Research Article

Congestion Prediction Based on Dissipative Structure Theory: A Case Study of Chengdu, China

Xiaoke Sun, Hong Chen, Yahao Wen, Zhizhen Liu, and Hengrui Chen

1College of Transportation Engineering, Chang’an University, Xi’an, China
2Guangdong Communication Planning & Design Institute Co., Ltd., Guangzhou, China

Correspondence should be addressed to Hong Chen; hongchen8233@163.com

Received 3 December 2020; Revised 29 December 2020; Accepted 5 January 2021; Published 19 January 2021

1. Introduction

Due to the rapid increase in urban expansion, traffic demand has also increased rapidly in recent years. Under such circumstances, traffic congestion has become a common problem in large cities, resulting in increased pollution emissions, traffic time, and casualties and slowing down the economy [1]. Traffic congestion is caused by various factors, such as the backwardness of transportation infrastructure, public transportation inability to meet general travel needs, and improper traffic control. Accurate traffic jam prediction can improve travelers’ satisfaction with traffic services and reduce related travel costs. There have been many methods used to predict traffic congestion in the current research process, including qualitative research, quantitative prediction, and various statistical techniques. However, most of the methods previously studied by scholars are still based on traditional transportation research methods. In the spatio-temporal analysis, some scholars used GPS data to compare the temporal and spatial patterns of taxi travel in Shanghai and New York City and established a regression model to study the relationship between urban land use, a permanent population, employment, and car ownership [2]. There are also studies using geographically weighted regression to model the spatial heterogeneity of taxi passenger capacity and visualize parameter estimates’ spatial distribution [3]. A visual analysis system of urban functions based on time-space taxi travel is proposed [4]. Some scholars have researched the interactive relationship between the taxi travel trajectory network’s topological structure and the spatial differentiation, revealing the spatial characteristics and movement laws of urban residents’ activities and interaction between the spatial layout of urban functions and residents’ activities. Provide a reference for optimizing the taxi transportation network and taxi operation management [5].

In terms of travel patterns, some scholars have studied how the travel pattern variables extracted from large-scale taxi GPS data can lead to the collapse of spatial agglomeration in urban areas and proposed a data-driven modelling method based on potential Dirichlet allocation of 50 travel modes [6]. Using the massive dataset of Didi Travel, including Didi Express and Didi Taxi services, some scholars analyzed the fluctuations in the number of orders in different urban areas after implementing travel restrictions in Shanghai in 2016 [7]. Some scholars analyzed Beijing taxis’ GPS trajectory data and found that taxi travel...
patterns have similar characteristics to individual travel patterns [8]. Massive car-hailing data have become a popular source for analyzing traffic operation and road congestion status [9]. Some studies have used clustering methods to detect different passenger distribution patterns in subways and taxis, examine the difference between passenger distribution patterns and cluster spatial distribution, and perform a two-step classification analysis to determine the factors affecting passenger patterns [10]. Some scholars have proposed a new model based on the hidden Markov model and contrast to define the traffic state during peak time in two-dimensional space. This model uses average speed and contrast to capture traffic patterns [11]. Other studies have proposed an expert system that detects traffic jams and accidents from real-time GPS data collected by GPS trackers or driver’s smartphones. The system will assign a traffic state to each section of the map according to the vehicle [12].

In terms of research methods, some scholars decompose the datasets of subway station entry and exit, weekdays, and weekends to obtain principal components and feature vectors [13]. Some studies have proposed a three-stage framework to explore the congestion correlation between road segments from multiple real-world data and found that the traffic congestion correlation has obvious directionality and transmission [14]. Some scholars have proposed a new method based on the entropy maximization theory, which uses the large-scale taxi GPS trajectory to model the OD distribution of Harbin city to verify the OD distribution of the taxi GPS data in the urban system [15].

In terms of congestion prediction, some scholars have proposed a method to detect traffic congestion from the taxi’s GPS trajectory at the turning level. Based on the analysis of GPS trajectory characteristics and identification of active trajectory segments, three congestion trajectories of different intensities are detected [16]. Some scholars have proposed a probabilistic model for predicting driving journey paths based on hidden Markov models. The prediction results show that this method is an accurate and feasible potential method [17]. Some scholars have proposed mining mixed temporal association rules to predict traffic congestion, apply the Dbscan algorithm to find the traffic environment, and generate qualified rules for predicting road network traffic congestion [18].

Nowadays, there is a new direction for studying traffic congestion. With the advent of the “big data” era, modern technology combined with traditional traffic management methods to effectively and timely obtain traffic demand information and identify road traffic conditions is an important idea to alleviate traffic congestion. With the rapid development of the mobile Internet, online car-hailing travels have been rapidly developed. It is also one of the supplementary travel modes of urban public transportation. A large amount of GPS data is generated in the daily operation process, attracting many scholars to develop big traffic data: analysis, mining, and application research. The famous physicist Prigogine introduced open systems to the second law of thermodynamics in 1969 and developed it to establish the dissipative structure theory. Dissipative structure theory explains how an open system changes from disorder to order, an extension of entropy theory. At present, dissipative structure theory has significantly impacted many fields of natural sciences and social sciences, such as physics, astronomy, biology, economics, and philosophy [19]. However, research of dissipative structure theory on the diffusion of pollutants is rarely studied.

This research uses dissipative structure and entropy theory, combined with the hidden Markov model, based on previous studies. Based on analyzing the temporal and spatial distribution of Chengdu, it predicts congestion in a part of Chengdu’s downtown area. It carries out verification, and the result shows that the prediction result is much better.

The rest of this article is organized as follows: after the Introduction, Section 2 introduces materials and methods, including data and the theory; Section 3 presents result analysis and discussion; Section 4 presents conclusion.

2. Materials and Methods

2.1. Data

2.1.1. Didi GPS Track Data. The data source used in this article is the Didi Gaia Data Open Platform, the data location is Chengdu City, the data time is November 2016, a total of 30 days, and the data scope is the trajectory data of the second ring road in Chengdu, so the research scope is divided. It is designated as the five central districts of Chengdu: Jinniu District, Chenghua District, Jinjiang District, Wuhou District, and Qingyang District. The data are stored in CSV file format and divided into GPS track data and travel OD data (recording the OD point information of each track data). The average order data from Sunday to Thursday is about 220,000 pieces a day and from Friday to Saturday is about 250,000 pieces a day. GPS track data record the operating status of each vehicle and records, and the interval is 3 s. The amount of data per day is about 3 GB, and the amount of track point data generated in one day is about 36 million. The formats of GPS track data and order data in the original data are shown in Table 1.

2.1.2. Road Network Data and POI Data. The road network data used are downloaded from the BIGEMAP mapper. The road network includes highways and urban roads (express roads, main roads, secondary roads, and branch roads). The coordinate system used WGS84 coordinates.

The POI data source is from BIGEMAP mapper, and the coordinate system adopted WGS84 coordinates. POI data include hotels, restaurants, roads, real estate communities, companies, enterprises, shopping, transportation facilities, finance, tourist attractions, car services, commercial buildings, life services, leisure and entertainment, medical care, and government agencies. POI data information includes longitude, latitude, address, and name.

2.2. Principal Component Analysis. Generally, in the study of multivariate research, we tend to use as few variables as possible to get more information. When the number of variables increases, the computational complexity will
increase geometrically. Similarly, in studying road traffic status, it is unnecessary to use all variables to analyze the traffic status. At the same time, it is impossible to check all variables. Therefore, it is essential to filter the variables. Thus, the principal component of the statistical method is used. Analysis can reduce the dimensionality of the data and transform a group of potentially correlated variables into linearly uncorrelated variables through orthogonal transformation. The changed group of variables is called the principal component. It can be used to analyze the variables to determine the more significant variables on traffic status and traffic congestion prediction, a standard method for dimensionality reduction. Typical steps for dimensionality reduction include the following: (1) standardize raw data; (2) calculate the covariance matrix and its eigenvalues and eigenvectors; (3) sort eigenvalues; and (4) keep the eigenvectors corresponding to the first $N$ eigenvalues and construct them into the new space.

### 2.3. Hidden Markov Model

Hidden Markov model (HMM) is a dynamic Bayesian network with a simple structure, a directed graph model. In this system, the state of the system at the next moment is determined only by the current state and does not depend on any previous state. It is a directed graph model, which is widely used in many fields. As shown in Figure 1, there are two types of variables in the hidden Markov model: hidden variables ($Y_1$, $Y_2$, $Y_3$, \ldots, $Y_n$) and observed variables ($X_1$, $X_2$, $X_3$, \ldots, $X_n$). $Y_n$ and $X_n$ represent the system status and observations of the $n$th time.

### 2.4. Entropy Theory and Dissipative Structure Theory

Entropy was used to describe the second law of thermodynamics. As a state parameter, entropy represents the uniformity of any kind of energy distribution in space.

$$S = K \times \ln N.$$  \hspace{1cm} (1)

The entropy $S$ of the material system is equal to the Bosch coefficient $K$ product and the number $N$ of states’ logarithm. The entropy value of the system directly reflects the degree of uniformity of the state. The smaller the entropy value of the system is, the more orderly and unevenly the state is. The more significant the system’s entropy value is, the more disordered and uniform it is. The system always tries to spontaneously change from a state with a small entropy value to a state with a significant entropy value (from order to disorder).

The entropy principle mainly studies the ideal isolated system, while the dissipative structure theory extends the system to open systems. The generalized dissipative structure can refer to a series of open systems far from equilibrium, physical, chemical, socioeconomic, and biological systems, which focuses on explaining how the open system moves from disorder to order [20].

### 3. Result Analysis and Discussion

#### 3.1. Spatial and Temporal Distribution of Congestion in Chengdu

#### 3.1.1. Time Distribution Characteristics

By counting the daily travel frequency within a month, as shown in Figure 2, it can be seen that the daily travel frequency is approximately a one-week cycle, with the highest travel frequency on Friday and Saturday. After the workweek, on Friday afternoon and Saturday, the frequency of people going out for entertainment and leisure will increase.

To understand the travel frequency at different times of the day, take an hour as a unit to count the frequency of pickup and drop-off points at various times during working days and nonworking days, as shown in Figures 3 and 4. As can be seen from the distribution diagram of the time of getting on and off the bus, there is a morning peak (8:00–10:00), noon peak (12:00–14:00), and evening peak (17:00–19:00) during work, while on rest days, there are only noon peaks and evening peaks. On Friday, the morning, afternoon, and evening peaks are similar to Monday to Thursday. Still, in the evening peak until midnight, the number of trips on Friday is higher than that on Monday to Thursday. There are some similarities and differences in travel time’s frequency distribution on Saturdays and Sundays on weekends. The difference is manifested in the apparent peak periods at noon (13:00–14:00) and evening
Figure 2: Travel frequency in November.

Figure 3: Distribution of boarding time.

Figure 4: Distribution of drop-off time.
(17:00–18:00). Besides, since the next day is a working day, people are more inclined to end the day’s itinerary earlier on Sunday evening. The overall travel frequency during the evening rush hour on Sunday is less than that on Saturday.

3.1.2. Spatial Distribution Characteristics Based on the Kernel Density Algorithm. The grid surface is generated based on the kernel density algorithm to establish the hot spot detection model. The parameter selection of the kernel density algorithm affects the clustering effect and effectiveness of the model. The larger the research scale and the larger the bandwidth, the smoother the result presented by the heat map; the smaller the research scale and the smaller the bandwidth, the more detailed the heat map’s effect in Table 2. Using multiple parameter selection and comparative analysis, selecting different parameters to produce clustering effect and hot spot detection, the hot spot area’s location can be preliminarily judged. Select the date of November 11 (Friday) as the representative days, visualize the data, analyze the trajectory data’s spatial characteristics, and display it in a heat map. The heat map in different parameter combinations is shown in Figure 5.

It can be known from the heat map that the demand for travel in the central area of the city is relatively stable. As the city center’s various supporting facilities are relatively complete, the population is much more. Hotels, restaurants, commercial buildings, shopping centers, residential areas, and other buildings are densely distributed. These places can generate a lot of transportation demand.

3.2. Analysis of Didi GPS Data Features. Due to Didi’s ride-hailing rules, some data need to be cleared. In collecting trajectory data, the sampling interval is 3 s, and the sampling frequency is relatively high. More redundant data are generated, which take up massive storage space and higher computer performance requirements, which significantly increases the processing runtime of the calculation. Therefore, it is necessary to compress the taxi trajectory data [21].

Calculate the length of the trajectory according to the compressed trajectory. Under the premise of ensuring accuracy, the travel distance statistics of 140,000 passenger trajectories extracted are made. The travel distance (a) and time (b) distribution are shown in Figure 6. The travel distance is mostly concentrated within 3–10 km, and this travel distance segment accounts for about 75% of the total travel ratio. The travel time is mainly focused on 6–26 minutes, accounting for about 75% of the total travel ratio.

After the invalid trajectory data are further eliminated and filtered, the travel speed diagram in the main urban area within one day by time intervals is obtained, as shown in Figure 7.

It can be seen from Figure 7 that, by calculating the average speed of the vehicles in each period, it can be found that the speed of the vehicles is slower in the morning peak and evening peak on working days. At night, the speed of the vehicle is faster.

Based on calculating the average travel speed of the entire road network by period, a single trajectory’s speed is calculated. The speed characteristics of the trajectory are further feature mined and analyzed based on different road sections. For a single trajectory, calculate the travel speed according to different sampling frequencies, and calculate the speed of different road sections in one travel. Take a trajectory on November 7 as an example, combined with the trajectory data points matched on the map, as shown in Figure 8. Vehicles drive from Fenglin Road near Zhongfeng Hongfengling and then pass through Shengdeng Road, Jianshe South Road, Hongguang Road, Xinhong North Branch Road, and Xiaolongqiao Road to reach the destination. Figure 9 shows the travel speed graph with sampling intervals of 3 s, 6 s, 15 s, and 30 s. When the sampling interval is 3 s, and the sampling frequency is dense, the speed-time image obtained can be regarded as an instantaneous speed graph. It can be seen that the rapid speed fluctuates wildly. It cannot reflect the road operation well. When the sampling interval is 15 s or 30 s, the effect of the displayed map is better.

After map matching, combining the road segment trajectory map, divide the trajectory into seven road segments, as shown in Figure 8, and calculate the seven road segments’ travel speed, respectively.

After adjusting and improving the instantaneous speed map, the road section speed map divided by road section is obtained, as shown in Figure 10. It can be seen from Figure 10 that the vehicle speed is low at the beginning of the start, and the vehicle is driving on road sections 5 and 6 with faster speed, and the overall driving speed is relatively smooth and unobstructed.

3.3. Analysis of Influencing Factors of Congestion in Chengdu

3.3.1. Extraction of Factors Affecting Traffic State. In the BIGEMAP map downloader, the downloaded POI data have the following 15 categories. The POI categories are readjusted and divided into six categories, namely, commercial, residential, office, transportation, leisure class, and life class. The classification situation and the proportion of each type are shown in Figure 11.

Due to the dense flow of people and the high demand for online car-hailing in the city’s central area, various regions have different degrees of influence on the pickup point. It is of considerable significance to analyze the possible impact of different POI types on the road state. Calculate the distance between the pickup point and the different POI types nearby, find the connection between the pickup point and the POI type, and further determine the impact on the road traffic state and traffic congestion the follow-up work [22].

Based on the principal component analysis method, this paper calculates the distance from the OD points of the trajectory data to various POI points. The Euclidean distance from each OD point to different POI points is calculated to facilitate statistics and classification.

After categorizing the POI data, combine the trajectory data’s spatial characteristics and the distance from the OD point
Table 2: Influence of different parameters on the clustering effect.

<table>
<thead>
<tr>
<th>Group</th>
<th>Order</th>
<th>Bandwidth (unit: m)</th>
<th>Pixel size</th>
<th>Number of hotspots (unit: number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>108.8</td>
<td>10</td>
<td>2104</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>317.0</td>
<td>10</td>
<td>623</td>
</tr>
</tbody>
</table>

Figure 5: Heat diagram in different parameter combinations.

Figure 6: Distance (a) and time (b) distribution.
Figure 7: Daily hourly average speed.

Figure 8: Divide the trajectory map based on road sections.

Figure 9: Continued.
Figure 9: Speed time of different time intervals.

Figure 10: Travel speed based on road segment division.

Figure 11: Percentage of POI classification.
of the order data to various POIs, perform a correlation test, and express it in a matrix.

From Table 3, the commercial and residential categories have a more significant impact on the frequency of getting on and off the bus.

In the city’s central area, travel demand is relatively strong near commercial buildings, shopping centers, and large residential areas. The distance between OD points and the commercial regions and residences is closer than other types. The relationship between OD points and commercial and residential areas is relatively close, so that it can be considered an influencing factor.

3.3.2. Construction of an Influencing Factor Set. This paper selects the area around the east section of the second ring road in Chengdu as the research area and divides the road sections and time periods. There are 4 road types in the divided road sections, and the time period and the road sections are both the influencing factors. The average travel time and the average speed of the road section are the main manifestations of whether the road is congested, and the road traffic state is the direct manifestation of whether the road is congested. Therefore, both are used as influencing factors. This paper uses principal component analysis to extract the types of POI that have a greater impact on the boarding point: the distance from the boarding point to the subway station, commercial area, and residential area, which are used as an influencing factor. The set of influencing factors selected in this paper is moderate. Too few would lead to unsatisfactory prediction results, and too many would easily lead to “overfitting,” and the error of the model will increase. Therefore, the influencing factors F1–F8 are finally determined.

The types and descriptions of each factor are as follows:

F1: type of road. The type of road has a significant influence on the vehicle’s speed and determines the upper limit of Table 4.

The speed is influencing the road traffic conditions and congestion. The corresponding relationship of road types is shown in Table 4. The road segments are classified and numbered according to the relationship table, which is represented by $S_{aa}$.

F2: average travel time of the road section (s).

The road segment’s average travel time is obtained according to the order data and is counted in seconds.

F3: average travel speed of the road section (km/h).

Based on the road segment division and speed calculation in front of the article.

F4: period.

According to the trajectory data’s temporal and spatial distribution characteristics, boarding at different times in a day is counted. On weekdays, the frequency of boarding at 8–10 am is relatively high. Simultaneously, considering the regularity of commuting passenger flow, this period is selected as the counter, divided into shorter periods. Choose 10 minutes as the time segment, divided into 12 time segments, respectively, represented by $1–12 (a−1)$.

F5: road traffic status. According to the road section, it is divided into five statuses, which are unblocked, basically unblocked, slightly congested, moderately congested, and severely congested, represented by 1–5.

F6: the distance to the subway station (m).

The subway can efficiently carry many passengers, which affects the travel mode of nearby shuttle vehicles and nearby people. The Euclidean distance is used here by calculating the distance from the boarding point to the nearest subway station.

The distance of F6–F8 is obtained through the distance calculation tool in the ArcGIS software. The distance is divided into three categories: 0–500 m, 500–1000 m, and distances higher than 1000 m, represented by 1, 2, and 3, respectively.

F7: distance to the business district (m).

By calculating the distance from the pickup point to the nearest commercial area, the commercial area’s judgment is obtained by dividing POI data. The Euclidean distance is used here.

F8: distance to a residential area (m).

By calculating the Euclidean distance from the pickup point to the nearest residential area, the residential area’s determination is obtained through the POI data division. The Euclidean distance is used for calculation. By dividing the influencing factor set, the final main parameter table is shown in Table 5.

3.4. Congestion Prediction Based on the Dissipative Structure

3.4.1. Dissipative Structure Condition. According to the dissipative structure theory established by Prigokin, the following four necessary conditions must be met to produce a dissipative structure.

The system must be open. The system must exchange material, energy, and information with the outside world and obtain negative entropy from the outside world to offset the increase in its extraction. The system can evolve from disorder to order and from simple to complex.

The system is far from equilibrium. The open system under the combined action of internal and external factors may destroy the original structure and form a new orderly structure when it is far from the equilibrium state.

Nonlinear interaction: there is a nonlinear mechanism in the interaction between the subsystems that make up the system, prompting the emergence of new properties in the system, leading to the system’s complexity and diversity. When a linear system changes, it is often carried out gradually; when a nonlinear system changes, there are often qualitative transformations and jumps. When affected by the outside world, a linear system will respond progressively, while a nonlinear
system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system changes continuously and changes state over time, while a nonlinear system can maintain its stability for a long time. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely. A linear system is very complicated. Sometimes it ignores external signals, and sometimes it reacts fiercely.

There are substantial random fluctuations in the road traffic system of Chengdu. Since Chengdu is a tourist-oriented city, a large amount of traffic will be generated during holidays. Due to the impact of other uncontrollable factors such as traffic accidents, the traffic system will have substantial random fluctuations in traffic volume.

In Chengdu’s road system, a specific type of state in the congestion state can be used as a critical point 3: slightly congested), a certain threshold reached when the system leaves the equilibrium state. When the threshold is reached, self-organization can be used to relieve the congestion state. As a part of the road system, human intervention belongs to the road system’s self-organization phenomenon. Social self-organization will transform the system from a disordered state to an orderly state.

3.4.2. Entropy Analysis. According to the dissipative structure theory and entropy theory, the influencing factors are divided into the primary deterministic entropy value and the uncertainty entropy value, affecting the total entropy value of the road traffic system. The deterministic entropy value consists of the influence of road type, period, distance to the subway station, distance to commercial area, and distance to the residential area on the degree of congestion. The uncertainty entropy value is composed of the road section’s average travel time, the average travel speed of the road section, and the road traffic conditions on the degree of congestion. Among them, the deterministic entropy value is fixed. This study will discuss the surrounding facilities when the surrounding facilities are relatively stable, without considering the impact of environmental factors and construction facilities changes.

$$S_d = - \sum_i^d K_i P_i \ln P_i,$$

$$S_n = - \sum_i^n K_i P_i \ln P_i,$$

$$S_c = S_d + S_n,$$

where $S_d$ is the deterministic entropy value and $d$ represents the type of deterministic entropy factors. $S_n$ is the uncertainty entropy value and $n$ represents the type of uncertainty entropy factors. $S_c$ is the total entropy. The congestion state changes as the total entropy value changes. When the total entropy value reaches a specific amount, when the system leaves the equilibrium state to reach a certain threshold (3: slightly congested), the system changes simultaneously, from state 1 or 2 changes to 3 or other states. At this time, human intervention is required for traffic, which is the phenomenon of systematic self-organization.

3.4.3. Congestion Prediction. This paper used Python language scipy, scikit-learn, hmmlearn, and other libraries, combined with the dissipative structure theory and the hidden Markov model for prediction.

### Table 3: Correlation analysis based on POI data classification.

<table>
<thead>
<tr>
<th>Type of road</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban expressway</td>
<td>1</td>
</tr>
<tr>
<td>Urban arterial road</td>
<td>2</td>
</tr>
<tr>
<td>Urban secondary road</td>
<td>3</td>
</tr>
<tr>
<td>Branch road</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Abbreviated representation</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road type</td>
<td>$R_i$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Average travel time (s)</td>
<td>$T_{ij}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Average travel speed (km/h)</td>
<td>$V_{ij}$</td>
<td>—</td>
<td>70</td>
<td>—</td>
</tr>
<tr>
<td>Period</td>
<td>$T_n$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Traffic status</td>
<td>$S_{out}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Distance to the subway station (m)</td>
<td>$D_{me}$</td>
<td>8.46</td>
<td>746.9</td>
<td>2875.41</td>
</tr>
<tr>
<td>Distance to the commercial area (m)</td>
<td>$D_{com}$</td>
<td>14.57</td>
<td>1379.06</td>
<td>5383</td>
</tr>
<tr>
<td>Distance to a residential area (m)</td>
<td>$D_{res}$</td>
<td>7.42</td>
<td>456.78</td>
<td>1348.95</td>
</tr>
</tbody>
</table>

These tables and figures illustrate the correlation analysis and parameter details for the road traffic system of Chengdu, demonstrating the influencing factors and conditions for self-organization and entropy changes in the traffic system.
We display the traffic congestion information of a trajectory in a congestion matrix. Each row represents various information of a road section. The first column to the eighth column represents F1–F8, representing the entropy’s influencing factors and its initial observation matrix. The representation is as follows:

\[
X_{ij} = \begin{bmatrix}
4 & 42 & 11 & 10 & 4 & 813 & 623 & 21 \\
3 & 27 & 22 & 10 & 1 & 997 & 764 & 53 \\
2 & 183 & 21 & 10 & 3 & 1123 & 536 & 163 \\
2 & 51 & 25 & 10 & 2 & 1387 & 123 & 178 \\
4 & 36 & 38 & 10 & 1 & 1523 & 96 & 231 \\
4 & 18 & 32 & 10 & 1 & 1459 & 321 & 93 \\
3 & 42 & 22 & 10 & 2 & 1203 & 413 & 46 
\end{bmatrix}
\]  \tag{5}

Use linear changes to standardize the matrix to obtain a standardized matrix of screening indicators:

\[
X_{ij} = \begin{bmatrix}
1.000 & 0.230 & 0.289 & 1.000 & 1.000 & 0.534 & 0.815 & 0.091 \\
0.750 & 0.148 & 0.579 & 1.000 & 0.250 & 0.656 & 1.000 & 0.229 \\
0.500 & 1.000 & 0.553 & 1.000 & 0.750 & 0.737 & 0.702 & 0.706 \\
0.500 & 0.279 & 0.658 & 1.000 & 0.500 & 0.911 & 0.161 & 0.771 \\
1.000 & 0.197 & 1.000 & 1.000 & 0.250 & 1.000 & 0.126 & 1.000 \\
1.000 & 0.098 & 0.842 & 1.000 & 0.250 & 0.958 & 0.420 & 0.402 \\
0.750 & 0.230 & 0.579 & 1.000 & 0.500 & 0.790 & 0.541 & 0.199 
\end{bmatrix}
\]  \tag{6}

Similarly, use Python language programming for batch processing and summarize all trajectories’ observation matrix to obtain the initial observation matrix \(X_{ij}\):

\[
X_{ij} = \begin{bmatrix}
X_{ij} \cdots X_{mj} \\
\vdots \cdots \vdots \\
X_{in} \cdots X_{mm} 
\end{bmatrix}
\]  \tag{7}

\(P(S_l|S_{l-1})\) is the current state transition probability, that is, the transition between the five traffic states of the road section; the steps to calculate the state transition matrix are as follows:

(a) Group the original data

(b) According to the previous period division, suppose the time series is \(S = \{S_0, S_1, \ldots, S_l\}\)

There are \(N\) observations, divided into five states, expressed as \(C = \{C_1, C_2, C_3, C_4, C_5\}\).

After dividing the data in the training set based on different periods and road sections, a matrix of periods and road section sets is obtained. Each row represents the road congestion state of a road section in a different period. The state transition matrix form is as follows:

\[
A_S = \begin{bmatrix}
S_{1a} \cdots S_{1l} \\
\vdots \cdots \vdots \\
S_{5a} \cdots S_{5l} 
\end{bmatrix}
\]  \tag{8}

Among them, \(a, b, c, \ldots, l\) are represented as 12 time periods, and 11–85 are designated as the division numbers of road sections.

3.4.4. Model Check. After the model training is over, output the fitted model, and compare it with the test value data (November 24th–November 30th).

Figure 12 shows the prediction results from 9:30 to 9:40. The second section of the East Section of the First Ring Road, Jianshe South Road, Xinhong Road, and the East Section of the Erxian Bridge of the Middle Ring Road are slightly congested. Some sections of the road are moderately congested, and the remaining routes are unblocked or basically unblocked. From 9:30 to 9:40, except for Xinhong Road and the second section of the East Section of the First Ring Road, which is slightly different from the predicted value, the remaining road sections’ actual congestion index is the same as the predicted value.

The congestion state 3 (light congestion) is set as the congestion threshold, the congestion is greater than or equal to 4, and the acceptable state is 1–3. When the predicted value reaches 3, the congestion threshold is reached, indicating that the dissipative structure’s critical state has been reached. The influencing factors have reached the critical state of the dissipative structure. However, since the hidden Markov model’s hidden variables are not visible, the hidden variables affect the predicted value. The influence is not visible, and only the dissipative structure can be used to analyze the influencing factors listed in this study. In terms of energy research, the dissipative structure can be formed or maintained only when the open system supports exchanging information, material, and energy with the outside world. In the system of this study, the exchange of traffic is the main form of energy exchange, and the increase in traffic flow is the formation of congestion. The main factors of traffic flow will affect the road section’s average travel time and the average travel speed of the road section and ultimately affect the road section’s traffic state. In terms of entropy research, due to the dissipative structure theory, when an open system undergoes a sudden change, it will transform from its original disordered state to a new state of order in time, space, or function. Entropy is a sign of order. It is a measure of system stability. In this study, the entropy value is combined with eight influencing factors: road type, average travel time of the road section, and hidden influencing factors. The most important influencing factors are the average travel time of the road section and the road section’s average travel speed influence to act.

To quantitatively test the prediction effect of the model, three indicators, mean square error (MSE), mean absolute error (MAE), and mean fundamental percentage error (MAPE), are introduced.

It can be seen from the above results that when taking the area near the Second Ring East Road in Chengdu as the research object, consider F1–F5 (road type, average travel time of road segment, the average travel speed of road segment, period, and traffic state of road segment). Hidden Markov model prediction has good accuracy when other influencing factors are used. If the POI-based influencing factors F6–F8 (distance to the subway station, distance to commercial area, and distance to residential area) are considered, the accuracy of prediction can be further
improved. The model’s accuracy can be further verified with the visualized road network traffic status, as shown in Table 6.

In summary, the prediction results of different evaluation index models are not the same. If the mean square error, average absolute error, and average fundamental percentage error are used as evaluation indicators, the model considers factors such as POI having higher accuracy. Within the region, there is better effectiveness and scope of application.

4. Conclusions

In this study, Chengdu was selected as the research object, using Didi car-hailing data to perform data cleaning, coordinate conversion, and map matching on approximately 6 million order data and about 90 GB of trajectory data. After completing the preprocessing of the trajectory data, the characteristic analysis of the trajectory data found that there were morning peaks (8:00–10:00), noon peaks (12:00–14:00), and evening peaks (17:00–19:00), while on rest days, there were only noon peak and evening peak. The demand for travel in the city’s central area is relatively stable in terms of spatial distribution. This paper also extracted the road sections’ speed characteristics based on the road sections’ division and divided the congested road sections according to the speed. Based on the principal component analysis of the POI data, the “principal components” that have the most significant impact on the road traffic state of the POI data are found to provide a basis for further research, such as subsequent predictions. According to the principal component analysis of spatial distribution characteristics, commercial and residential areas significantly impact pickup points. In this study, the area around the east section of the Second Ring Road in Chengdu was selected as the study area, and the road segment and period were divided. Simultaneously, road types, average travel time, and road speed were considered influencing factors. When considering the set of influencing factors of POI, the distance from the boarding point to the subway station, commercial area, and residential area is considered. The initial observation matrix is constructed for the two situations, and the state transition matrix is established according to various influencing factors. Finally, combine the hidden Markov model and the dissipative structure theory, using Python to transform and solve the matrix, and verify the training set’s results and the test set. The results show that the model has a certain degree of prediction accuracy. The accuracy of the prediction model based on the GPS data of car-hailing and considering the impact of POI is higher than that without considering its impact. As commercial and residential areas have a more significant effect on pickup points, public transportation should be vigorously developed to reduce passengers’ travel demand in private cars and taxis in commercial and residential areas. Traffic organization should be strengthened near congestion points to promote traffic flow circulation and reduce congestion entropy, like intersection signal timing.

Data Availability

All data are collected from the websites http://outreach.didichuxing.com/research/opendata/ and http://www.bigemap.com/.

Conflict of 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.

Authors’ Contributions

Xiaoke Sun wrote the paper, Hong Chen provided the research ideas, Yahao Wen and Zhizhen Liu processed the data, and Hengrui Chen revised the paper.

Acknowledgments

This work was supported by the 111 Project of Sustainable Transportation for Urban Agglomeration in Western China under Grant number B20035, the National Natural Science Foundation of China under Grant number 51878062, and Xi’an Science and Technology Bureau under Grant number SZJJ2019-22.
References


