor to evaluate the reasonability of urban spatial structure and connectivity of urban traffic networks. Therefore, understanding travel spatiotemporal characteristics is significant to better explore travel patterns and provide theoretical support for urban planning and traffic management. With the rapid development of information and communication technologies (ICT), large quantities of trajectory data that record residents' activities at both temporal and spatial scales have become available, which

include mobile phones [1–4], license plate recognition (LPR) [5, 6], Global Positioning System (GPS) [7–10], and so on. Given its strength in rich information, trajectory data attracts more and more attention in revealing spatiotemporal characteristics of residents' travel. The mobile phone data has been applied for studying the general human travel behavior due to its huge samples [2]. Especially, special categories of users, such as different age groups, can also be studied without the limit of the sample coverage. Furthermore, supporting the convenient collection of large volumes of individual trajectory data, GPS and LPR provide great opportunities to study various mobility patterns. As a main part of the transportation sector, taxis provide accessible and flexible travel services for people living in urban centers [5, 8]. A careful analysis of taxi GPS data can provide an innovative strategy to improve the quality of public transit

services and facilitate urban public transit planning and operational decision-making. Recently, much research has been conducted to investigate the spatial-temporal patterns of human mobility in urban cities such as Shanghai city [11] and Shenzhen city in China [12].

Furthermore, large amounts of works have focused on traffic state estimation and the prediction using travel trajectories. Tang et al. [13] developed a data-driven approach based on an HMM to predict the future links on which the vehicle may travel. Zhan et al. [14] proposed a complete solution using LPR data for link-based traffic state estimation and prediction for arterial networks. Naserian et al. [15] proposed a novel group discovery approach taking the people movement behavior into consideration to predict personalized location for group travelers from spatial-temporal trajectories. Several approaches [16–18] addressed the location prediction as a historical movement matching problem. They considered the resident's movement trajectory as a sequence of locations and then extracted the frequent movement patterns from the set of trajectories. These frequent patterns were used as the prediction rules to be matched with the previous movement of the residents.

Travel patterns are influenced by many variables, such as land type [19–21], travel purpose [22, 23], and road networks [24, 25]. As one of the important branches of data mining technology, multivariable influence analysis has attracted more attention and application in the field of transportation to explore the influence degree of variables [1, 26, 27]. In fact, the performance of the goal is the embodiment of the combinations of multivariable. The results of multivariable influence analysis reflect travel behavior under different combinations of variables. There is also a series of researches focusing on the influence of variables on the target [28–30]. Truong et al. [30] developed a quantitative approach using hierarchical variance analysis, which deals with the exploration of the relevant factors and the confirmation of their significant contribution to analyzing the residents' perception of tourism impacts. In the previous study, various researches have been implemented focusing on multivariable influence analysis on statistical and regression levels [31–33]. Arora et al. [32] used a hierarchical regression prediction model to explore the influence of celebrity endorsement factors or consumer factors on shoppers' purchase intentions. Wang et al. [33] developed univariate negative binomial conditional autoregressive (NB-CAR) and bivariate negative binomial spatial conditional autoregressive (BNB-CAR) models to analyze the influencing factors.

However, there are still several limitations and challenges that need further exploration and discussion.

Firstly, the conventional matrix can only be used to analyze the joint influence of two variables. Li et al. [34] put forward a kind of multivariable emotion mode that is based on factors such as character, mood, and emotion. They generated the emotional factor matrix through the statistical analysis of the corresponding rules, which are fused into the changes in character and mood, emotional activation threshold, and emotional interaction. If variables are represented by vectors separately, the structural features of the

original samples cannot be extracted, and the correlation and complementarity between factors cannot be explored. Furthermore, the correlation between variables affects the analyzing accuracy. If some related variables are analyzed as independent components directly, it may lead to the multicollinearity of the model. Then we use a simple matrix model to model these two variables, the matrix is likely to have a low rank, and the accuracy will be greatly reduced when calculating eigenvalue or singular value decomposition. Therefore, it is necessary to design an improved method to enrich the understanding of travel patterns and the influence of multivariable on vehicle trips from a large amount of data.

Secondly, in the previous studies, researchers generally analyze the degree of independent influence and joint influence of variables separately [35, 36]. The existing methods cannot analyze the degree of independent influence and the degree of joint influence at the same time. The conventional independent influence analysis methods, such as the unary linear regression method, obviously cannot perform multivariable joint influence analysis. For the joint influence degree analysis method, although multivariable regression [36] can be degraded into the unitary regression to carry out independent influence degree analysis after reducing the number of selected variables, this is equivalent to a repeated analysis process. So, constructing a hybrid analysis method is helpful to make a reasonable evaluation of the degree of independent influence and joint influence of variables.

Thirdly, some meaningful researches have focused on using machine learning methods to explore multivariable influence degree analysis [37, 38]. However, methods based on machine learning methods, such as support vector regression, whose calculation of model parameters are closely related to each sample data, a large data set will increase the calculation cost, making it difficult for the algorithm to analyze the results.

Compared with the conventional traffic data sources, the LPR data set can be used to monitor urban traffic networks with extensive coverage areas. Meanwhile, the greatest advantage of LPR data lies in the rich and unique information provided. LPR data records almost all the passing vehicles at an intersection, whereas the GPS vehicle data only covers a small fraction of vehicles in the traffic. The information in an LPR record generally includes vehicle ID, intersection ID, record time, and so on. The same vehicle can be tracked at multiple intersections in the road network. Therefore, it allows us to reconstruct the trajectory chain of each individual vehicle by summarizing a series of spatiotemporal records. Given extensive coverage, high frequency, and great precision, LPR data is crucial for supporting studies in the field of transportation. Furthermore, the tensor model can summarize multivariate categorical data into a multidimensional array [39]. The goal of tensor decomposition is to efficiently reproduce the higher-order interactions between different orders in multivariate data by using simple structures with relatively few parameters. Given its strength in dealing with multidimensional data, the tensor model becomes increasingly important in interpreting the structure of complex data sets. It has been successfully applied in a

variety of fields, such as signal processing [40, 41] and image processing [42, 43]. Recently, the tensor model has also attracted more attention in the field of transportation data analysis. Integrating taxi trajectory data with other urban sensing data, Zhang et al. [44] modeled urban refueling events as a three-order tensor (gas station  $\times$  hour  $\times$  day) and adapted Tucker decomposition to take features of gas stations into account. Using a large taxi trajectory data set in Beijing, Wang et al. [45] modeled collective mobility as a three-way (origin  $\times$  destination  $\times$  time, with a large size of  $651 \times 651 \times 24$ ) tensor, with each cell  $(i, j, k)$  corresponding to traffic volume from zone  $i$  to zone  $j$  at time domain  $k$ . Although many scholars have applied a tensor model in many fields, there are few works to explore travel patterns using LPR data and the influence of variables on travel patterns based on the tensor model.

Above all, we apply the tensor model to explore travel characteristics in the temporal and spatial dimensions based on the LPR data of the Changsha city. Then a method framework is proposed to analyze the influence of variables on travel characteristics. The clustering algorithms are performed on all variables to construct feature variables for each cluster, which achieves the purpose of dimensionality reduction and eliminates the complex collinearity among variables. Next, tensor decomposition and reconstruction are used to identify important variables in vehicle travel characteristics.

The remainder of this paper is organized as follows. In Section 2, we describe the data sets of this paper. In Section 3, we introduce a method framework based on the clustering algorithm, PCA, and tensor decomposition and reconstruction model. The case study of spatiotemporal characteristics analysis for travel patterns in the Changsha city is introduced in Section 4, and the influence of variables on vehicle travel characteristics is further discussed in Section 5. The last section summarizes the conclusions of this study.

## 2. Data Description

**2.1. LPR Data Set.** The data are collected from the LPR system in the Changsha city, China, which can detect over 100 million records in a week, involving 782 intersections. Figure 1 shows the distribution of the LPR detector in the road network, where the red points represent the detectors and the black lines represent the road network.

The data collected from the LPR system used in this study including vehicle ID, intersection ID, record time, direction number, and lane number are shown in Table 1. Considering protecting the privacy of travelers, the vehicle IDs are transferred to a unique number. It is worth noting that the direction number indicates the direction of a vehicle running through the intersection, and the “1” represents the east, “2” represents the west, “3” represents the south, and “4” represents the north. The lane number indicates the number of lanes from outside to inside in different directions.

In this study, data collected from the LPR system in the Changsha city, from July 1st to 7th, 2019, for seven days (a week) are used in the experiment. However, data

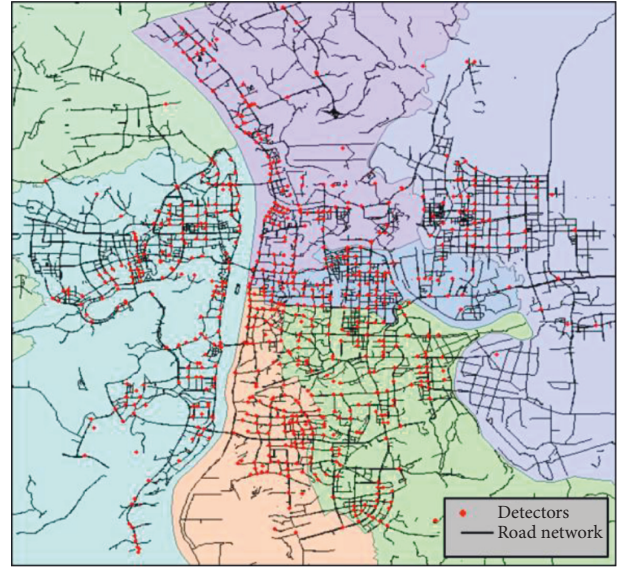


FIGURE 1: Road network and distribution of LPR system in the Changsha city.

transmission error is a common and unavoidable issue that exists in the LPR system. Due to device malfunctions and transmission distortion, incomplete data, wrong data, and duplicate data exist in LPR records. It is necessary to conduct a data cleaning procedure before further data analysis and application. We establish a systematic framework and method for data cleaning in this section, which includes eliminating noise data, identifying individual trips of all vehicles, and extracting the traffic characteristics of intersections in the whole road network.

We identify individual trips of all vehicles running in the whole road network from the real large-scale LPR data. Firstly, we delete the erroneous LPR data. There are two kinds of erroneous data in this study. (a) The LPR records are incomplete or duplicate. If there are two identical records, only one needs to be kept. (b) The detected time in the record exceeds the time of 0:00–24:00, and the detected intersections in the record are beyond the scope of the Changsha city. Secondly, the trip trajectory chain is extracted from the LPR data of the vehicle within one day; the trajectory data can be represented as following forms: vehicle ID, time (1), position (1), time (2), position (2), . . . , time (n), position (n). We distinguish single trips by setting the trip time threshold (30 minutes), which means that if two consecutive detected times meet a certain threshold, this trip is considered two individual trips.

**2.2. POI Data Set.** Urban points of interest (POI) data refer to landmarks buildings and geographic entities that are closely related to people’s daily life, which can reflect urban spatial structure. In this study, the POI data of the Changsha city are used to calculate land use around the intersection to reflect the attraction to the drivers. The POI data are acquired through the requests module in Python using the API provided by the Baidu map development platform. The POI data are divided into 11 categories according to presetting

TABLE 1: Row data sample.

Vehicle ID	Intersection ID	Record time	Direction number	Lane number
XLTY075ARP	67519316465721	July 1, 2019, 0:01:02	1	2
XEPP337PKE	12462432645262	July 1, 2019, 5:00:18	2	1
XKKB186ASD	58912564216451	July 1, 2019, 3:12:08	2	3
XYGD998HKL	26597891561720	July 1, 2019, 7:21:01	4	4
XOMA034CNA	26145906505920	July 1, 2019, 10:45:03	3	6

TABLE 2: POI category classification.

category	Contents
Traffic facilities	Airports, railway stations, bus stations, ports, etc.
Leisure	Cinema, KTV, theater, shopping center, etc.
Company	Company, factory, etc.
Medical facilities	General hospitals, specialized hospitals, clinics, etc.
Residential area	Residential area, dormitory
Education institution	Colleges, elementary schools, kindergartens, etc.
Tourist sites	Parks, zoos, gardens, amusement parks, museums, etc.
Living service	Post offices, ticket offices, laundries, photo shops, etc.
Food	Chinese restaurants, foreign restaurants, cake, etc.
Shopping	Shopping centers, supermarkets, markets, etc.
Hotels	Star hotels, apartment, residential accommodation, etc.

functional attributes, which include residential area, corporate enterprise, shopping, transportation facilities, education and training, hotel, tourist attraction, food, life service, leisure and entertainment, and medical treatment. The records in each category include category attribute, POI name, and longitude and latitude information; 17,858 pieces of POI data in the Changsha city are acquired. We then calculate the number of POI (indicated by POI) within a range of 1 km around the intersection. Table 2 shows the POI category classification.

Travel patterns are influenced by many variables. Comparing the performance of the model with and without POI data, Bao et al. [46] verified the importance of POI data in exploring travel patterns and trip purposes. Meanwhile, POI data can further refine the characteristics of residents' travels in the spatial dimension [47]. The quantitative characterization of travel origin and destination indicates the potential dynamic demand in the travel demand network and is evolved by the distribution of activities and interaction between places in cities [48]. Wang et al. [7] analyzed the spatial distribution of the pick-up points and the drop-off points to characterize the spatial patterns of trajectory travels. The characteristics of the road network, such as the number of regulated and unregulated intersections and the number of lanes, have a significant effect on the travel of urban vehicles [49]. The peak hour factor (PHF) characterizes the fluctuations of traffic flow based on the busiest 15 minutes during the peak hour. This parameter is used in the process of evaluating the traffic flow conditions such as capacity and level of service [50]. The section non-equilibrium factor of vehicle flow (NEF) is an important index for measuring the state of vehicle travel at intersections.

The O-point data represents the trip generation; the D-point data represents the trip attraction capacity. PHF<sub>15</sub> reflects the proportion of vehicle trips at intersections in

peak hours; NEF reflects the equilibrium degree of vehicle trips in each time period at intersections. The number of POI within a radius of 1 km at intersections represents the attraction of land use to the vehicle, and the number of lanes at the intersections is also one of the important variables affecting the travel patterns. Taking into account the influence of the characteristics of land use, traffic distribution, and road network, we selected O, D, PHF<sub>15</sub>, NEF, POI, and Lane in this paper as the original variables to explore the influence on travel patterns.

### 3. Methodology

The framework of the proposed model is illustrated in Figure 2, which consists of two phases, that is, explore travel patterns and analyze the influence of variables on travel patterns. In the first phase, a fourth-order tensor is constructed, and the basic travel patterns are explored by tensor CP decomposition. In the second phase, 6 variables are calculated from LPR data and POI data. We use clustering algorithms and the principal component analysis (PCA) method on six variables to extract the main feature variables. The tensor decomposition and reconstruction algorithms are performed based on extracted feature variables to analyze their influence on travel patterns.

*3.1. Clustering Algorithms.* The  $K$ -means clustering algorithm originated from a vector quantization method in signal processing, which is popular in the field of data mining [51–53]. Given a data set, the  $K$ -means clustering algorithm can find  $K$  different clusters, and the center of each cluster is calculated by using the mean value of the values contained in the clusters. Assuming that we extract  $n$  objects  $(x_1, x_2, \dots, x_n)$  of the original data and each object  $x_i$  is a vector of  $d$  dimensions, which consists of  $d$  features of the

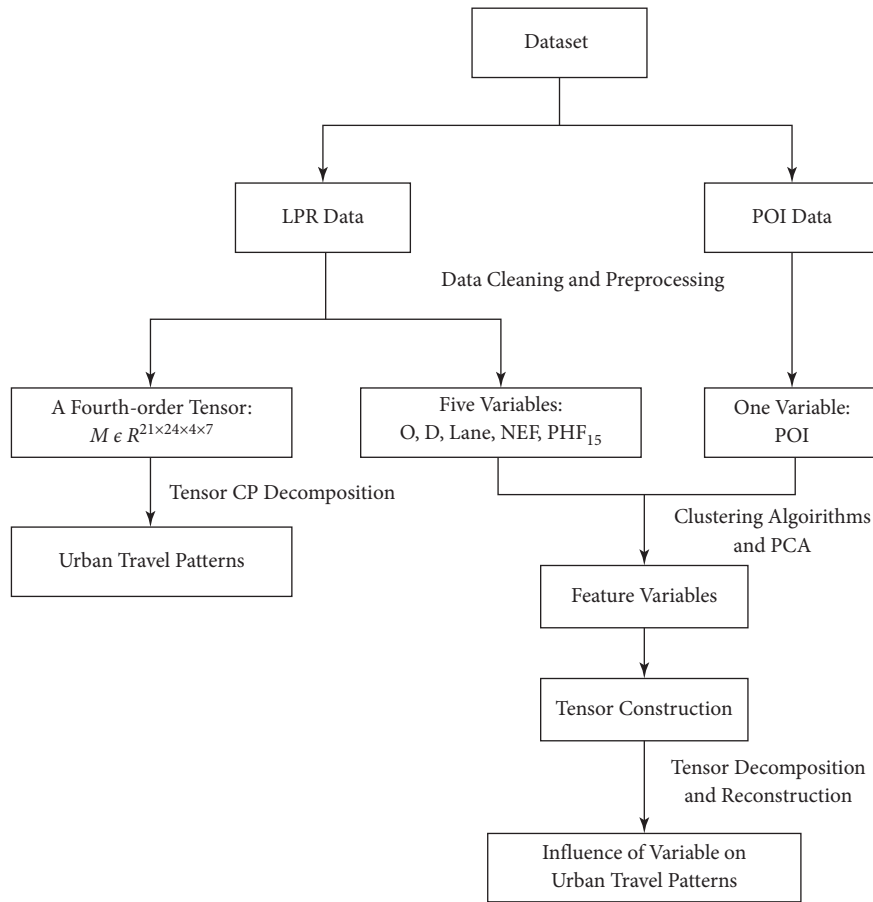


FIGURE 2: Model framework.

original data. Given the value of the number of classification groups  $k \leq n$ , the purpose of the  $K$ -means clustering algorithm is to divide the original data into  $k$  clusters ( $S = \{S_1, S_2, \dots, S_k\}$ ). However,  $K$ -means clustering algorithm may cause errors in the final clustering results when unreasonable extreme values appear in the sample data.

The  $K$ -medoids clustering algorithm [54, 55] has advantages in weakening the influence of unreasonable extreme values. Different from calculating the mean value of the values contained in the clusters when the  $K$ -means clustering algorithm updates the cluster center, the  $K$ -medoids clustering algorithm firstly calculates the distance between each point and other points in the cluster and then selects the point with the smallest distance as the new cluster center.

The maximum and minimum distance clustering algorithm [56, 57] takes the far object as the cluster center based on Euclidean distance. There are  $n$  objects  $(x_1, x_2, \dots, x_n)$ ; a random sample object  $x_i$  is used as the first cluster center. Then another object  $x_j$  farthest from the first cluster center is selected as the second cluster center. Next, other cluster centers are determined based on this principle until no new cluster centers are generated. Finally, the samples are classified into the nearest cluster according to the minimum distance. This clustering algorithm can avoid the situation that the initial cluster centers may be too closed.

The three clustering algorithms mentioned above are chosen in this study to divide the original variables into  $k$  feature variables to achieve the purpose of dimensionality reduction. The original variables include O, D, PHF<sub>15</sub>, NEF, Lane, and POI. We set several cluster numbers and apply the Elbow method to find the optimal cluster numbers. The variables are related to each other within the same cluster but independent of other clusters. Therefore, the variables that belong to the same cluster can be combined to obtain a new variable, which is called the feature variable of this cluster.

**3.2. Principal Component Analysis.** PCA [58–60] transforms a group of variables that may have a correlation into several new uncorrelated comprehensive indicators by orthogonal transformation. A few principal components are extracted to reveal the internal structure of multiple variables. They retain as much information as possible about the original variables and are not related to each other. Assuming that certain research involves  $P$  indicators, use  $X = (X_1, X_2, \dots, X_n)$  to represent them. The mean of  $X$  is  $\mu$ , and the covariance matrix of  $X$  is  $\Sigma$ . New uncorrelated comprehensive indicators are expressed in mathematical notation as equation (1). To simplify the model structure, only comprehensive indicators with large variances are usually selected in research.

$$\begin{aligned}
\mathbf{Y}_1 &= \mathbf{u}_{11}\mathbf{X}_1 + \mathbf{u}_{12}\mathbf{X}_2 + \cdots + \mathbf{u}_{1p}\mathbf{X}_p, \\
\mathbf{Y}_2 &= \mathbf{u}_{21}\mathbf{X}_1 + \mathbf{u}_{22}\mathbf{X}_2 + \cdots + \mathbf{u}_{2p}\mathbf{X}_p, \\
&\dots \\
\mathbf{Y}_p &= \mathbf{u}_{p1}\mathbf{X}_1 + \mathbf{u}_{p2}\mathbf{X}_2 + \cdots + \mathbf{u}_{pp}\mathbf{X}_p.
\end{aligned} \tag{1}$$

The PCA is a common method to reduce dimension and eliminate multicollinearity of data. However, the practical significance of feature variables generated by PCA is not clear, which does not meet the requirements of modeling and index in the subsequent analysis process of this study. Therefore, we utilize clustering algorithms to reduce dimension and apply the PCA method to generate feature variables by a linear combination of original related variables in the same cluster.

**3.3. Tensor Model.** Feature variables reflect the characteristics and attributes of this cluster. It is crucial to characterize the relationship between feature variables with an appropriate mathematical model. However, vectors and matrices are difficult to deal with the issues concerning the multi-dimensional features of original data, which cannot retain the structure information of the original data. Tensor is a high-order generalization of a matrix, which can be considered to be constructed by the product of multiple vector spaces. Given its strength in retrieving and storing information, the tensor model is applied in this study to explore travel patterns from LPR records.

**3.3.1. Tensor Construction.** Assuming that there are  $n$  variables ( $n \geq 1$ ) in the sample data set, which are denoted as  $x_1, x_2, \dots, x_n$  ( $x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n$ ).  $X_1, X_2, \dots, X_n$  are the value sets of these  $n$  variables, respectively. An  $n$ -order tensor  $X \in R^{I_1 \times I_2 \times \dots \times I_n}$  can be constructed to describe the relationship between  $n$  variables, where  $I_1 \times I_2 \times \dots \times I_n$  represent the dimensions of these  $n$  variables. The value of the elements  $X_{i_1 i_2 \dots i_n}$  is equal to the average values of the target data. Therefore, a certain tensor can be conducted to explore the spatiotemporal characteristics of travel patterns.

Furthermore, we also want to explore the influence of several variables on travel patterns. The problem is transformed into finding the value of the dependent variable in the case of a corresponding value combination of several independent variables. Assuming that there are  $m$  new feature variables ( $1 \leq m \leq n$ ) after clustering, which are denoted as  $x_{i_1}, x_{i_2}, \dots, x_{i_m}$  ( $x_{i_1} \in X_{i_1}, x_{i_2} \in X_{i_2}, \dots, x_{i_m} \in X_{i_m}$ ). Considering each specific sample, we create an  $m$ -tuple of the form  $(x_{i_1}, x_{i_2}, \dots, x_{i_m})$  for each sample, and the  $j_{th}$  sample that corresponds to the  $m$ -tuple is expressed as  $(x_{j i_1}, x_{j i_2}, \dots, x_{j i_m})$ , where  $x_{j i_k}$  represents the value of the  $i_k$  ( $1 \leq k \leq m$ ) new independent variable of the  $j_{th}$  sample, and the value of the  $m$ -tuple is equal to the value of the dependent variable in the sample  $y_j, y_j \in Y$ . Taking the actual LPR data as an example, the  $m$ -tuple  $x_{j i_1}, x_{j i_2}, \dots, x_{j i_m}$  represents the combination of the influencing variables of the sample, and the value of the  $m$ -tuple represents the actual trip volume of the intersection under the combined conditions.

**3.3.2. Tensor Decomposition.** We refer to the CANDECOMP/PARAFAC decomposition as CP decomposition. Tensor CP decomposition is a high-order generalization of singular value decomposition (SVD). The CP decomposition decomposes a high-order tensor into a sum of  $R$  rank-one tensors, which can be expressed by the outer product of  $N$  vectors [61]. In this study, tensor CP decomposition is applied to analyze the spatiotemporal characteristics of travel patterns and explore the influence of several variables on travel patterns.

A third-order tensor  $X \in R^{I \times J \times K}$  is used to explain the CP decomposition process;  $I, J, K$  are the dimensions of three orders. Its approximate decomposition can be defined as follows:

$$X \approx \sum_{r=1}^R u_r v_r w_r, \tag{2}$$

where  $R$  is a positive integer and  $u_r \in R^I, v_r \in R^J, w_r \in R^K, r = 1, 2, \dots, R$ . The symbol  $\circ$  means vector outer product, and  $u_r v_r w_r$  is a rank-one tensor. The factor matrix ( $U = [u_1, u_2, \dots, u_R], V = [v_1, v_2, \dots, v_R],$  and  $W = [w_1, w_2, \dots, w_R]$ ) refers to the combination of the vectors from the rank-one component.

The element of the tensor  $X$  can be expressed as follows:

$$x_{ijk} \approx \sum_{r=1}^R u_{ir} v_{jr} w_{kr}, \tag{3}$$

where  $i = 1, 2, \dots, I, j = 1, 2, \dots, J$  and  $k = 1, 2, \dots, K$ , and  $i, j,$  and  $k$  are used to denote the index of the dimension of three orders.

The columns of  $U, V,$  and  $W$  can be normalized to length one with the weights of lambda.

$$X \approx \sum_{r=1}^R \lambda_r u_r v_r w_r = [\Lambda; U, V, W]. \tag{4}$$

The three-dimensional tensor can be extended to an  $N$ -dimensional tensor and written as follows:

$$X \approx \sum_{r=1}^R \lambda_r a_r^{(1)} a_r^{(2)} \dots a_r^{(N)} = [\Lambda; A^{(1)}, A^{(2)}, \dots, A^{(N)}]. \tag{5}$$

Negative values are difficult to explain, so we apply non-negative CP (NCP) with alternating Poisson regression from the tensor toolbox in MATLAB to perform NCP decomposition, which adds non-negative constraints to the tensor decomposition process. There is no straightforward algorithm to determine the rank  $R$  of a specifically given tensor, so we iterate through  $R$  from 1 to find a suitable solution.

**3.3.3. Tensor Reconstruction.** The tensor model reflects the distribution of data in the existing samples. However, the original tensor  $X$  is very sparse, which shows the value of many elements is zero. High sparsity means that there is no corresponding vehicle travel record in the LPR data set, but in reality, it is still possible to have vehicle records at intersections. In order to explore the complete data value

distribution and analyze the degree of joint influence of multiple variables on the target, it is necessary to learn from existing samples and predict the overall data distribution of the model. Tensor reconstruction is considered to apply in this process. Reconstructing the decomposed results can predict and fill most original zero elements while retaining the distribution of the original sample data.

Taking the third-order tensor as an example, the process of tensor reconstruction for CP decomposition can be expressed as follows:

$$\bar{X} = \sum_{r=1}^R u_r v_r w_r. \quad (6)$$

This is a process of learning based on actual samples. Equation (6) can be interpreted as inferring the data distribution characteristics of the whole situation from the distribution of the known sample data. The result of tensor reconstruction reflects the degree of influence of multiple variables on the target. The value of the element in the new tensor represents the estimation of the travel characteristics of the vehicle under the condition of the combination of these variables.

#### 4. Spatiotemporal Characteristics Analysis for Travel Patterns

The trajectory of vehicle trips generally expresses a strong temporal and spatial distribution characteristic, which is significant to the development of transportation and urbanization. The enriched LPR data provides us the opportunity to deeply understand the characteristics of vehicle mobility based on the tensor model. In the experiment, the large-scale LPR data are used to explore the spatiotemporal characteristics of travel patterns. In order to analyze the traffic flow in four directions at the intersections, we extract 21 intersections in the Wuyi Square area in the Changsha city as an example, which is the center of the city. The intersections are shown in Figure 3.

The daily samples are divided into 24 periods at time dimension. Based on the statistical analysis, we accumulate the numbers of the trips that happened in a specific period of time and driving direction at intersections. Then, a fourth-order tensor  $M \in R^{21 \times 24 \times 4 \times 7}$  (intersections, periods of the day, driving direction, days of the week) can be constructed. The element of the tensor indicates the trip volume passes through the intersection in a specific direction and time period. We conduct CP decomposition to the constructed tensor. There is no straightforward algorithm to determine the rank  $R$  for a specifically given tensor, so we iterate the value of  $R$  from 1 to find a suitable solution. Through repeated experiments and analysis of tensor composition results, we find that when the number of travel mobility patterns  $R = 3$ , three modes including morning peak pattern, evening peak pattern, and night peak pattern can be extracted at the temporal dimension. Thus, the collective model is decomposed into three modes.

The temporal distribution characteristics of vehicle trips under three modes in the driving direction dimension are shown in Figure 4. It shows that mode-1 can recognize

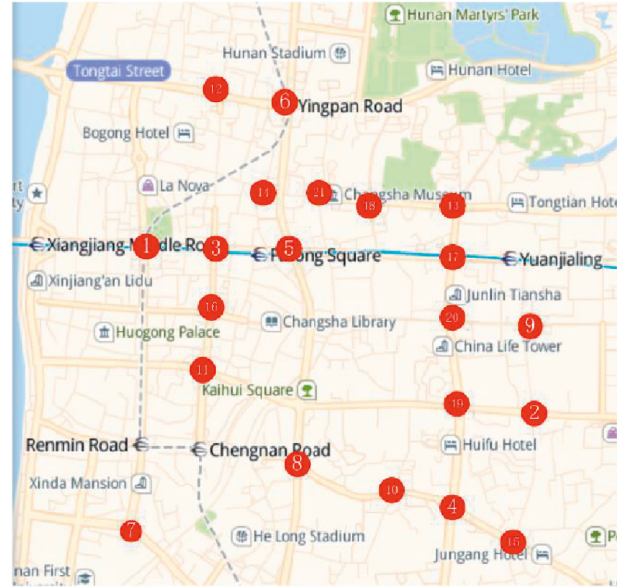


FIGURE 3: Intersections in Wuyi Square.

vehicle trip distribution of the morning peak hours in four driving directions, and the morning peak hours are concentrated at 7:00–9:00. Mode-2 indicates that there are obvious vehicle trips in the evening peak hours, and the evening peak hours in the four directions are all concentrated at 17:00–19:00. Mode-3 can identify vehicle trip distribution of the night peak hours in four driving directions, and the night peak hours are concentrated around 21:00.

The spatial distribution characteristics of vehicle trips corresponding to the temporal distribution of the three modes can be identified simultaneously based on the decomposition result. In order to have a comprehensive understanding of the spatial characteristics of vehicle trips, we calculate the proportion of vehicle trips in four driving directions under three modes in Figure 5.

We can explore that there are three characteristics of vehicle trips during peak hours in the spatial dimension. Firstly, the main traffic direction of some intersections is the same during the morning and evening peak hours, such as intersection 15. The main direction of the vehicle trips at some intersections during the morning and evening peak hours is the same in one direction and opposite in another direction, such as intersections 6, 7, and so on. The main driving directions of some intersections are all opposite during the morning and evening peak hours, such as 1, 2, 3, and so on. According to the different characteristics of vehicle trips during peak hours, some widely popular methods such as tidal lanes can effectively improve the efficiency of lane use. Secondly, the proportions of vehicle trips in four driving directions during the three peak hours are close at intersections 4, 8, and so on. Thirdly, there are some intersections (such as 1, 2, 5, 13, etc.) that show a great imbalance in the four driving directions of vehicle trips at certain peak hours. The intersections with this feature can

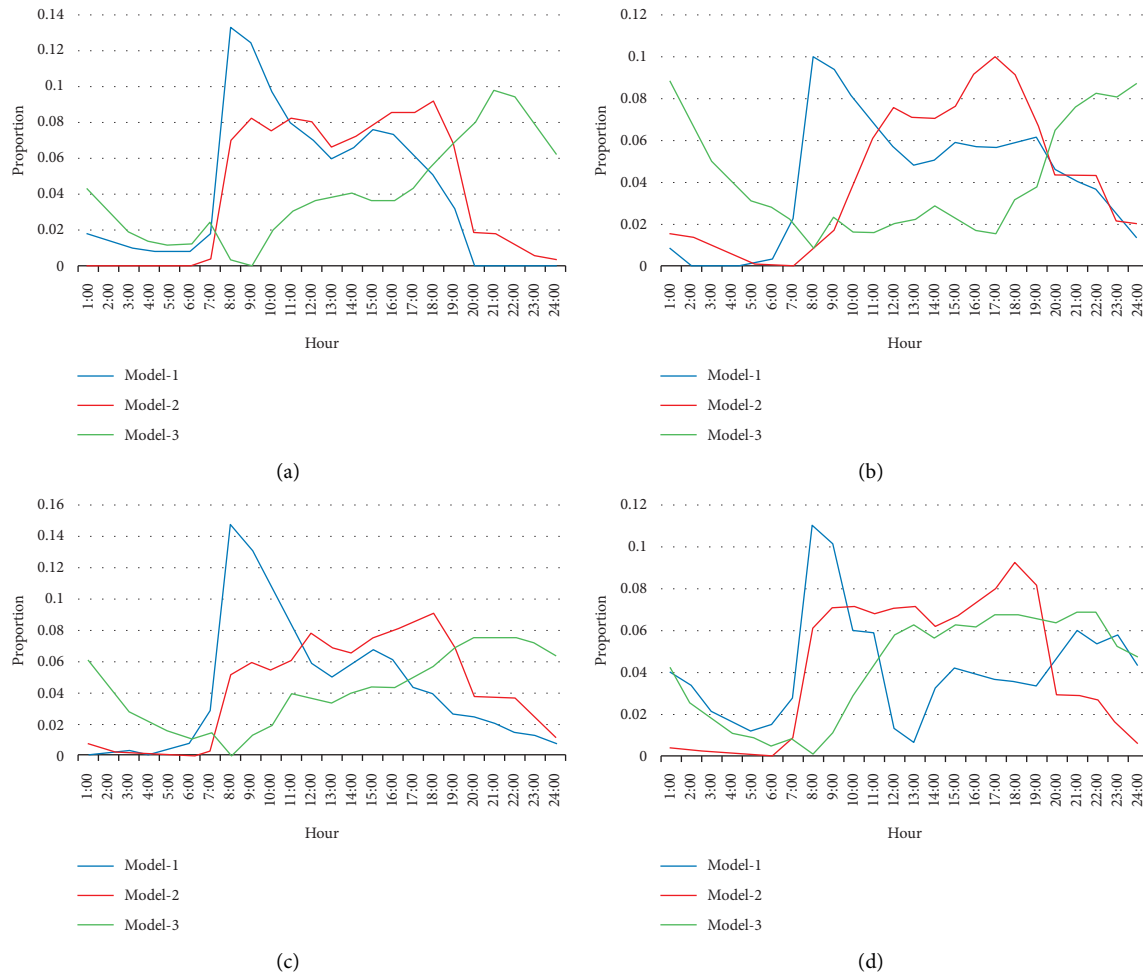


FIGURE 4: Three modes extracted at temporal dimension in four directions: (a) east to west, (b) west to east, (c) south to north, and (d) north to south.

conduct strategy by adjusting the signal timing scheme to the driving direction with more vehicle trips a long time to meet the traffic demand.

## 5. Influence of Variables on Travel Characteristics

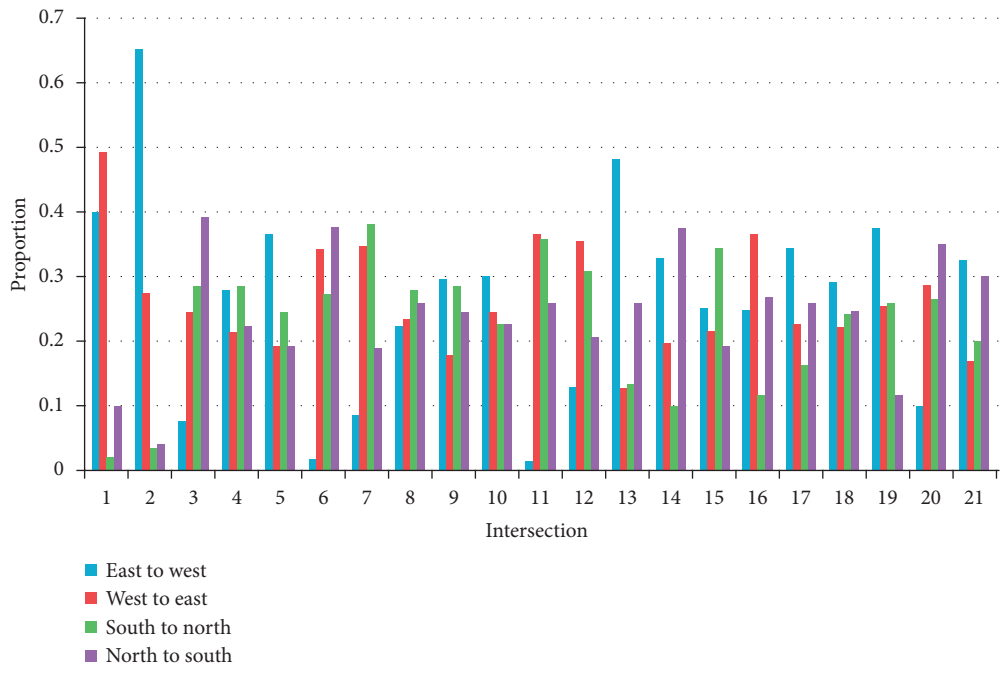
As an important branch of data mining technology, multivariable influence degree analysis has attracted more and more attention and application in various fields. The multivariable influence degree analysis used in this study is to explore the degree of influence of several variables on travel characteristics.

The original variables include O, D,  $PHF_{15}$ , NEF, POI, and Lane, which are introduced in Section 2. The O-point data represents the trip generation; the D-point data represents the trip attraction capacity.  $PHF_{15}$  reflects the proportion of vehicle trips at intersections in peak hours; NEF reflects the equilibrium degree of vehicle trips in each time period at intersections. The number of POI within a radius of 1 km at intersections represents the attraction of land use to the vehicle, and the number of

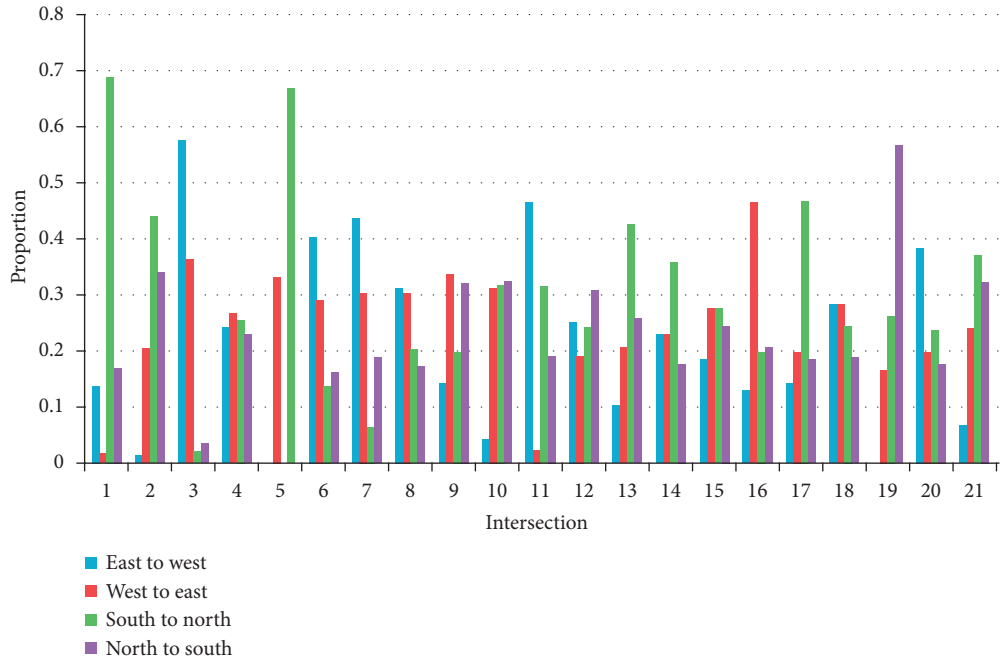
lanes at the intersections is also one of the important variables affecting the travel patterns. The problem in this study is to explore the degree of influence of several variables on travel characteristics. We abstract the problem by taking the six variables as independent variables and the target as dependent variables. Therefore, the problem is transformed into finding the value of the dependent variable in the case of the corresponding value combination of several independent variables.

Matrix is difficult to consider the joint influence of more than two variables at the same time. If these variables are expressed separately by vectors, the structural characteristics of the original sample cannot be preserved. So we propose the influence degree analysis method based on the tensor model, which has great advantages in retrieving and storing data. If all variables are directly modeled in tensor, the order of the model may be high, which will seriously affect the operation speed and accuracy. On the other hand, there may be a strong correlation between the variables. We treat these variables as independent components that may lead to the occurrence of multicollinearity between the variables, which will greatly reduce the accuracy of the operation results.



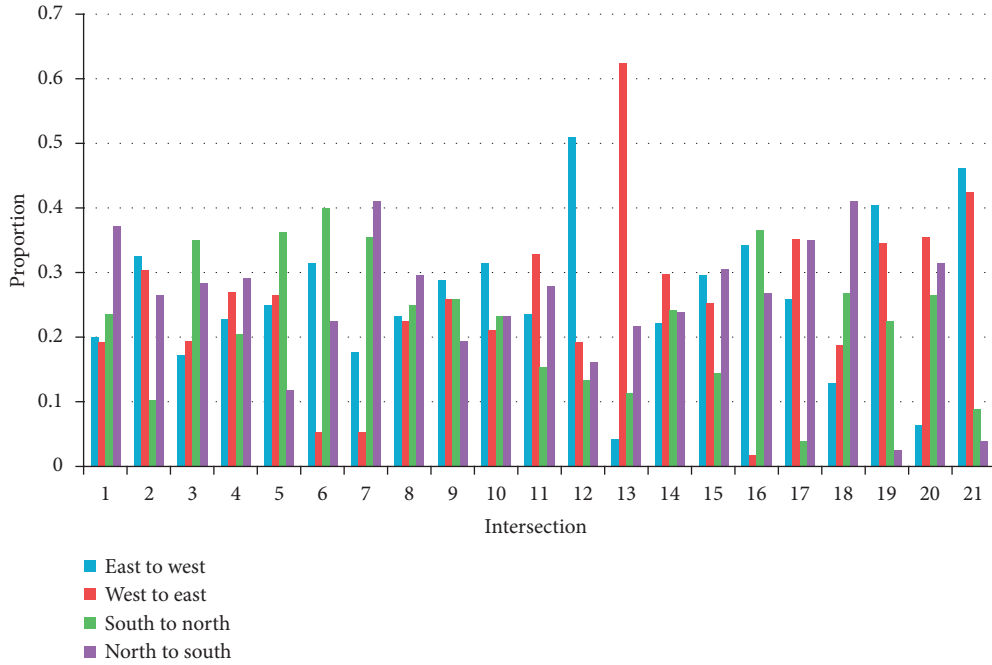


(a)



(b)

FIGURE 5: Continued.



(c)

FIGURE 5: The proportion of vehicle trips in four driving directions: (a) mode-1, (b) mode-2, and (c) mode-3.

Therefore, we perform clustering algorithms on all variables to construct feature variables for each cluster to achieve the purpose of dimensionality reduction and eliminate the complex collinearity among variables. Then, we utilize the tensor to model the feature variables generated from cluster analysis and PCA. Through tensor decomposition and reconstruction, the joint influence degree of each feature variable on the target can be analyzed. The method framework in this paper optimizes the data model and improves the accuracy of the analysis.

**5.1. Variables Clustering Results.** Clustering algorithms are an important technical branch in the field of data mining. There are two kinds of clustering methods that are very suitable for this study: partition clustering method and distance clustering method. The partition clustering algorithm divides the data set into  $k$  clusters by optimizing the evaluation function, which needs  $k$  as the input parameter. The distance clustering algorithm can divide the data set into several classes based on a distance threshold; the distance between classes is greater than the threshold; and the distance of classes is as small as possible. This kind of clustering method needs a certain distance threshold as an input parameter. In this paper,  $K$ -means clustering algorithm,  $K$ -medoids clustering algorithm, and maximum-minimum distance clustering algorithm are selected to process the six variables that affect vehicle travel characteristics.

Based on the roulette BET method, the  $k$  value in the  $K$ -means clustering algorithm is selected as 2 and the variables are divided into the following two categories:

- (1) Feature variable 1: O, D, PHF<sub>15</sub>, POI, and Lane

- (2) Feature variable 2: NEF

The clustering results of the  $K$ -medoids clustering algorithm are as follows:

- (1) Feature variable 1: O, D, Lane, and POI
- (2) Feature variable 2: PHF<sub>15</sub>
- (3) Feature variable 3: NEF

By setting an appropriate distance threshold ( $t = 0.8$ ), the clustering results of the variables in the maximum and minimum clustering algorithm are as follows:

- (1) Feature variable 1: O, D, PHF<sub>15</sub>, and Lane
- (2) Feature variable 2: NEF
- (3) Feature variable 3: POI

PCA is a common method to reduce dimension and eliminate multicollinearity of data. However, the practical significance of feature variables generated by PCA is not clear, which cannot meet the requirements of modeling in the subsequent analysis process of this study. Therefore, we utilize clustering algorithms to reduce dimension and apply the PCA method to generate feature variables by a linear combination of original related variables in the same cluster.

*K*-means clustering algorithm

$$X_1 = 0.4206O + 0.5315D + 0.4265PHF_{15} + 0.3141POI + 0.7166Lane. \quad (7)$$

*K*-medoids clustering algorithm

$$X_1 = 0.62O + 0.6293D + 0.412POI + 0.2231Lane. \quad (8)$$

Maximum and minimum clustering algorithm

$$X_1 = 0.5878O + 0.5999D + 0.4159PHF_{15} + 0.3488Lane. \quad (9)$$

*5.2. Analysis of Joint Influence Degree of Variables.* The clustering results mean that the variables within the cluster are related to each other, while the clusters are independent of each other. Therefore, the variables belonging to the same cluster can be linearly combined to obtain a new variable, which is called the feature variable of this cluster. Feature variables reflect the characteristics and attributes of the cluster and have clear practical significance. Joint influence degree analysis focuses on the joint influence of multivariable on the same goal. Using the influence distribution of existing sample data, we analyze and predict the joint influence of variables based on the tensor model.

We take the results of the  $K$ -medoids clustering algorithm as an example to explain the process of constructing the tensor model. The characteristics of vehicle trips are mainly affected by three feature variables. Therefore, we establish a third-order tensor for modeling and analysis. Let  $U$ ,  $V$ , and  $W$  be the value sets of these three feature variables and establish a three-order tensor  $X \in R^{I \times J \times K}$  to describe the relationship among the three feature variables, where  $I$ ,  $J$ , and  $K$  represent the dimensions of feature variables  $U$ ,  $V$ , and  $W$ , respectively. The value of the elements  $X_{ijk}$  is equal to the average value of the target data  $(i, j, k)$  mentioned in the previous description.

We combine three clustering algorithms with the tensor CP decomposition method and obtain the following specific methods:

- (1)  $K$ -means-CP:  $K$ -means algorithm is used for clustering, and CP decomposition is used for tensor decomposition
- (2)  $K$ -medoids-CP:  $K$ -medoids algorithm is used for clustering, and CP decomposition is used for tensor decomposition
- (3) D-CP: The maximum and minimum distance method is used for clustering, and CP decomposition is used for tensor decomposition

The reconstructed tensor is like a large table, and the travel characteristics of vehicles at the intersection can be found in the reconstructed tensor by index. Since the data of the reconstructed tensor cannot be represented intuitively, this paper uses the mean absolute percentage error (MAPE) to show the accuracy of the analysis of the degree of influence of multiple factors on residents' travel characteristics, as shown in Table 3. Mean absolute percentage error (MAPE) is shown in equation (10). MAPE is a relative value and is not influenced by actual and estimated values; thus, it can reflect the difference between actual and estimated values. A smaller MAPE represents the higher accuracy.

$$MAPE = \frac{\sum_1^n X_i - \bar{X}_i / X_i}{n} \times 100\%, \quad (10)$$

where  $n$  is the total number of samples and  $X_i$  and  $\bar{X}_i$  are the actual value of samples and their estimated value, respectively. The real values are the traffic travel volume under the

TABLE 3: The accuracy comparison of joint influence degree of variables.

Algorithm	MAPE (%)
D-CP	10.12
$K$ -medoids-CP	11.33
$K$ -means-CP	9.66

TABLE 4: The accuracy comparison in single variable influence degree analysis.

MAPE	O	D	Lane	PHF <sub>15</sub>	NEF	POI
$K$ -means-CP	10.10	8.04	12.78	10.87	7.24	11.78
D-CP	8.54	9.13	13.14	10.98	8.64	12.79
$K$ -medoids-CP	9.78	11.60	18.27	12.82	9.98	13.27

combination of feature variables obtained from cluster analysis and principal component analysis. The prediction values are the values predicted and filled by the tensor reconstruction model.

The tensor decomposition model can efficiently reproduce the higher-order interactions between different orders in multivariate data by using simple structures with relatively few parameters. The tensor model reflects the distribution of data in the existing samples. In order to explore the influence of variables on travel patterns completely, it is necessary to learn from existing samples and predict the overall data distribution of urban traffic networks. The tensor reconstruction model can predict and fill most original zero elements while retaining the distribution of the original sample data. Therefore, the method we proposed in this paper can greatly improve the accuracy of the joint influence degree of variables when analyzing travel characteristics.

*5.3. Analysis of Single Influence Degree of Variables.* The tensor model shows strong advantages in the analysis of the joint influence degree of variables. The reconstruction of tensor decomposition results can predict and fill original zero elements while retaining the distribution of the original sample data. The reconstructed tensor is just like a big table, the vehicle trips' characteristics at an intersection can be found in the tensor through the index. In order to explore the influence of a single variable on travel characteristics, we removed one variable in each model to calculate the MAPE. The accuracy comparison in single variable influence degree analysis is shown in Table 4.

It can be seen that the performance of  $K$ -means-CP is better than  $K$ -medoids-CP on the analysis of a single influence degree of variables. Compared to the  $K$ -means-CP model, D-CP produces obvious better accuracy in D and Lane and lower accuracy in O, PHF<sub>15</sub>, NEF, and POI. According to prior knowledge, the number of clusters can be roughly determined, so  $K$ -means is more appropriate to the purpose of this study. Therefore, the  $K$ -means-CP model in this paper can greatly improve the accuracy of the influence degree of variables on travel characteristics. The greater the influence of variables on travel characteristics, the lower the

accuracy of tensor reconstruction. We can find that Lane, O, and POI have a greater influence on travel characteristics, while PHF<sub>15</sub> has less influence on it.

## 6. Conclusions

Multivariable influence analysis method based on tensor decomposition and reconstruction can analyze joint influence degree of variables on vehicle trips' characteristics effectively and accurately. In this paper, we perform  $K$ -means,  $K$ -medoids, and the maximum and minimum distance clustering algorithms on several variables to construct feature variables for each cluster, which achieve the purpose of dimensionality reduction and eliminate the complex collinearity among variables. Modeling the feature variables into a tensor can ensure the integrity of the data and reflect the relationship between the variables. Tensor decomposition and reconstruction can identify important variables on travel characteristics and greatly improve the accuracy of the joint influence degree of variables when analyzing travel characteristics. The main conclusions of this paper can be summarized as follows:

- (1) The characteristic of vehicle trips can be decomposed into three modes in temporal dimension based on tensor decomposition. The morning peak pattern is concentrated at 7:00–9:00; the evening peak pattern is concentrated at 17:00–19:00; and the night peak pattern is concentrated around 21:00. We further explored the spatial characteristics of vehicle trips and find three characteristics of vehicle trips during peak hours.
- (2) The method framework we proposed in this paper achieves the purpose of dimensionality reduction, eliminates the complex collinearity among variables, and improves the accuracy of the joint influence degree of variables when analyzing travel characteristics.
- (3) We explored the influence of a single variable on travel characteristics, and Lane, O, and POI have a greater influence on travel patterns, while PHF<sub>15</sub> has less influence.

With the rapid development of urbanization, urban structures are becoming more complex. Rich traffic data provide opportunities to explore urban human travel characteristics and patterns. The research in this paper is dedicated to exploring travel patterns under temporal and spatial dimensions based on the LPR data and analyzing the influence of variables on travel characteristics. The methods provided in this paper are also suitable for the analysis of other cities with similar data sources. However, it should be noted that the travel patterns based on LPR data cannot represent the travel characteristic distribution of urban residents. Residents can choose different transportation modes, such as taking the metro for various trip purposes. Most existing data are only able to show city characteristics from a specific perspective. Further studies may expand the data source to include metro trips and shared electric vehicle

trips. This combination of diverse data can comprehensively and accurately describe travel mobility and urban structure in more dimensions. Meanwhile, the dimensions of the tensor constructed in this paper are relatively low. To take the advantage of the tensor model, the tensor will be constructed high-dimensionally to fully utilize the inner correlation. Besides, one potential question is how to choose an appropriate rank  $R$  in different situations. As a future research topic, we would like to propose an adaptive rank calculation method to figure out rank selection issues.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was funded in part by the Shandong Provincial Department of Transportation Technology Project (No. 2021B68), the National Natural Science Foundation of Shandong Province (No. ZR202103040494), the Humanities and Social Sciences Foundation of the Ministry of Education (No. 21YJCZH147), and the Innovation-Driven Project of Central South University (No. 2020CX041).

## References

- [1] S. Razzaq, F. Riaz, T. Mehmood, and N. I. Ratyal, "Multi-factors based road accident prevention system," in *Proceedings of the 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, IEEE, Quetta, Pakistan, April 2016.
- [2] G. Zhong, T. Yin, J. Zhang, S. He, and B. Ran, "Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data," *Transportation*, vol. 46, 2018.
- [3] L. Ni, X. C. Wang, and X. M. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.
- [4] Y. Wang, G. H. de Almeida Correia, B. van Arem, and H. J. P. Timmermans, "Understanding travellers' preferences for different types of trip destination based on mobile internet usage data," *Transportation research, Part C. Emerging technologies*, vol. 90, no. May, pp. 247–259, 2018.
- [5] H. Chen, C. Yang, and X. Xu, "Clustering vehicle temporal and spatial travel behavior using license plate recognition data," *Journal of Advanced Transportation*, vol. 2017, Article ID 1738085, 14 pages, 2017.
- [6] J. Tang and J. Zeng, "Spatiotemporal gated graph attention network for urban traffic flow prediction based on license plate recognition data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, pp. 3–23, 2021.
- [7] H. Wang, H. Huang, X. Ni, and W. Zeng, "Revealing spatial-temporal characteristics and patterns of urban travel: a large-scale Analysis and visualization study with taxi GPS data,"

- ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 257, 2019.
- [8] J. Tang, J. Liang, F. Liu, J. Hao, and Y. Wang, "Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network," *Transportation Research Part C: Emerging Technologies*, vol. 124, no. 102951, pp. 1–8, 2021.
- [9] A. Nasri, L. Zhang, J. Fan et al., "Advanced vehicle miles traveled estimation methods for non-federal aid system roadways using gps vehicle trajectory data and statistical power analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 11, pp. 296–308, 2019.
- [10] F. Mao, J. I. Minhe, and T. Liu, "Mining spatiotemporal patterns of urban dwellers from taxi trajectory data," *Frontiers of Earth Science*, vol. 10, no. 2, 2016.
- [11] L. Yu, C. Kang, G. Song, X. Yu, and T. Yuan, "Understanding intra-urban trip patterns from taxi trajectory data," *Journal of Geographical Systems*, vol. 14, no. 4, pp. 1–21, 2012.
- [12] J. Tang, X. Wang, F. Zong, and Z. Hu, "Uncovering spatio-temporal travel patterns using a tensor-based model from metro smart card data in shenzhen, China," *Sustainability*, vol. 12, 2020.
- [13] J. Tang, J. Liang, S. Zhang, H. Huang, and F. Liu, "Inferring driving trajectories based on probabilistic model from large scale taxi gps data," *Physica A: Statistical Mechanics and Its Applications*, vol. 506, pp. 566–577, 2018.
- [14] X. Zhan, R. Li, and S. V. Ukkusuri, "Link-based traffic state estimation and prediction for arterial networks using license-plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 117, Article ID 102660, 2020.
- [15] E. Naserian, X. Wang, K. Dahal, Z. Wang, and Z. Wang, "Personalized location prediction for group travellers from spatial-temporal trajectories," *Future Generation Computer Systems*, vol. 83, no. JUN, pp. 278–292, 2018.
- [16] Y. Yang, K. He, Y. p. Wang, Z. z. Yuan, Y. h. Yin, and M. z. Guo, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning methods," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, Article ID 127083, 2022.
- [17] M. Morzy, "Prediction of moving object location based on frequent trajectories," *Computer and Information Sciences – ISICIS 2006. ISICIS 2006*, vol. 4263, pp. 583–592, Springer, Berlin, Heidelberg, 2006.
- [18] J. J. Ying, W. C. Lee, and V. S. Tseng, "Mining geographic-temporal-semantic patterns in trajectories for location prediction," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 1, 2013.
- [19] T. H. Gim, "An analysis of the relationship between land use and weekend travel: focusing on the internal capture of trips," *Sustainability*, vol. 10, no. 2, 2018.
- [20] T. H. T. Gim and T. H. Tommy, "The relationship between land use and automobile travel utility: a Multiple Indicators Multiple Causes approach," *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 188–204, 2015.
- [21] T. H. T. Gim, "The relationships between land use measures and travel behavior: a meta-analytic approach," *Transportation Planning and Technology*, vol. 36, no. 5, pp. 413–434, 2013.
- [22] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartography and Geographic Information Science*, vol. 43, no. 2, pp. 103–114, 2015.
- [23] F. Zong, H. C. Qi, M. Tang, J. Y. Lyu, and P. Yu, "Identifying daily travel pattern-trip purpose using gpg data," *Jilin DaxueXuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)*, vol. 48, no. 5, pp. 1374–1379, 2018.
- [24] K. J. Button, *Travel Behavior and Travel Demand*, 2013.
- [25] S. Wang, D. Yu, M. P. Kwan, L. Zheng, H. Miao, and Y. Li, "The impacts of road network density on motor vehicle travel: an empirical study of Chinese cities based on network theory," *Transportation Research Part A: Policy and Practice*, vol. 132, 2020.
- [26] G. D. Han, Y. Wang, W. Zhao, L. Y. Xiao, C. L. Zhao, and G. Xu, "Simulation analysis of suv-pedestrian accident based on multi-factor influence," *Procedia Computer Science*, vol. 154, pp. 657–662, 2019.
- [27] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering and Management Journal*, vol. 16, no. 10, pp. 2413–2423, 2017.
- [28] K. Park, R. Ewing, B. C. Scheer, and G. Tian, "The impacts of built environment characteristics of rail station areas on household travel behavior," *Cities*, vol. 74, pp. 277–283, 2018.
- [29] J. Li, Y. Chen, Z. Li, and Z. Liu, "Quantitative analysis of the impact factors of conventional energy carbon emissions in Kazakhstan based on lmdi decomposition and stirpat model," *Journal of Geographical Sciences*, vol. 28, no. 7, pp. 1001–1019, 2018.
- [30] Q. H. Truong, T. N. An, Q. A. Trinh, T. Ngoc, and L. Hens, "Hierarchical variance analysis: a quantitative approach for relevant factor exploration and confirmation of perceived tourism impacts," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, 2020.
- [31] N. Arora, S. Prashar, C. Parsad, and S. V. Tata, "Influence of celebrity factors, consumer attitude and involvement on shoppers' purchase intention using hierarchical regression," *Decision*, vol. 46, no. 3, pp. 179–195, 2019.
- [32] X. Wang and M. Feng, "Freeway single and multi-vehicle crash safety analysis: influencing factors and hotspots," *Accident Analysis & Prevention*, vol. 132, Article ID 105268, 2019.
- [33] J. Yu Gong, W. Yu, T. Qu, Y. Liu, X. Fu, and Y. Xu, "Multi-factor Analysis assisting T-image design for tactile cognition," *Journal of Computer-Aided Design & Computer Graphics*, vol. 30, no. 8, pp. 1438–1445, 2018.
- [34] S. Li Emotion Model Based on Multi Factor Matrix, 2016.
- [35] P. Kenny, P. Dumouchel, G. Boulianne, and P. Ouellet, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Transactions on Audio Speech and Language Processing*, vol. 15, no. 4, pp. 1435–1447, 2007.
- [36] K. Khine and T. Nyunt, *Predictive Big Data Analytics Using Multiple Linear Regression Model*, Springer, Singapore, 2018.
- [37] W. Bian and D. Tao, "Max-min distance analysis by using sequential sdp relaxation for dimension reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 1037–1050, 2011.
- [38] S. Liu, N. Wang, and X. Wang, "Multi-factor influenced integrated driving behavior clustering analysis evaluation model," in *Proceedings of the 2019 International Conference on Computer, Information and Telecommunication Systems (CITS)*, Beijing, China, August 2019.
- [39] T. G. Kolda, B. W. Tamara Bader, and B. W. Brett, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

- [40] A. Cichocki, D. Mandic, L. De Lathauwer et al., "Tensor decompositions for signal processing applications from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.
- [41] M. Sørensen and L. D. Lathauwer, "Tensor decompositions with Vandermonde factor and applications in signal processing," in *Proceedings of the 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, IEEE, Pacific Grove, CA, USA, November 2012.
- [42] X. Zhang, X. Yuan, and L. Carin, "Nonlocal low-rank tensor factor analysis for image restoration," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, UT, USA, June 2018.
- [43] Y. Liu, Y. Yang, Q. Zhao, and S. Aeron, "Editorial to the special issue on tensor image processing," *Signal Processing: Image Communication*, vol. 73, pp. 1-2, 2019.
- [44] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 13–22, Zurich Switzerland, September 2013.
- [45] J. Wang, G. Fei, C. Peng, L. Chao, and X. Zhang, *Discovering Urban Spatio-Temporal Structure from Time-Evolving Traffic Networks*, Springer, Cham, 2014.
- [46] J. Bao, C. Xu, P. Liu, and W. Wang, "Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests," *Networks and Spatial Economics*, vol. 17, no. 4, pp. 1231–1253, 2017.
- [47] J. j. Tang, J. Hu, Y. w. Wang, H. l. Huang, and Y. h. Wang, "Estimating hotspots using a Gaussian mixture model from large-scale taxi GPS trace data," *Transportation Safety and Environment*, vol. 1, no. 2, pp. 145–153, 2019.
- [48] M. Saberi, H. S. Mahmassani, D. Brockmann, and A. Hosseini, "A complex network perspective for characterizing travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks," *Transportation*, 2016.
- [49] J. O. Poltavskaya and O. A. Lebedeva, "Assessment of the impact of the characteristics of the road network on the duration of the route," *Journal of Physics: Conference Series*, vol. 1680, no. 1, Article ID 012043, 6 pages, 2020.
- [50] S. Bassan, "Modeling of peak hour factor on highways and arterials," *KSCE Journal of Civil Engineering*, vol. 17, no. 1, pp. 224–232, 2013.
- [51] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, 2003.
- [52] Y. Kumar and G. Sahoo, "A new initialization method to originate initial cluster centers for k-means algorithm," *International Journal of Advanced Science and Technology*, vol. 62, pp. 43–54, 2014.
- [53] S. F. Hussain and M. Haris, "A k-means based co-clustering (kcc) algorithm for sparse, high dimensional data," *Expert Systems with Applications*, vol. 118, no. MAR, pp. 20–34, 2019.
- [54] J. Deng, J. Guo, and Y. Wang, "A novel k-medoids clustering recommendation algorithm based on probability distribution for collaborative filtering," *Knowledge-Based Systems*, vol. 175, no. JUL.1, pp. 96–106, 2019.
- [55] A. Bostan and É. Schost, *A Simple and Fast Algorithm For*, 2009.
- [56] K. Wang, B. Cheng, and T. Long, "An improved SIFT feature matching algorithm based on maximizing minimum distance cluster," *International Conference on Computer Science & Information Technology*, 2011.
- [57] J. A. Rose and C. C. Seldev, "Performance analysis of minimum distance classifier and clustering algorithm in cbir," *Applied Mechanics and Materials*, vol. 626, pp. 14–19, 2014.
- [58] L. E. Richards and I. T. Jolliffe, "Principal component analysis," *Journal of Marketing Research*, vol. 25, no. 4, p. 410, 1988.
- [59] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [60] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, Berlin, Heidelberg, Germany, 2005.
- [61] T. G. Kolda and B. W. Brett, "Tensor decompositions and applications," *SIAM Review*, vol. 51, 2009.