IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 3, 2020, accepted December 11, 2020, date of publication December 16, 2020, date of current version December 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045182

A Flood-Discharge-Based Spatio-Temporal Diffusion Method for Multi-Target Traffic Hotness Construction From Trajectory Data

TAO WU[1,2], PENGFEI ZHANG[2], JIANXIN QIN[1,2], DI WU[3,5], (Member, IEEE), LONGGANG XIANG[4], AND YILIANG WAN[1,2]

[1]Hunan Key Laboratory of Geospatial Big Data Mining and Application, Hunan Normal University, Changsha 410081, China
[2]College of Resources and Environmental Sciences, Hunan Normal University, Changsha 410081, China
[3]Key Laboratory for Embedded and Network Computing of Hunan Province, Hunan University, Changsha 410082, China
[4]State Key Laboratory of LIESMARS, Wuhan University, Wuhan 430079, China
[5]School of Design, Hunan University, Changsha 410082, China

Corresponding authors: Jianxin Qin (12495@hunnu.edu.cn) and Yiliang Wan (wanylir@hunnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41771474, and in part by the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 19I05.

ABSTRACT With the significant updates of location-acquisition technologies, there are more spatial-temporal trajectory data available. Spatial-temporal stream plays a crucial role in the research and applications of intelligent transportation. However, there are still some problems in the study of multi-target hotness diffusion in spatial-temporal stream scenarios: such as forecast of POIs' (Points of Interest) visitors or short-term traffic. Passenger and traffic flow can be called spatial-temporal hotness, the law of hotness diffusion in spatial-temporal of multi-hot-spots can be detected by studying their trajectory. Spatial-temporal trajectory data often presents multi-core (or multi-target) characteristics around events or behaviors, it is still a challenge to carry out multi-target modeling at the spatial-temporal level. Here, we provide a compelling method for dealing with multi-target hotness diffusion. To excavate the hotness in the movement and diffusion laws in spatial-temporal, we treat movements as a long network infrastructure flood from its source. Through modeling and analysis of OD (Origin-Destination) stream, the hotness prediction is finally achieved. Finally, two groups of experiments were used to demonstrate our method from the perspectives of passenger flow and traffic flow respectively and the experiments based on real-world data show that the effectiveness of our method in predicting the spatial diffusion state of multi-target hotness in different spatial scales. Based on the R-square, MAE (Mean Absolute Error), MSE (Mean Squared Error), and other evaluation indexes compared to the traditional prediction method and ARIMA (Autoregressive Integrated Moving Average) model. Thus, these findings suggest that our method shows more advantages than others.

INDEX TERMS Spatial-temporal trajectory data, multi-target hotness, flood-discharge model.

I. INTRODUCTION

Spatial-temporal data play a vital role in the field of intelligent transportation. As an important part of spatial-temporal data, spatial-temporal trajectory data reflect the patterns of the moving objects. Based on the trajectory data, the prediction of traffic flow and research of multiple hot spots attracted a lot of attention. Methods based on time series have been widely used for traditional hot spot analysis in the field of intelligent transportation [1]–[3]. These methods are often appropriate for single targets, and not for multiple models that needed to be trained within complex scenarios. Multi-hot-spots issues still needed to be researched further in LBS (Location Based Service), intelligent transportation, and other fields. Recently, with a large number of moving objects integrated with positioning devices, a massive amount of trajectory data are recorded to obtain the location, speed, and other information [4]. The recorded trajectory data portray numerous spatial and temporal features that are beneficial to serve intelligent transportation applications. The trajectory

The associate editor coordinating the review of this manuscript and approving it for publication was Jerry Chun-Wei Lin.

data can easily reflect the characteristics of the movement path and OD of the population and vehicles. Combined with the stream computing methods [5], [6], research on traffic flow prediction and hotness diffusion have made a major step ahead [7]–[9], and provide ideas for solving multi-hot-spots issues.

These studies of multi-hot-spots issues can be broadly divided into two categories: one is the research method based on time series which focuses on hotness prediction; the other is the research on hotness motion based on streaming data. The former describes the similarities between historical events and current events based on time series. These methods have a high prediction accuracy, but ignores the spatial correlation between multi-target. Therefore, it cannot achieve good results when multi-target are involved in prediction or emergency response [10]–[12]. The latter explores the law of multi-target hotness diffusion in the space by analyzing people's behavior characteristics through the trajectory. However, some scholars paid more attention to the limitations of hotness diffusion path and speed, ignoring the changes of hotness itself. Their research fields mainly focus on predictions of bus arrival time and short-term traffic flow [13].

To solve the multi-hot-spots issues better, we add the hotness change factor based on the second category in this paper. A flood-discharge-based spatial-temporal diffusion method is proposed for multi-target traffic hotness construction from trajectory data. In particular, we regard the process of spatial-temporal diffusion of hotness as the process of flood flowing through the river and finally into the flood storage area. The amount of flood reflects the scale of the hotness of the traffic flow, and we treat the flow of passengers as a special kind of traffic flow in this paper. The core of the flood-discharge method is to determine the source of hotness generation, the law of hotness release, the route of hotness diffusion, the wastage during hotness movement, the convergence hotness of the destination. Combined with stream computing, this method can be used to predict the hotness change of multi-hot-spots like the region of hotness intersection or destination.

In summary, the research of multi-target hotness diffusion based on trajectory data focuses on both hotness change and diffusion. Different from other studies, this study exploring hotness change at the source. The source of the hotness means the origin of the traffic generated. The source of canteen hotness is every teaching buildings, while the source of traffic hotness is every ROIs (Regions of Interest). The change of hotness at the source reflects the law of passenger flow and traffic released by the teaching building and ROIs. Combined with the law of hotness diffusion in spatial-temporal, hotness state of hot-spots can be predicted. Therefore, the requirement is different in our study:

1) This paper focuses on the correlation between geographical factors, rather than the similarity with historical data, and predicts the hotness from the spatial diffusion law of hotness.
2) This paper not only focuses on the hotness change of hot spots but also pays attention to the generation and diffusion of hotness.
3) This paper uses prediction methods of hotness change based on time series at source rather than results. The hotness change at the source is relatively simple, and the change of a similar category of targets has common characteristics.

The rests of this article are arranged as follows. Section II involves the related work in the field of multi-target hotness diffusion; Section III introduces the flood-discharge model and its combination with this study. Section IV shows two case studies on real data for evaluating the proposed approach; Finally, section V summarizes and extends the discussion.

II. RELATED WORKS

Current studies on multi-target hotness detection mainly focus on two aspects: one is the prediction method of event hotness based on time series, which mainly focus on the similarity between historical time and current events; the other one is the overall trend of hotness diffusion in space, with OD(Origin-Destination) stream which tends to ignore the actual path or short-term single-target motion state research.

Hotness prediction methods are often associated with time series. By summarizing the time series of hot spots for a long time, scholars trained the stable hotness change trend of hot spots, which has been applied in many fields such as traffic flow and network traffic prediction [14]–[21]. Karimpour [22] proposed a time series model to predict the traffic flow for a certain intersection. Lu.Deng [23] proposed a hotness prediction method based on similar topics and co-occurrence topics by combining genetic algorithm and topic hotness time series. Yaslan.Y [24] processed the historical time series of power usage based on EMD (Empirical Mode Decomposition) and SVR (Support Vector Regression), and predicted the change of future power load demand. J. Chen [25] proposed nonlinear learning integrated LSTM method for deep learning time series prediction based on LSTMs, SVRM (Support Vector Regression Machine), and EO (Extremal Optimization) to predict the future wind speed. In addition, it is worth mentioning that ARIMA method is more efficient and more accurate than traditional hotness prediction method [26].

In the studies of spatial-temporal hotness diffusion, scholars proposed some methods for two aspects: one is the overall hotness diffusion trend of space hotness, which does not include the real routes as OD stream; the other is an estimate of the motion state of the moving single target [27]–[35]. J.Qiu [36] proposed a deep learning method based on neighbors for travel time estimation (TTE), called the Nei-TTE method. Yisheng Lv [37] proposed a novel deep-learning-based traffic flow prediction method, which considers the spatial and temporal correlations inherently. Von.Landesberger [38] used a density-based clustering method to aggregate strongly connected adjacent locations into regions. C.Chen and Y.Ling [39], [40]simplified the clustering of taxi operation track data at night, revealed people's travel rules at night, and optimized the route selection of night buses. Another part of scholars predicted the movement and diffusion state of hotness based on historical data, regression model, and artificial neural network model [41], [42]. J.Amita [43] predicted the actual running status of public transport according to the artificial neural network model. According to the historical behavior and current behavior of public transport, T.Cristobal [44] combined with k-medoids clustering algorithm to predict the running state of public transport.

In summary, these methods are not well effective in the study of multi-objective space diffusion of hotness. On the one hand, the previous hotness prediction methods are not appropriate to solve multi-hot-spots problems as they focused on single target problems. On the other hand, the OD-based analyses on spatial-temporal hotness diffusion failed to consider specific routes in the real world. Therefore, this paper proposes a flood-discharge-based spatial-temporal diffusion method for multi-target hotness construction from mobility data. The process of flood-discharge is very similar to the process of hotness diffusion. However, to our best knowledge, there has been no research on traffic flow or spatial-temporal flow detection taking account of the flood-discharge model. Current researches on the flood-discharge model are mainly applied in water conservancy projects, disaster prevention, and control, and other fields [45], [46]. This study explored the law of hotness release, routes of hotness diffusion, the law of hotness diffusion, and other characteristics to analyze and solve multi-hot-spots problems.

III. METHODOLOGY

A complete flood discharge process begins with the release of the flood from the spillway. The flood proceeds along the planned channel toward the preset flood storage area and the intersection of the floods create hot spots, which need to be controlled with emphasis. At the same time, in the process of flood movement, there will be an infiltration phenomenon, that is, the flood is absorbed by the river soil, share part of the flood 'hotness'. The confluence area appears at the place where the flood converges, it can be seen as a 'hot spot' for floods. As the movement of transportation flow in the road network, it also includes the origin and destination points, routes, wastage of hotness, and a time series of vehicles depart from its origin. We hypothesized a scenario to describe this process in more detail. A team of trucks transported goods from the warehouse to the mall in batches, but some of the trucks failed to arrive at the mall due to vehicle failure, resulting in the wastage of goods. This process involves the OD of trucks' flow, the time series at the source of trucks depart, the path of trucks' movement, and the possible wastage. Hot spots will be formed at the intersection of trucks when we lengthen the distance and add different warehouses and shopping malls. Rest stations can be set to facilitate drivers in the hot spot area. In order to make the flood-discharge model truly feasible with the data of different scenarios, each part of the model is specified and parameterized. At the source of the model, we consider the law of hotness release. The release of hotness is not completed instantly, Newton's law of cooling is employed to simulate the release of hotness. The origin and destination of the hotness diffusion must be clear, clustering methods are used to determine fuzzy OD points in large scale space. Hotness diffusion along the real routes on the space, we attempt to explore the best route of hotness diffusion to estimate the movements of hotness. Hotness converges and gathers in space to form multi-hot-spots, and the hotness prediction of multi-hot-spots can be achieved by calculating the hotness at different time periods. Wastage of hotness is accompanied by hotness diffusion, hotness may no longer be attracted to its target as it moves. For example, few students may be attracted by the roadside restaurants or snacks bars and do not choose the planned canteen, resulting in a part of hotness wast.

We propose a novel framework for a flood-discharge-based spatial-temporal diffusion method for multi-target hotness construction from mobility data in this paper. Figure 1 shows the entire workflow of the study. Two different sets of traffic flow data were used to demonstrate our work in different scale space, one was canteen visitors data in school, and the other was operational vehicle data in urban. The school dataset includes canteen visitors data, schedule data, POI data (buildings in school), and school network data. The urban dataset includes vehicle trajectory data, road network data, and urban maps. After completing multi-source data collection, a data pre-processing step was performed to remove erroneous data. We applied different methods to obtain the hotness OD points of different scale space. POIs are extracted as the OD points of hotness diffusion in school data. In the trajectory data of the operating vehicle, after the stop points are extracted, OD points are obtained by HDBSCAN clustering method. Path analysis of both data is based on the actual road network. Finally, a flood-discharge model is used to predict the hotness of multiple hot spots, results will be compared with test data and other methods to verify its accuracy.

A. FLOOD-DISCHARGE MODEL OF HOTNESS

The flood-discharge model is an abstraction of the whole process of flood-discharge. In our model, the origin points of hotness regard as the source of flood-discharge, the destination points of hotness regard as the storage area of the flood, the routes of hotness diffusion regard as the river of flood movements, and the wastage of hotness diffusion regard as the infiltration of flood movements. In addition, the multi-storage-area corresponds to multi-hot-spots in this study. A hot spot is a confluence of multiple streams or floods.

1) FLOOD-DISCHARGE MODEL

The flood-discharge model mainly consists of five parts: source O_i, flood storage D_i, discharge F, channel R and

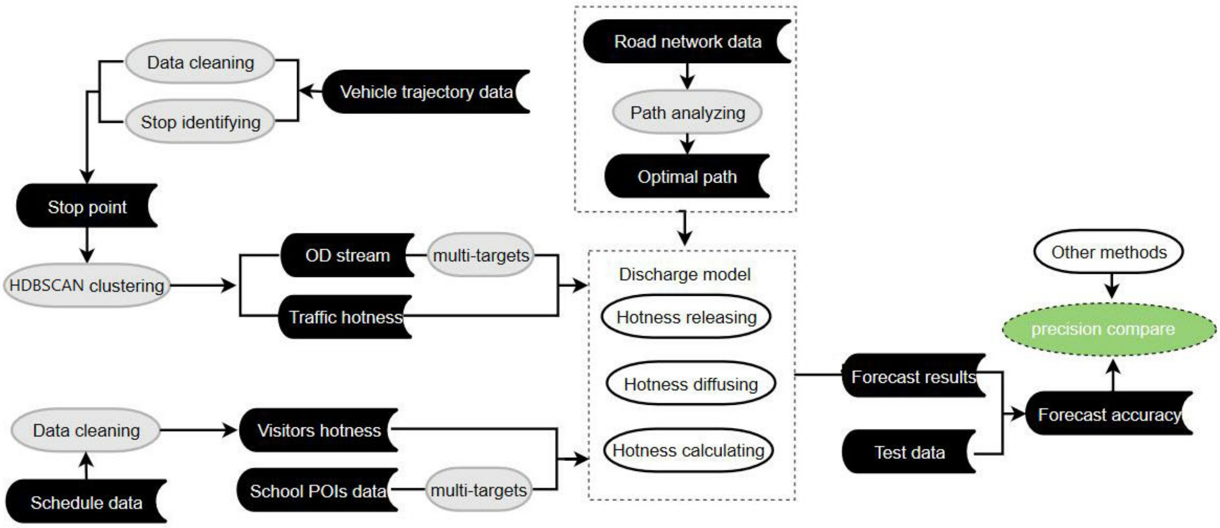


FIGURE 1. Workflow of this study.

summary infiltration *Wast*. *Wast* is the summary infiltration in the process of flood movements. This part of the flood will be absorbed by the river during its moves. This amount of flood should be removed in the actual calculation.

When moving multi-source flood converges through river channels, the discharge process can be accumulated from multi-period flood quantities by applying the following formula:

$$F = \sum_{i=1}^{n}\sum_{j=1}^{n} F_{ij} \tag{1}$$

where F_{ij} is the flood quantity discharged by O_i in spillway at time j.

As the infiltration phenomenon will occur in the whole river, the wastage of flood will increase gradually with the advance of time. The shorter the distance between the current location of the flood and storage area, the greater the total the wastage will be. It's important to note that the wastage here is the amount of all steps rather than wastage of an individual step. The amount of wastage each step is proportional to the distance, because the farther the distance, the less attractive it is. *Wast* is the total wastage of flood, it can be calculated by applying the following formula:

$$Wast = h_i^{-\rho} \tag{2}$$

where h_i is the distance from the current position of the flood to the storage area D_i, and ρ is the decay factor of the distance.

The final flood volume S into the storage area D_i is the total amount of water released F by each spillway that reaches each storage area after diversion minus the infiltration volume:

$$S = F - Wast \tag{3}$$

2) FLOOD-DISCHARGE OF HOTNESS ISSUES

The spatial-temporal variation of hotness inside the transportation network is, to some extent, analogous to the process of flood discharge. Similar characteristics between flood-discharge and hotness diffusion are mainly reflected in the following four aspects:

1) OD stream of hotness in transportation network: the diffusion of hotness and the movements of flood have definite OD points, the OD points of hotness is not always determined. The origin and destination of hotness may be a region in a larger spatial scale. In this study, points were used as the origin and destination of hotness. Therefore, OD flow information needs to be extracted and processed from traffic data.
2) Spatial-temporal dependent feature: hotness or flood are not released regularly, not instantaneously. A suitable hotness algorithm is needed to reproduce the actual sequence of hotness release. Trends at the source are simpler and easier to track, both flood-discharge and hotness diffusion.
3) Path-dependent feature: in the real world, diffusion of hotness or movements of flood have preset channels. In the traffic network, paths of hotness diffusion are affected by more factors and require more detailed path analysis.
4) Distance-dependent wastage: wastage occurs during the process of hotness diffusion or flood movements. The distance between hotness and targets affects the actual attractiveness of targets, long distances are more likely to cause hotness wastage.

3) SPATIAL-TEMPORAL CHANGE MODEL OF HOTNESS IN TRANSPORTATION NETWORK

This paper proposed a spatial-temporal change model of hotness in the transportation network. The ith source point of hotness generation is O_i, the ith destination point is D_i. In small scale Spaces, O_i and D_i represent different buildings. In large scale Spaces, O_i and D_i represent the centroids of different ROIs.

The calculation of hotness release is derived from Newton's law of cooling, it can be calculated by applying the following formula:

$$T_{ij} = T_t^{\alpha(j-t)} \tag{4}$$

where T_{ij} is the hotness released by the ith O point at time j, T_t is the hotness at the previous time t, and α is the change coefficient of hotness.

The hotness diffuse over a wide area and different path levels in large scale space. Weighted shortest path analysis is used to find the best path of hotness diffusion:

$$R = Dijkstra\,(w_i r) \tag{5}$$

where R is the best path for hotness diffusion, r is the sub-path, w_i is the weight of the sub-path, and $Dijkstra$ represents the shortest path analysis. Although R is not directly involved in subsequent calculations, it affects the time t when the hotness reaches its target.

The possible wastage of hotness also needs to be considered. wastage will appear in the process of hotness diffusion, and the quantity will increase with the diffusion process. IDW (inverse distance weight method) method is introduced to our model for this case. The formula is as follows:

$$Wast = \mathrm{per}\left(h_i^{-\rho}\right) \tag{6}$$

where $Wast$ is the entire wastage in the process of hotness diffusion, h_i is the path distance from D_i, ρ is the distance attenuation coefficient, and $per()$ is the symbol for the percentage of the result.

The final results can be calculated by applying the following formula:

$$P = \begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{n} T_{ij}^{(\alpha * \Delta t^*)} * W, & j \geq \Delta t_i \\ 0, & j < \Delta t_i \end{cases} \tag{7}$$

where,

$$T_{ij} = T_t^{\alpha(j-t)} \tag{8}$$

$$\Delta t^* = j - t \tag{9}$$

$$W = 1 - Wast \tag{10}$$

where P is the final hotness prediction result of the target region or object. Δt^* is the time interval between last time node t and current time j. Time interval Δt_i is the time of hotness moves from origin point to destination. W is the weight of actual hotness. To be noted, the hotness value is 0 when the time interval is greater than j, which means that the hotness has not to reach targets.

An illustrative example is presented to demonstrate the method. Suppose there are two team trucks carry goods from two warehouses to the mall. The time series of trucks depart from the warehouse can be modeled by Newton's cooling law as shown in Figure 2. The dotted line represents the time it takes for trucks depart from the warehouse, and the horizontal distance between the solid line represents the time

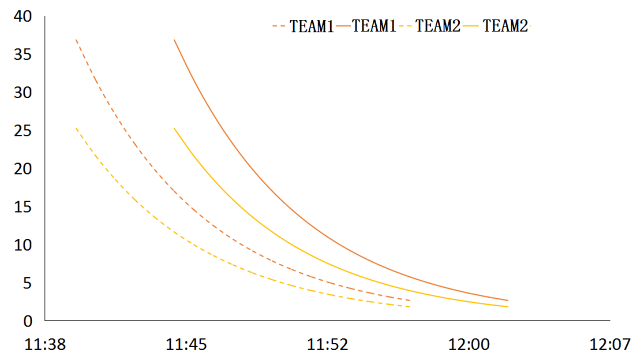


FIGURE 2. Calculated results of Newton's law of cooling and time of trucks arrival.

it takes to arrive at the warehouse. The distance between them is affected by the actual road network and path analysis methods.

The relationship between distance and hotness wastage is shown as follows Figure 3. In the process of goods transportation, there may be vehicle failure, cargo scheduling, and other factors, resulting in the wastage of goods actually delivered to the mall. The farther the truck's current location is from the mall, the higher the probability of these factors.

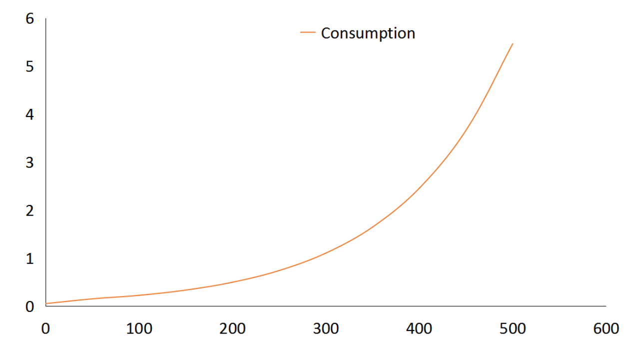


FIGURE 3. The relationship between distance and trucks wastage.

As the trucks toward the mall, this wastage accumulates with each step, and the distance between the current position of trucks and the mall is gradually decreasing. The relationship between the wastage and the distance between the current position and the target point satisfies the condition of the IDW method. It is important to note that this relationship is different from the relationship between hotness wastage and distance shown in Figure 3. The accumulation of hotness wastage in IDW is not the amount of hotness wast in each step, but the sum of them. It can be explained by the following formula:

$$Wast = per(\sum_{n=1}^{N} f(d)) = per(h^{-\rho}) \tag{11}$$

where $f(d)$ is the hotness wastage of trucks in different distances, h is the distance parameter of IDW, and ρ is the distance attenuation coefficient, $per()$ is the symbol for the

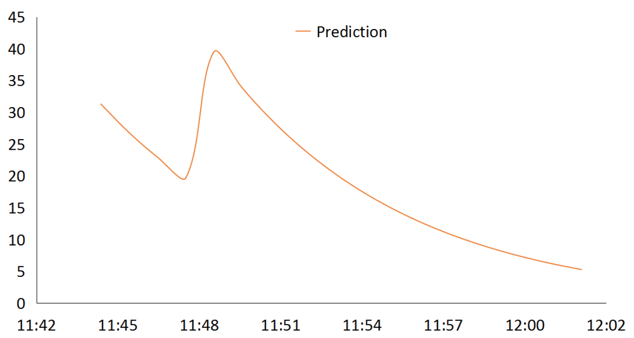


FIGURE 4. **The relationship between distance and trucks wastage.**

percentage of the result, *Wast* is the accumulation of trucks wastage in IDW.

The final prediction results of trucks arriving at the mall are shown in the following formula and Figure 4:

$$P = \sum_{n=1}^{N} (f_i(t) * (1 - Wast)) \tag{12}$$

where $f_i(t)$ is the calculated results by Newton's law of cooling from different warehouses.

B. EXTRACTION OF ORIGIN-DESTINATION MOVEMENT DATA

The origin and destination of hotness diffusion is a point in our model, it is necessary to preprocess data of large scale space. In this paper, OD is extracted by clustering the stop points. The stop points reflect the origin and destination locations of hotness, the region of hotness concentrated is our study object in large scale space. Common clustering methods can be divided into the following four categories: hierarchical clustering, partitional clustering, density-based clustering, and grid-based clustering. Density clustering is widely used in the analysis of traffic flow and spatial-temporal stream because it accepts non-spherical clustering and can reflect hot spots. In this paper, we compared three different density clustering methods of DBSCAN (Density Based Spatial Clustering of Application with Noise) [47], [48], HDBSCAN (Hierarchical Density Based Spatial Clustering of Application with Noise) [50] and CFSFDP (Clustering by Fast Search and Find of Density Peaks) [51], [52].

An example is presented here to compare these clustering methods. Figure 5 shows the results of the different clustering algorithms, which are evaluated by means of point density analysis. In point density analysis, lower density pixels are removed to facilitate comparison with different clustering results. The clustering result of HDBSCAN algorithm is the closest to the actual density, and the CFSDFP algorithm has the worst effect, while DBSCAN algorithm retains too many noise points. To sum up, the HDBSCAN algorithm is used to extract OD in this paper. HDBSCAN performs DBSCAN over varying epsilon values and integrates the result to find a clustering that gives the best stability over epsilon. This allows HDBSCAN to find clusters of varying densities (unlike DBSCAN), and be more robust to parameter selection.

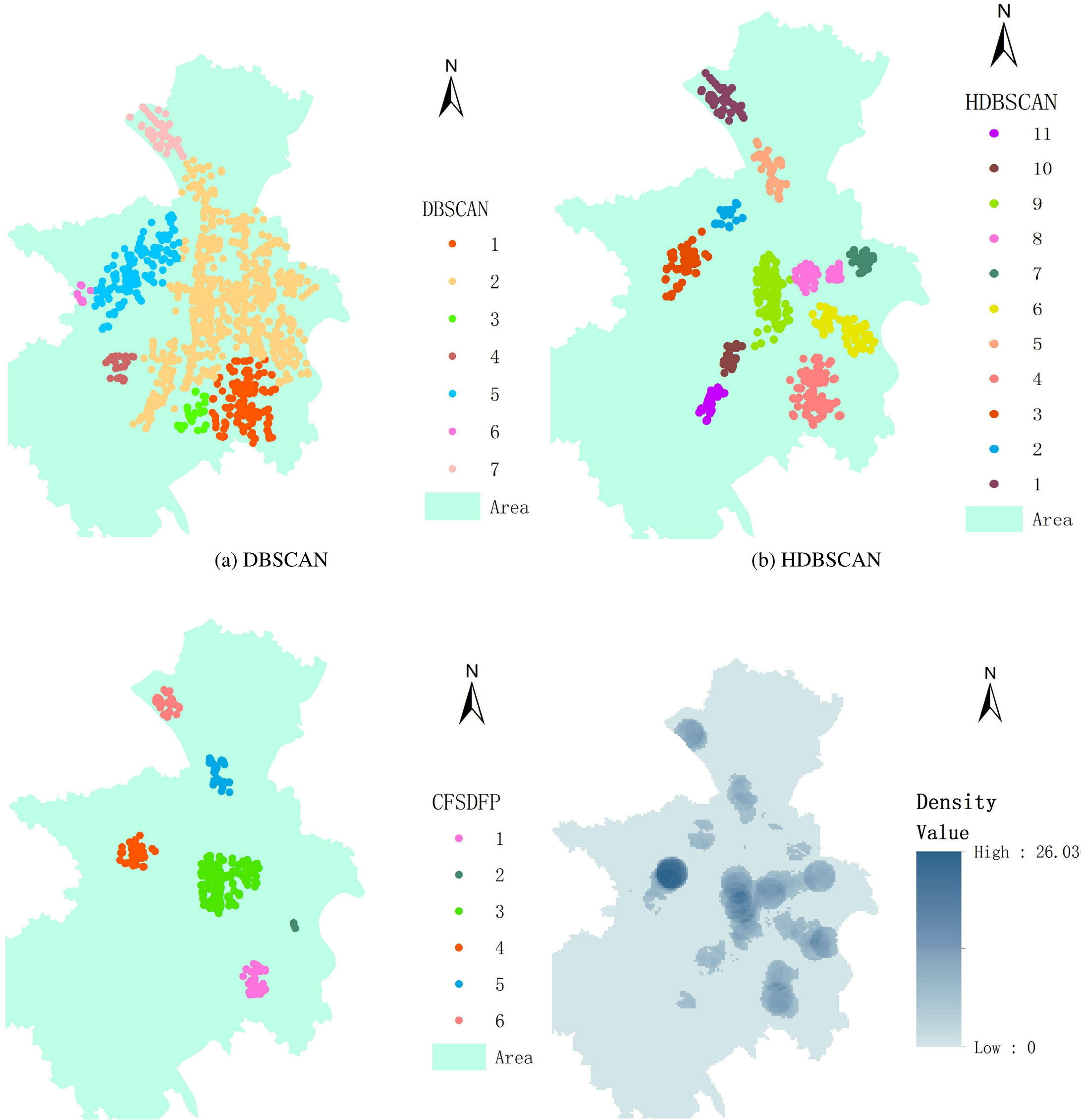
Although HDBSCAN algorithm can cluster the stops of the same attribute, it also brings some trouble to our work. HDBSCAN clustering results on the space of the area may be large in Figure 6(a). In order to reduce its influence on OD extraction, the points within these regions need to be clustered again. Therefore, this paper combines the advantages of HDBSCAN and CFSDFP methods to perform a second clustering of the HDBSCAN clustering results. CFSDFP is a clustering algorithm based on peak density. It believes that there are points with lower density around the clustering center, and the distance between these points and the clustering center is the closest compared with other clustering centers. The results of this algorithm are also non-spherical, and the complexity of the algorithm is lower. The refined result of HDBSCAN is shown in Figure 6(b), in which clusters 4,6, and 9 are divided into two small clusters.

It is important to extract centroids of the refined clusters to represent these regions or ROIS. In our study, clustering results of stop points can be any shape such as the clustering results of commercial streets is banded. However, two methods often used in polygon-to-point processes are minimum enclosing rectangle and minimum circumcircle, which are better suited to convex polygons than concave polygons. An approach of building TIN (Triangulated Irregular Network) to find the centroid is proposed in our study, we get the actual scope of the data by build TIN. This method is suitable for both convex polygons and concave polygons.

Figure 7 shows the different locations of the centroid, our method can better reduce the redundancy in the construction polygon and the result more approximate the actual data area. It is obvious that the centroid obtained by constructing TIN method on a concave surface is more reasonable.

C. DETECTION OF MULTI-TARGET HOTNESS

Combined with the flood-discharge model, hotness detection is carried out for multi-target from the whole process of hotness diffusion in our study. Different from traditional methods, hotness detection methods were used to detect the source of hotness release rather than detect the hotness of targets in our study. Newton's law of cooling is introduced in this paper to detect the change of hotness source. The release of hotness is not completed in an instant and has a certain law. It is important to find a suitable hotness prediction method to get the release law of hotness. There are few influence factors of hotness change in its source, the simple prediction algorithm can predict the hotness change quickly and efficiently. Newton's law of cooling is widely applied to hotness predictions because it can well reflect the hotness change between two time nodes and does not require too many parameters [49]. Newton's law of cooling is formulated



(a) DBSCAN (b) HDBSCAN (c) CFSDFP (d) Points Density

FIGURE 5. Clustering results of different algorithms and actual points density.

as follows:

$$T_t' = -\alpha\left(T_t - H\right) \tag{13}$$

where T_t' represents the rate of hotness change of events or POI, T_t represents the function of event hotness T with respect to time t, H represents the surrounding hotness, and α represents the proportional relationship between the surrounding hotness and hotness change rate. The formula for the integral transform is as follows:

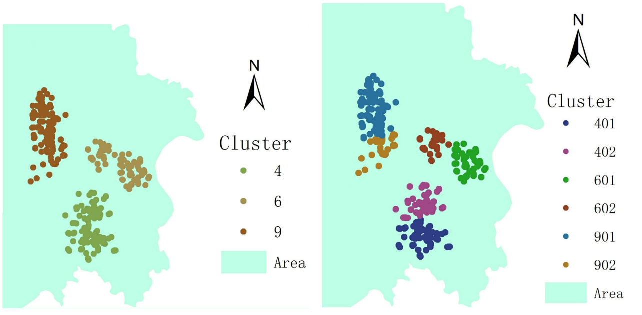
$$T = H + \left(T_0 - H\right) e^{-\alpha(t-t_0)} \tag{14}$$

where T represents the temperature of the object at the moment, T_0 represents the temperature at the previous time, and $t - t_0$ represents the time difference between the previous time and this time.

The hotness prediction at the source reflects the hotness change in the time of multi-target. Detecting the hotness of multi-target also requires attention to the diffusion process of hotness. The hotness diffuses along the actual paths and converges to form multi-hot-spots, multi-target hotness detection can be achieved by hotness prediction of these hot spots.

D. PATH ANALYSIS

The spatial-temporal diffusion of multi-target hotness is based on the actual paths. In this paper, we use the shortest path algorithm and adjust it according to the actual situation to get the best path. The speed and current location of hotness diffusion can be grasped effectively by path analysis. As a classical algorithm, the shortest path algorithm is suitable for most scenarios.



(a) The result of HDBSCAN (b) The result of our refined clustering

FIGURE 6. Results of HBSCAN and its refined clustering.

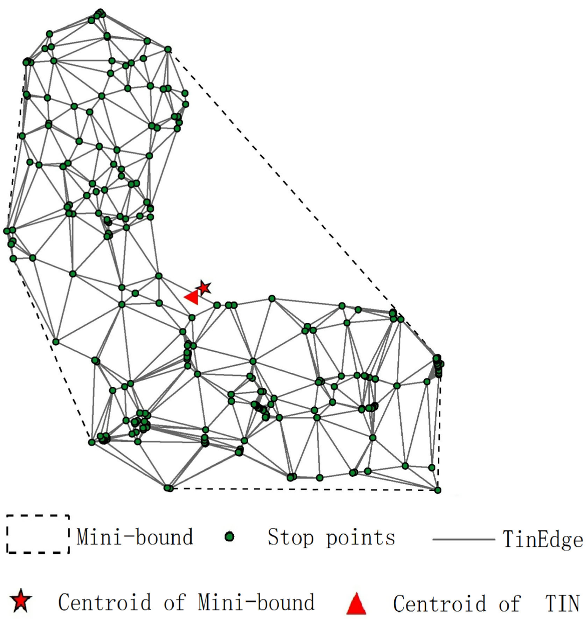


FIGURE 7. The different centroid of build TIN and minimum enclosing rectangle.

Two scenarios, small scale space and large scale space, were used in our study. In small scale space, the road network is characterized by sparsity, small grade difference, and short distance. The path of hotness diffusion is not complicated, and the shortest path analysis can be applied to this scenario.

In large scale space, the road network is characterized by well-developed, hierarchical road levels, and long distance. We combine the weighted shortest path analysis with buffer construction. Based on the actual situation, the weighted shortest path analysis modify the results of the shortest path analysis by setting the weights. The weights include levels and attractiveness of the road. Hotness diffusion in different roads has different speed because of the speed limits. Different roads also have different attractions for different types of vehicles. For instance, the taxis may prefer places with more passengers on the way, while the trucks tend to take the expressway with less traffic. Although weighted shortest path analysis can find the optimal path for hotness diffusion, it should also consider the case of user detour. People often choose the road around the optimal path for a short detour when the current road is congested. Building buffers for this optimal path can be applied in this case. In order to more accurately grasp the extent of hotness diffusion, we conducted buffer analysis on the optimal route according to the distance of road intersections. Diffusion and computation of hotness based the on buffer path in large scale space.

E. WASTAGE OF HOTNESS

Wastage occurs during hotness diffusion. Hotness may no longer be attracted to the target as it diffuse. As the increases time of diffusion, the wastage of hotness is increasing and the distance between the current location of hotness and its targets is shrinking. IDW is adopted here to solve this problem. It takes the distance between the interpolating point and the sample point as the weight to carry out the weighted average, and the inverse distance weighting method mainly depends on the power value of the inverse distance. By adjusting the power value, the influence of the distance factor on POI attraction or attention can be changed. The process of IDW is as follows:

1) Calculating the distance from a point (or position) to the target POI point

$$h = distance(D, hot) \tag{15}$$

where h is the distance between current position *hot* of hotness and target D.

2) Calculating the weight of each point: the weight is the function of the inverse of the distance.

$$\rho = f(1/h) \tag{16}$$

where ρ is the distance attenuation coefficient, $f(1/h)$ represent ρ is is a function of the inverse of h.

3) Calculating the final result.

$$Wast = h_i^{(-\rho)} \tag{17}$$

where *Wast* is the total wastage of hotness in the process of hotness diffusion.

IV. EXPERIMENT AND ANALYSIS

In this section, the proposed method is evaluated in two different scenarios, based on real trajectory data and road network data. The first experiment is research on multi-target hotness diffusion in school, while the second one is in urban traffic. The hotness detection object of small scale space is the number of canteen visitors, while the object of large scale space is the number of vehicles in different hot-spots.

A. CASE 1: MULTI-HOT-SPOTS ISSUES IN SMALL SCALE SPACE ROAD NETWORK

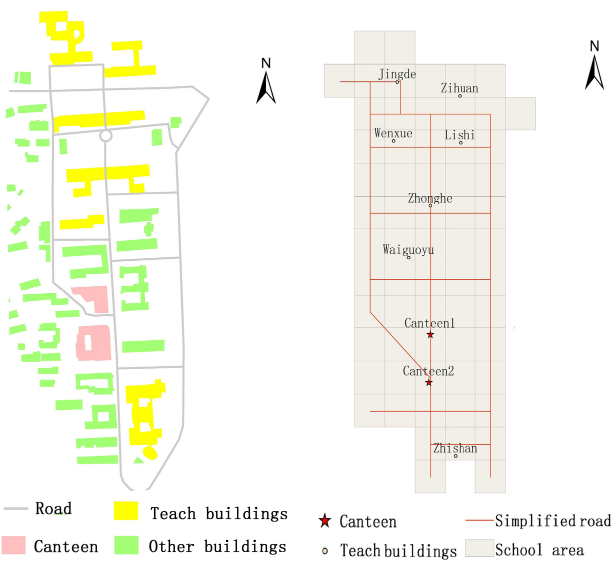
We apply the flood-discharge model to real canteen visitors' data to show its effectiveness in small scale space. The dataset includes the schedule data and canteen visitors data. The

TABLE 1. An example of campus cards' access records in a canteen.

User ID	Time	Machine ID	others
S20******14	2019/4/15 12:02:08	HQ016	...
S20******05	2019/4/15 12:03:18	HQ002	...
S20******34	2019/4/15 12:03:35	HQ002	...
...	...	...	...

schedule data records the number of students and the corresponding classroom from April 15th, 2019 to April 26th, 2019. The canteen visitors data records the number of canteen visitors between 11:45 am-12:00 am on the corresponding dates. The canteen visitors dataset comes from the campus cards' access records of the school's logistics department. An example of campus cards' access records in a canteen is shown in Table 1. Unnecessary and private data is hidden or processed in this example. The machine ID can distinguish different canteens.

In addition, we also obtained the basic vector data of the school. The study region is divided into regular grids, and the actual path is simplified to facilitate the understanding of the model. The size of the fishing net was set as 50m*50m based on the students' walking speed and at an interval of one minute in Figure 8.



(a) Wed-canteen2 (b) Wed-canteen1

FIGURE 8. School map and its simplified school map.

1) OD AND PATH ANALYSIS

In this case, we explore the spatial-temporal diffusion of canteen visitors. Since the road network on the campus is sparse and the difference in road grade is small, the shortest path algorithm can meet the research requirements of this case. The roads in school are relatively straight, and we carry out the shortest path analysis through the simplified road network, which will only produce a small error. Origin of hotness is the teaching building, and destination is the canteen in Figure 9. Except for Zhishan Building, most of the teaching buildings are far away from the canteen.

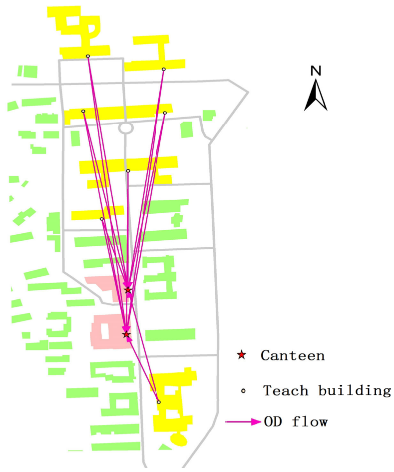


FIGURE 9. OD stream of canteen visitors in school.

Figure 10 shows the potential canteen visitors of each building on different weekdays. The Zhishan Building is the building of the basic courses, and has the highest potential visitors. They are small differences in the number of visitors from different workdays.

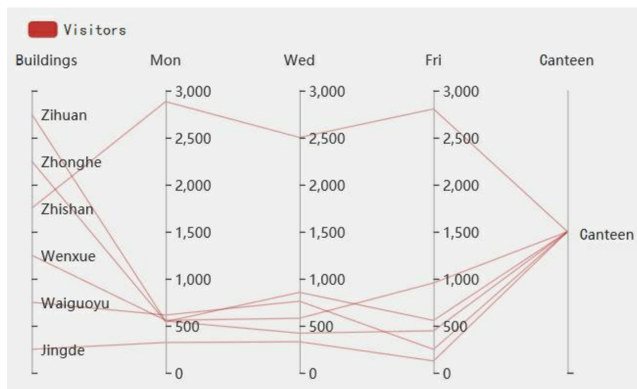
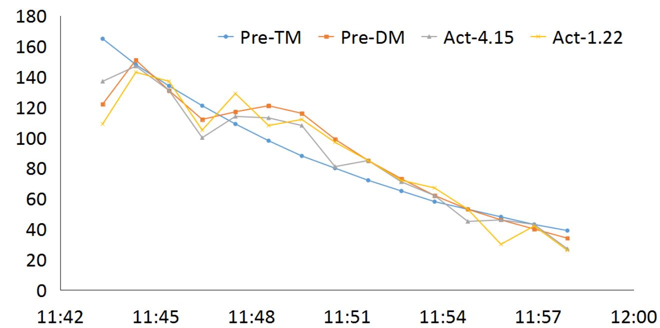


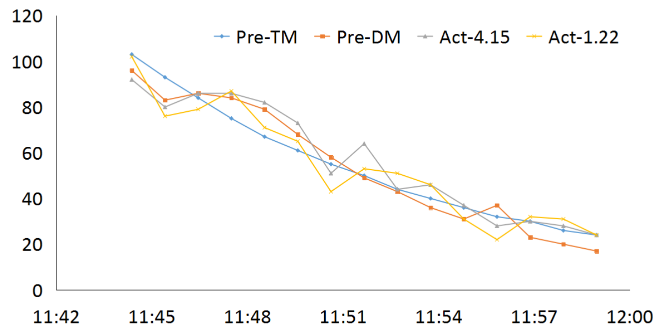
FIGURE 10. Potential visitors of each building on different weekdays.

2) ANALYSIS OF RESULTS

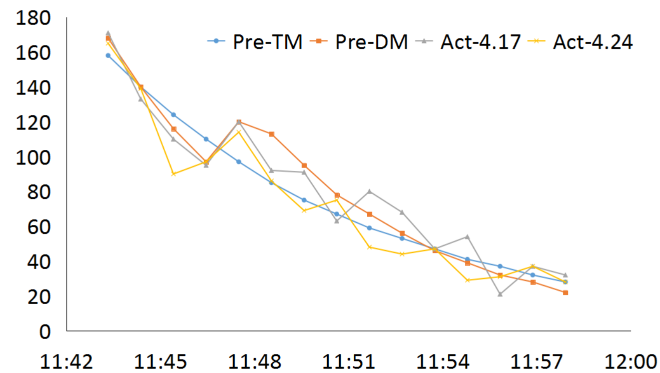
To verify the effectiveness of our method in small scale space, this experiment compares the calculated results of this method with the time series curve of actual canteen visitors. The relevant parameters were obtained from multiple experiments: Newton cooling rate $\alpha = -0.155$; distance decay rate $\rho = -0.142$; the ratio of attraction between the two canteens is 0.625:0.375. The traditional method(TM) is introduced to compare with our method. The traditional method uses Newton's cooling law to directly fit the historical data of the number of canteen visitors, without considering the generation and diffusion of hotness. Combined with the multi-day



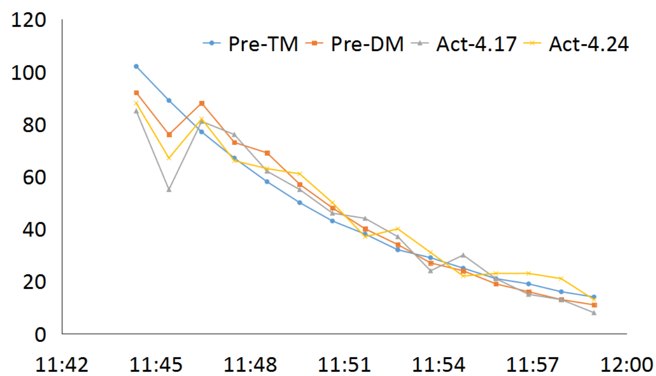
(a) Mon-canteen1



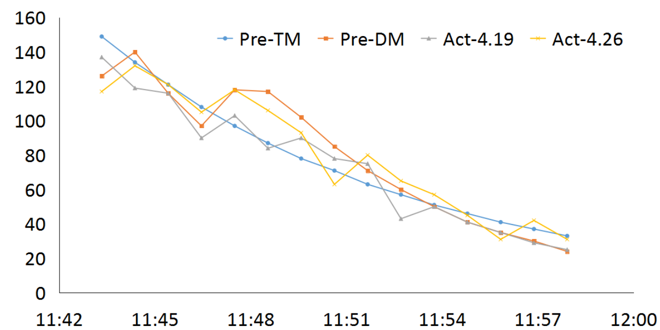
(b) Mon-canteen2



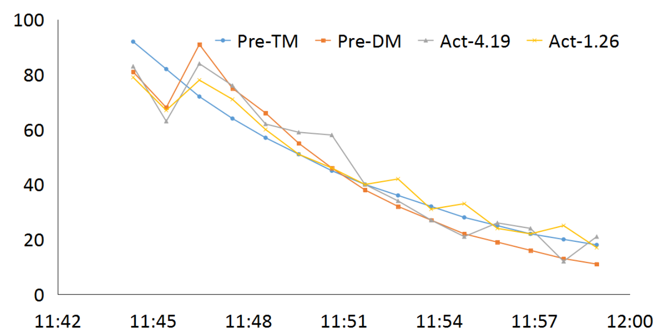
(c) Wed-canteen1



(d) Wed-canteen2



(e) Fri-canteen1



(f) Fri-canteen2

FIGURE 11. Predicted results of canteen visitors and actual visitors in different day.

data, the average coefficient method was used to optimize the coefficient, and the final prediction result was obtained.

Figure 11 shows the prediction of two restaurants on different working days. Each subgraph contains four curves: prediction of our method, the actual values, and the prediction of the traditional method. Our prediction method is closer to the actual curve in the variation of details and has more reference value. Traditional methods without considering the laws of heat generation and diffusion, cannot adapt to the fluctuations caused by the arrival time of hotness and other factors. It can only find the laws of such fluctuations through constant training models, which is less efficient. Different periods and different hot spots also require constant training models, so it is not advisable to detect multi-hot-spots in this way. To be noted, there are still some differences between our prediction and actual value because of the unified parameters in the experiment. These differences are more pronounced at

TABLE 2. R-suqare of Dredicted and Dctual Disitors for Different Date.

Date	Method	Canteen1	Canteen2
4.15	DM	0.932	0.953
4.15	TM	0.891	0.883
4.17	DM	0.947	0.884
4.17	TM	0.873	0.895
4.19	DM	0.889	0.858
4.19	TM	0.914	0.913
4.22	DM	0.893	0.949
4.22	TM	0.875	0.720
4.24	DM	0.923	0.931
4.24	TM	0.811	0.914
4.26	DM	0.944	0.927
4.26	DM	0.869	0.844

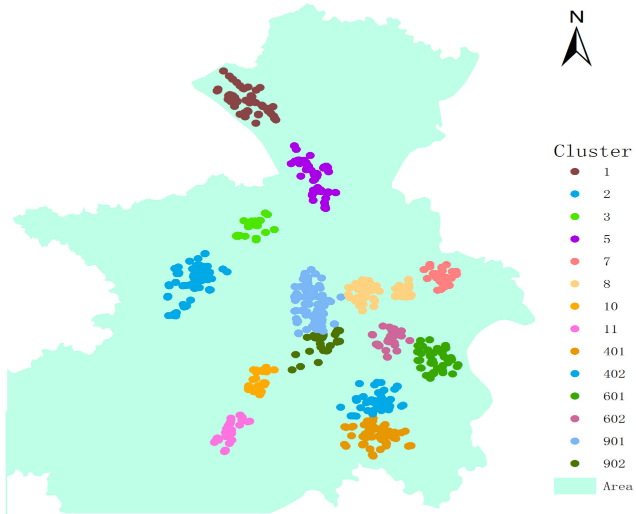


FIGURE 12. Clustering results of refined HDBSCAN.

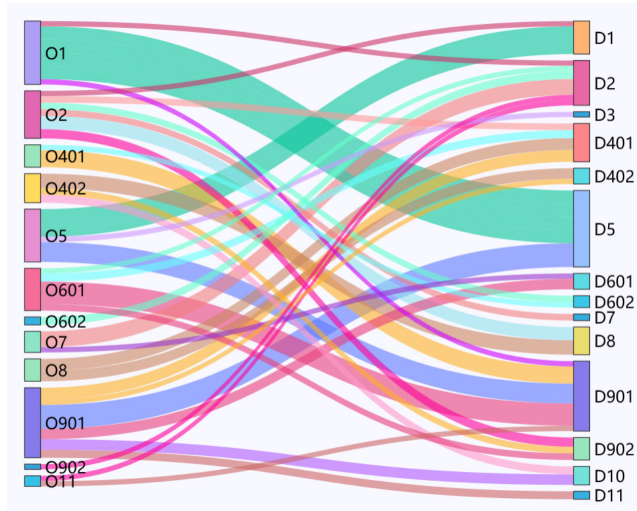


FIGURE 13. The main movement of traffic flows between different ROIs.

TABLE 3. Type of stop points of DBSACN clustering result.

ROIs	TYPE	POIs
901 902		Wuyi square Furong square
8 3	Business district	Dongying commercial plaza
7 601		Jintai square et al
2 11		Ginkgo park Meixihu park
401	Tourist area	Juzizhou park
3 10		Xiangjiang river landscape et al.
2 5 602		Datang community
401 402	Residential area	Shanyucheng community et al
3 1		Bafang community
8 5		Changsha south bus station
601 7	Transportation junction	Changsha railway station
401 105		Changsha HSR station et al

the end of the curve due to a gradual increase in the proportion of outsiders. The distance factor causes the hotness change of canteen2 to start and end one minute later than the canteen1. Table 2 shows the R-square values predicted by the two methods on different dates. The accuracy of our method(DM) is generally higher than that of traditional methods(TM), and the R-square value reaches more than 8.5. Table 2 shows that our overall prediction accuracy is better than the traditional method, and R-square is greater than 0.8. DM had a lower forecast accuracy than TM on April 19th, possibly because there were fewer people visiting the canteen than on April 26. In general, our method can achieve good results of hotness prediction in small scale space.

B. CASE 2: MULTI-HOT-SPOTS ISSUES IN LARGE SCALE SPACE ROAD NETWORK

We applied the flood-discharge model to real operating vehicle trajectory data to show its effectiveness in large scale space. The research area is the main urban of Changsha and dataset is an operating vehicle GPS dataset recorded in the city of Changsha on January 1, 2015, 1.04 million GPS records of 560 vehicles were extracted from the original dataset. The basic road network data of Changsha is from OSM.

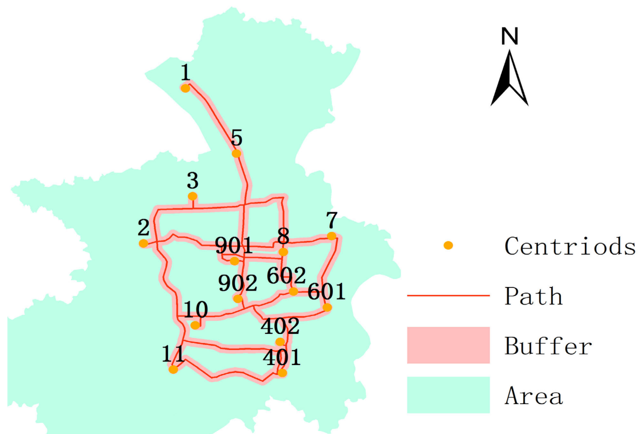


FIGURE 14. Result of weighted shortest path analysis and its buffer.

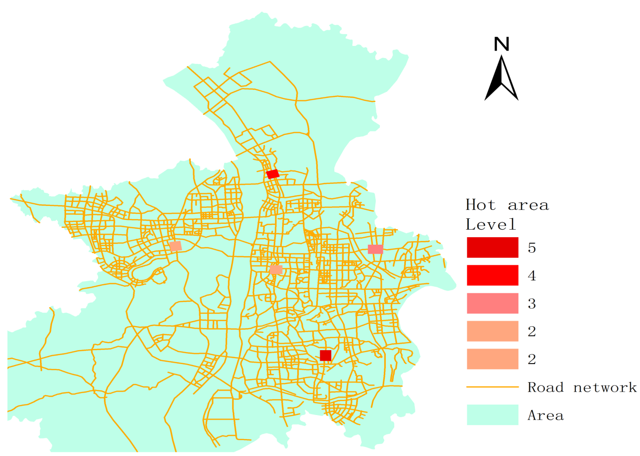
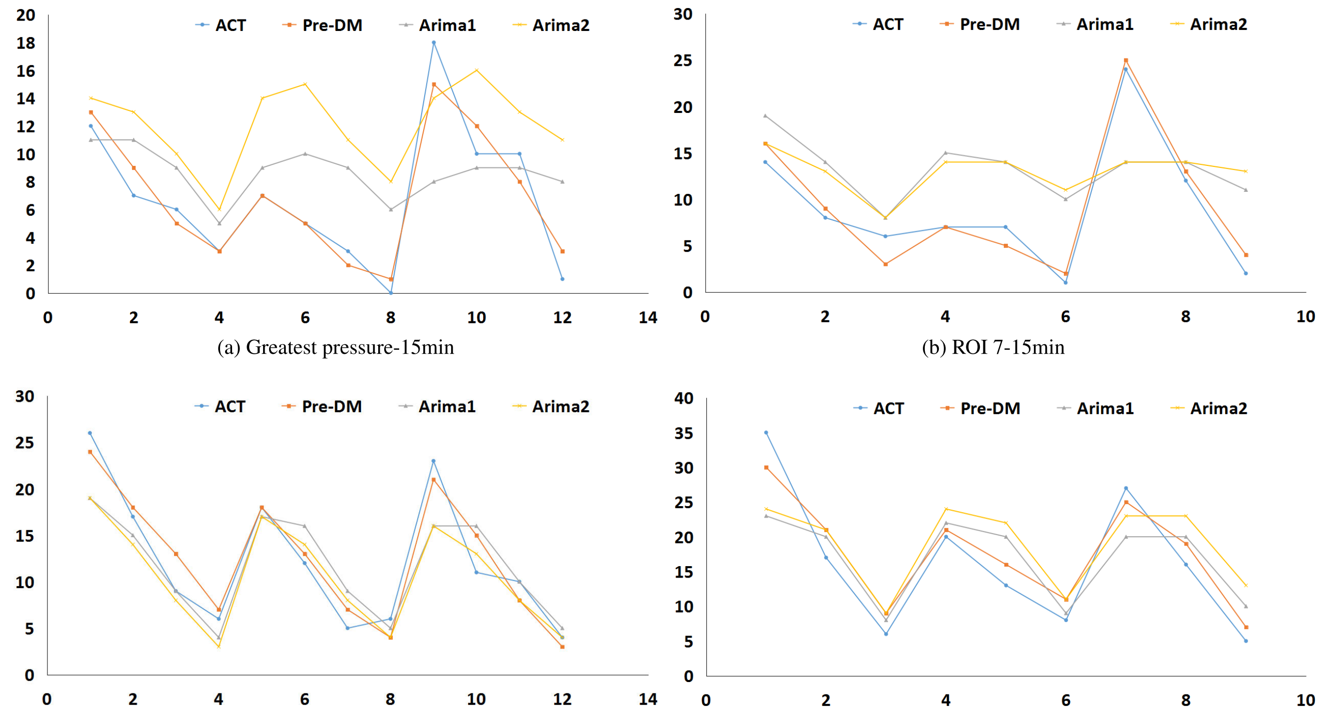


FIGURE 15. The main hot area of traffic in urban.



(a) Greatest pressure-15min (b) ROI 7-15min (c) Greatest pressure-20min (d) ROI 7-20min

FIGURE 16. The predicted results of different time interval by two methods.

1) STOP POINTS AND OD EXTRACT

We use the time-distance threshold method to extract stop points. The first point of a short trajectory will be regarded as a stop point if the trajectory did not move more than 60m in 30 minutes. The experiment extracted about 3600 spots from the original trajectory. A complete OD includes at least two stop points, vehicles with only one stop record will not participate in OD extraction. In this paper, OD is extracted by HDBSCAN clustering of stop points. The parameter of HDBSCAN is set according to multiple experiments: MinPts=35. Figure 12 shows the results of refined HDBSCAN. According to the method proposed in this paper, type 4 is refine clustered as 401 and 402, and the same goes for categories 6 and 9. Table 3 shows the categories of regions represented by centroid and the main POIs in the regions. POIs are mainly divided into four categories: Business district, Tourist area, Residential area, and Transportation junction. These categories also fit the characteristics of the operating vehicle.

Figure 13 shows the main movement of traffic flows between different ROIs. The 'O' and 'D' represent the different origins and destinations (e.g 'O1' and 'D1' means 'origin' and 'destination' with the centroid of cluster 1), while the colors and widths of the bands represent different OD stream and stream sizes. As can be seen from the figure, the major colors of the band mean that the operating vehicles move mainly between business districts and residential areas.

2) PATH ANALYSIS

The weight of the shortest path analysis in large scale space is determined by the speed limit and the attractiveness of the road. The speed limit on the road is 80km/h for expressway, 60km/h for the main road, 40km/h for secondary road, and 30km/h for branch road. The attract coefficients of the road are set as follows: 2.0 for expressway, 1.4 for the main road, 1.0 for secondary road, and 0.6 for branch road.

Figure14 shows the optimal path from the weight shortest path analysis and the buffer of the optimal path. The buffer radius is set to 800m according to the interval between the intersections of main roads in Changsha.

3) ANALYSIS OF RESULTS

To verify the effectiveness of our method in large scale space, this experiment predicted the hotness variations of areas with the greatest traffic pressure. Figure 15 shows the areas with the main traffic pressure. These areas contain a number of commercial or transportation hub POIs, which meet the activity characteristics of operating vehicles. The area with the greatest pressure is near the Changsha north bus station, which is an important traffic junction and close to mainly outflow area 'O1' and inflow area 'D5' as shown in Figure 13.

We compared the effectiveness of our method with the ARIMA model to detect hotness, and the greatest hotness area in Figure 15 and ROI-7 was selected as the test area. Figure 16 shows the prediction accuracy of the two methods at different time intervals at 2:00 am-4:00 am. ARIMA1 does not include seasonal factors, and ARIMA2 set the seasonal factors as 3 hours. This experiment compares the effectiveness of the three methods in different time intervals: 15 minutes and 20 minutes.

Each subgraph in Figure 16 contains four curves: prediction of our method, the actual values, prediction of ARIMA1 method, and prediction of ARIMA2 method. The results predicted by our method are closer to the actual data. Comparing the subgraphs, we can find that ARIMA model is more suitable for the curve with regular fluctuation, and the effect of ARIMA model is poor when the fluctuation regularity is vague. With the increase of time interval, the law of curve change tends to be stable, and the effects of these methods are improved.

TABLE 4. Hotness prediction accuracy by different method.

AREA	METHOD	MAE15	MSE15	MAE20	MSE20
	DM	1.25	2.42	1.44	2.78
Greatest pressure	ARIMA1	4	23.5	6.44	49.33
	ARIMA2	5.92	41.92	6.22	50.67
	DM	1.83	4.67	2.89	9.56
ROI-7	ARIMA1	2.83	13.83	4.78	33.44
	ARIMA2	2.75	11.92	5.88	42.33

Table 4 shows that our approach in general better than ARIMA method. MAE and MSE are two evaluation methods based on the difference between real and predicted values. Overall, the results of ARIMA1 is better than ARIMA2 because the regularity of 'ACT' curve fluctuation is not obvious, thus the seasonal factors do not play a role in ARIMA2. With the increase of time interval, the regularity of the curve is more obvious, and the effect of ARIMA2 is gradually approaching that of ARIMA. Our method is more accurate, because our method does not rely on curve fitting, but on the law of hotness diffusion to predict the hotness.

V. CONCLUSION AND DISCUSSION

The construction of the multi-hot-spots problem is still a challenge in the field of intelligent transportation. In this paper, we present a flood-discharge-based spatial-temporal diffusion method for multi-target traffic hotness construction from trajectory data. This study detects hotness changes of multi-target from the perspective of hotness diffusion, rather than predicting hotness changes directly from the target itself. According to the similarity between the process of flood-discharge and hotness diffusion, this paper explored the law of hotness change, and find the actual path of hotness diffusion. The hotness change state of multi-hot-spots can be predicted by calculating the hotness in the confluence area. Besides, this study also considers the wastage of hotness in the process of diffusion. This method can detect the hotness of multi-target simultaneously, which provides references for related research. Finally, we applied the method to real schedule data and trajectory data, and the effectiveness of this method is demonstrated by several cases in different scale space. The two experiments prove the validity of using the flood-discharge model to solve multi-hot-spots problems and the necessity of studying hotness generation and diffusion. In the small scale space, our method has shown more excellent predicted results than the traditional method. In the large scale space, our method is adaptable and accurate for predicting traffic flow hotness inside road network, and verify the effectiveness of our method by comparing with the ARIMA model.

As future work, we intend to improve the flood-discharge model from many aspects. First, novel clustering methods such as GCN-based clustering algorithm will be considered to optimize our model. Second, novel methods will be employed to refine hotness characteristics. By further dividing the characteristics of operating vehicles, the behavioral characteristics of drivers of different operating vehicles can be better distinguished, more appropriate routes can be planned, and the accuracy of multi-hot-spots prediction can be improved. Finally, the law of hot wastage can be further discussed. Road intersections, POIs, and other factors will be concerned with the analysis of hotness wastage.

ACKNOWLEDGMENT

(Tao Wu and Pengfei Zhang are co-first authors.)

REFERENCES

[1] Y. Xiao, J. J. Liu, J. Xiao, Y. Hu, H. Bu, and S. Wang, "Application of multiscale analysis-based intelligent ensemble modeling on airport traffic forecast," *Transp. Lett.*, vol. 7, no. 2, pp. 73–79, Mar. 2015.

[2] Y. Chen, L. Shu, and L. Wang, "Traffic flow prediction with big data: A deep learning based time series model," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, May 2017, pp. 1010–1011.

[3] Y. Li, X. Jiang, H. Zhu, X. He, S. Peeta, T. Zheng, and Y. Li, "Multiple measures-based chaotic time series for traffic flow prediction based on Bayesian theory," *Nonlinear Dyn.*, vol. 85, no. 1, pp. 179–194, Jul. 2016.

[4] T. Wu, J. Qin, and Y. Wan, "TOST: A topological semantic model for GPS trajectories inside road networks," *ISPRS Int. J. Geo-Information*, vol. 8, no. 9, p. 410, Sep. 2019.

[5] T. Hunter, T. Das, M. Zaharia, P. Abbeel, and A. M. Bayen, "Large-scale estimation in cyberphysical systems using streaming data: A case study with arterial traffic estimation," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 4, pp. 884–898, Oct. 2013.

[6] Z. Galić, *Spatio-Temporal Data Stream Clustering*. New York, NY, USA: Springer, 2016.

[7] P. Sun, N. Aljeri, and A. Boukerche, "Machine learning-based models for real-time traffic flow prediction in vehicular networks," *IEEE Netw.*, vol. 34, no. 3, pp. 178–185, May/Jun. 2020.

[8] F. Zhao, G.-Q. Zeng, and K.-D. Lu, "EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 101–113, Jan. 2020.

[9] T. Anwar, C. Liu, L. V. Hai, and M. S. Islam, "Roadrank: Traffic diffusion and influence estimation in dynamic urban road networks," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2015, pp. 1671–1674.

[10] W. Song, L. Wang, Y. Xiang, and A. Y. Zomaya, "Geographic spatiotemporal big data correlation analysis via the Hilbert–Huang transformation," *J. Comput. Syst. Sci.*, vol. 89, pp. 130–141, Nov. 2017.

[11] R. Beiranvand, A. Karimi, A. Delpisheh, K. Sayehmiri, S. Soleimani, and S. Ghalavandi, "Correlation assessment of climate and geographic distribution of tuberculosis using geographical information system (GIS)," *Iranian J. Public Health*, vol. 45, no. 1, pp. 86–93, 2016.

[12] A. Ermagun and D. Levinson, "Spatiotemporal traffic forecasting: Review and proposed directions," *Transp. Rev.*, vol. 38, no. 6, pp. 786–814, Nov. 2018.

[13] A. Achar, D. Bharathi, B. A. Kumar, and L. Vanajakshi, "Bus arrival time prediction: A spatial Kalman filter approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1298–1307, Mar. 2020.

[14] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Hoboken, NJ, USA: Wiley, 2015.

[15] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Comput. Surv.*, vol. 51, no. 4, 2018, Art. no. 3161602.

[16] C. K. Moorthy and B. G. Ratcliffe, ''Short term traffic forecasting using time series methods,'' *Transp. Planning Technol.*, vol. 12, no. 1, pp. 45–56, Jul. 1988.
[17] U. Mori, A. Mendiburu, and J. Lozano, ''TSdist: Distance measures for time series data,'' *R Package Version*, vol. 3, 2016.
[18] F. H. Alqahtani and S. F. Crone, ''Multivariate k-nearest neighbour regression for time series data—A novel algorithm for forecasting UK electricity demand,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–8.
[19] Haviluddin and R. Alfred, ''Short-term time series modelling forecasting using genetic algorithm,'' *Adv. Sci. Lett.*, vol. 24, no. 2, pp. 1219–1223, Feb. 2018.
[20] X. Jingting, F. Jun, S. Xia, Z. Lei, and L. Xiaoning, ''Hot news click rate prediction based on extreme learning machine and grey verhulst model,'' in *Proc. ELM*. Cham, Switzerland: Springer, 2018, pp. 89–97.
[21] O. Laib, M. T. Khadir, and L. Mihaylova, ''A Gaussian process regression for natural gas consumption prediction based on time series data,'' in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Jul. 2018, pp. 55–61.
[22] M. Karimpour, A. Karimpour, K. Kompany, and A. Karimpour, ''Online traffic prediction using time series: A case study,'' in *Integral Methods in Science and Engineering: Practical Applications*, vol. 2. Cham, Switzerland: Springer, 2017, pp. 147–156, doi: 10.1007/978-3-319-59387-6_15.
[23] L. Deng, Q. Liu, J. Xu, J. Huang, B. Zhou, and Y. Jia, ''Predicting Popularity of Topic Based on Similarity Relation and Co-occurrence Relation,'' in *Proc. Int. Conf. Intell. Interact. Syst. Appl.*, 2018, pp. 163–170.
[24] Y. Yaslan and B. Bican, ''Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting,'' *Measurement*, vol. 103, pp. 52–61, Jun. 2017.
[25] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, and K.-D. Lu, ''Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization,'' *Energy Convers. Manage.*, vol. 165, pp. 681–695, Jun. 2018.
[26] M. H. Amini, A. Kargarian, and O. Karabasoglu, ''ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation,'' *Electric Power Syst. Res.*, vol. 140, pp. 378–390, Nov. 2016.
[27] A. Hofleitner, R. Herring, and A. Bayen, ''Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning,'' *Transp. Res. B, Methodol.*, vol. 46, no. 9, pp. 1097–1122, Nov. 2012.
[28] J. Alonsomora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, ''On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment,'' *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 3, pp. 462–467, 2017.
[29] W. Tian, M. Lauer, and L. Chen, ''Online multi-object tracking using joint domain information in traffic scenarios,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 374–384, Jan. 2020.
[30] B. Zhu and X. Xu, ''Urban principal traffic flow analysis based on taxi trajectories mining,'' in *Proc. International Conf. Swarm Intell.*, 2015, pp. 172–181.
[31] N. G. Polson and V. O. Sokolov, ''Deep learning for short-term traffic flow prediction,'' *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.
[32] G. Andrienko, N. Andrienko, G. Fuchs, and J. Wood, ''Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 9, pp. 2120–2136, Sep. 2017.
[33] Z. Zhou, L. Meng, C. Tang, Y. Zhao, Z. Guo, M. Hu, and W. Chen, ''Visual abstraction of large scale geospatial origin-destination movement data,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 43–53, Jan. 2019.
[34] J. He, H. Chen, Y. Chen, X. Tang, and Y. Zou, ''Diverse visualization techniques and methods of Moving-Object-Trajectory data: A review,'' *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 2, p. 63, Jan. 2019.
[35] X. Yao, D. Zhu, Y. Gao, L. Wu, P. Zhang, and Y. Liu, ''A stepwise spatio-temporal flow clustering method for discovering mobility trends,'' *IEEE Access*, vol. 6, pp. 44666–44675, 2018.
[36] J. Qiu, L. Du, D. Zhang, S. Su, and Z. Tian, ''Nei-TTE: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city,'' *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2659–2666, Apr. 2020.
[37] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, ''Traffic flow prediction with big data: A deep learning approach,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[38] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren, ''MobilityGraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 11–20, Jan. 2016.
[39] Y. Ling, J. Zong-fu, J. Shou-xu, R. Xiang-min, and Z. Fu-sheng, ''Urban night bus routes planning with taxi traces,'' in *Proc. 12th Int. Conf. Comput. Sci. Edu. (ICCSE)*, Aug. 2017, pp. 375–379.
[40] C. Chen, D. Zhang, Z.-H. Zhou, N. Li, T. Atmaca, and S. Li, ''B-planner: Night bus route planning using large-scale taxi GPS traces,'' in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Mar. 2013, pp. 225–233.
[41] F. Zhang, X. Zhu, T. Hu, W. Guo, C. Chen, and L. Liu, ''Urban link travel time prediction based on a gradient boosting method considering spatiotemporal correlations,'' *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 11, p. 201, Nov. 2016.
[42] T. Yin, G. Zhong, J. Zhang, S. He, and B. Ran, ''A prediction model of bus arrival time at stops with multi-routes,'' *Transp. Res. Procedia*, vol. 25, pp. 4623–4636, 2017.
[43] J. Amita, J. Sukhvir Singh, and G. Pradeep Kumar, ''Prediction of bus travel time using artificial neural network,'' *Int. J. Traffic Transp. Eng.*, vol. 5, no. 4, pp. 410–424, Dec. 2015.
[44] T. Cristóbal, G. Padrón, A. Quesada-Arencibia, F. Alayón, G. de Blasio, and C. R. García, ''Bus travel time prediction model based on profile similarity,'' *Sensors*, vol. 19, no. 13, p. 2869, Jun. 2019.
[45] R. Cheah, L. Billa, A. Chan, F. Y. Teo, B. Pradhan, and A. M. Alamri, ''Geospatial modelling of watershed peak flood discharge in selangor, malaysia,'' *Water*, vol. 11, no. 12, p. 2490, Nov. 2019.
[46] M. Higashino and H. G. Stefan, ''Variability and change of precipitation and flood discharge in a japanese river basin,'' *J. Hydrol.: Regional Stud.*, vol. 21, pp. 68–79, Feb. 2019.
[47] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, ''Dbscan revisited, revisited: Why and how you should (still) use DBSCAN,'' *ACM Trans. Database Syst.*, vol. 42, no. 3, p. 19, Aug. 2017.
[48] K. Mahesh Kumar and A. Rama Mohan Reddy, ''A fast DBSCAN clustering algorithm by accelerating neighbor searching using groups method,'' *Pattern Recognit.*, vol. 58, pp. 39–48, Oct. 2016.
[49] Y. Wang, F. Jin, H. Su, J. Wang, and G. Zhang, ''Reasearch on user profile based on User2vec,'' in *Proc. Int. Conf. Web Inf. Syst. Appl.*, 2018, pp. 479–487.
[50] N. Mou, R. Yuan, T. Yang, H. Zhang, J. J. Tang, and T. Makkonen, ''Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China,'' *Tourism Manage.*, vol. 76, pp. 103955-1–103955-14, Feb. 2020.
[51] X. Wu, Z. Huang, X. Peng, Y. Chen, and Y. Liu, ''Building a spatially-embedded network of tourism hotspots from geotagged social media data,'' *IEEE Access*, vol. 6, pp. 21945–21955, 2018.
[52] A. Rodriguez and A. Laio, ''Clustering by fast search and find of density peaks,'' *Science*, vol. 344, no. 6191, p. 1492, 2014.

TAO WU received the B.S. degree in resources environment and the management of urban and rural planning from the Wuhan University of Technology, Wuhan, China, the M.S. degree in cartography and geography information system from Central South University, Changsha, China, and the Ph.D. degree in geographic information science from the Joint Doctoral Program between Wuhan University and Central South University. He is currently working as a Postdoctor with the Hunan Key Laboratory of Geospatial Big Data Mining and Application, Changsha. His research interests include geographic information systems, large scale spatio-temporal trajectory data mining, and smart cities.

PENGFEI ZHANG received the B.S. degree in geographical science from Anqing Normal University, in 2017. He is currently pursuing the M.S. degree in cartography and geographic information systems with Hunan Normal University. His research interests include hotness of vehicle trajectory, geographic data mining, and traffic optimization.

JIANXIN QIN received the B.S. degree in geography education from Hunan Normal University, Changsha, China, the M.S. degree in physical geography from Peking University, Beijing, China, and the Ph.D. degree in cartography and geography information system from the Chinese Academy of Science, Beijing, China. He is currently a Professor of cartography and geography information system with Hunan Normal University, Changsha. His research is focused on visualization of spatial information, spatial information systems, and spatial data mining.

DI WU (Member, IEEE) received the Ph.D. degree in computer science from the University of California, Irvine, USA, in 2013. He was a Researcher with the Intel Collaborative Research Institute for Sustainable Connected Cites, a Research Associate with Imperial College London, a Staff Research Associate with the University of California, Irvine, a Visiting Researcher with IBM Research, and a Student Research Associate with the SRI International. He is currently a Professor with Hunan University, China, and an Adjunct Researcher with the University of California, Irvine, USA. His research interests include future networking, intelligent analytics, and smart architecture. He has actively served on many conference committees and is currently an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

LONGGANG XIANG received the B.S. degree in optoelectronic information science and engineering from Wuhan University, Wuhan, China, and the M.S. degree and the Ph.D. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China. He is currently a Professor of geography information science with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan. His research is focused on trajectory data analysis and spatio-temporal data management.

YILIANG WAN received the B.Eng. degree in remote sensing science and technology and the Ph.D. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2009 and 2015, respectively. He is currently working with Hunan Normal University, China. His current research interests include GISci and remote sensing, with focusing on uncertainty modeling and quality control for spatial data, machine learning for volunteered geographic information (VGI), and change detection from high resolution satellite images.