Postprint of Urban Area Accessibility Evaluation Model Based on Multi-source Data

Authors: Shan Xiaochen, Qu Haicheng, Liu Wanjun

Date: 2018-05-02T00:00:00+00:00

Abstract

Urban regional accessibility evaluation has long been a hot topic of considerable interest in the field of intelligent transportation. Traditional regional accessibility evaluation models generally support only single-source data such as GIS or GPS as the basis for accessibility assessment, and cannot avoid the problem of inaccurate evaluation of regional accessibility caused by external factors. To address this issue, we construct an urban regional accessibility evaluation model that supports multi-source data, using multi-dimensional data such as taxi GPS trajectory data, time periods, and weather as the basis for regional accessibility assessment. On this foundation, we design a multi-source data regional accessibility rate calculation method based on multi-dimensional OD matrices, and employ accessibility rate as the quantitative standard for regional accessibility to improve the accuracy of accessibility evaluation. Furthermore, to address the problems of effective information omission and inaccurate data correction caused by overly coarse traditional GPS data cleaning methods, we utilize a sequence data cleaning method based on statistical theory, employing speed and acceleration information from taxi GPS data to correct potential erroneous data and improve the effectiveness of GPS data cleaning. Experimental results demonstrate that the proposed multi-source data urban regional accessibility evaluation model improves the accuracy of accessibility evaluation by 9.1%-37.8%, wherein the accuracy of calculated regional accessibility rate improves by 12.6%-35.5% compared with traditional methods, and the accuracy of average travel time improves by 18.5%-31.6%.

Full Text

Evaluation Model of Urban Area Accessibility Based on Multivariate Data

Shan Xiaochen, Qu Haicheng†, Liu Wanjun
School of Software, Liaoning Technical University, Huludao, Liaoning 125000,

China

Abstract: Urban area accessibility assessment has long been a focal issue in intelligent transportation research. Traditional regional accessibility evaluation models typically rely solely on single-source data such as GIS or GPS, which cannot avoid inaccuracies caused by external factors. To address this limitation, we propose a multivariate data-supported urban area accessibility evaluation model that incorporates multidimensional data including taxi GPS trajectory data, time periods, and weather conditions. We design a multivariate data regional accessibility ratio calculation method based on multidimensional OD matrices, using the accessibility ratio as a quantitative standard for regional accessibility to improve assessment accuracy. Additionally, to address the problems of effective information loss and inaccurate data correction caused by overly coarse traditional GPS data cleaning methods, we employ a statistical theory-based sequential data cleaning approach that utilizes speed and acceleration information from taxi GPS data to correct potential errors and enhance cleaning effectiveness. Experimental results demonstrate that the proposed multivariate data urban area accessibility evaluation model improves assessment accuracy by 9.1%–37.8%, with regional accessibility ratio accuracy increasing by 12.6%–35.5% and average travel time accuracy improving by 18.5%–31.6% compared to traditional methods.

Keywords: GPS; accessibility; accessibility ratio; silhouette method; OD matrix

0 Introduction

As urban populations continue to grow and city areas expand, people have become increasingly concerned about travel efficiency. Many cities have evolved from single-center to multi-center structures, yet road network development has not kept pace with urban modernization. This mismatch creates difficulties in accessing certain areas, and the key to resolving these challenges lies in identifying regions with poor accessibility and analyzing the underlying causes. By improving accessibility in specific areas, overall urban traffic flow can be enhanced. Furthermore, regional accessibility assessment provides decision-making support for traffic management departments in addressing congestion and road construction planning.

Current urban area accessibility assessment typically begins by dividing the city into uniform regions, then clustering travel information to identify hotspots (areas with high travel frequency). Regions containing hotspots are designated as learning areas. The assessment primarily involves mining travel information within these learning areas, calculating average travel times between regions, and finally comparing these times against predetermined thresholds to quantify accessibility.

Existing quantification strategies for accessibility fall into two main categories. Traditional baseline models analyze GIS data to evaluate accessibility based on average travel time between regions. After obtaining average travel times from a starting region to all destination regions, these methods compute the mean of these averages to assess regional accessibility. However, this baseline approach calculates inter-regional travel time by dividing straight-line distance by average speed, which introduces inherent errors. This method only works when travel routes approximate straight lines, but in reality, most routes are longer than the straight-line distance between endpoints, making traditional baseline assessments inevitably inaccurate.

With the proliferation of satellite positioning technology, taxis in major cities are now equipped with GPS systems for dispatch and data analysis. Taxi GPS data can reveal a city' s travel distribution patterns, as taxi drivers typically select the most time-efficient routes based on experience, thereby avoiding route selection bias in accessibility assessment. To reduce errors caused by GIS data' s reliance on straight-line distances, research has gradually shifted toward GPS data-based analysis methods. These approaches quantify accessibility by counting the number of accessible hotspots, with the silhouette method being the most representative algorithm. This method generally assumes that if the average travel time from a learning region to another exceeds the citywide average, all hotspots in the target region are considered inaccessible to the starting region, and vice versa. While this can identify regions with clearly good or poor accessibility, it performs poorly for areas with uneven travel distributions or ambiguous accessibility levels, as each region has varying numbers of connected hotspots and travel times.

To reduce quantification errors, this paper employs regional accessibility ratio as the evaluation metric—defined as the proportion of trips from a region that complete within a threshold time range relative to the total number of trips originating from that region. This approach not only improves upon the traditional silhouette method' s accuracy but also accounts for regional travel distribution characteristics, avoiding erroneous estimates for regions with uneven travel patterns. Moreover, trips extracted from GPS data require no route consideration; calculating the time difference between start and end points yields accurate travel times, overcoming the baseline method' s inability to accurately determine average travel times. Building upon this foundation, our model uses taxi GPS trajectory data, time periods, weather, and other multidimensional data as assessment criteria, constructs a multivariate data-supported urban area accessibility evaluation model, and designs a multivariate data regional accessibility ratio calculation method based on multidimensional OD matrices.

1 Related Work and Research

Accessibility was initially defined as the degree of communication opportunity between regions in a transportation network. Later, Shen et al. approached accessibility from an urban spatial perspective, arguing that it is closely related to citizens' socio-economic activities and their geographic relationships, serving as a measure of these spatial associations. As the concept has evolved, accessibility is now generally understood as the ease with which individuals can reach destinations using certain transportation systems or modes. Regional accessibility is recognized as a method for measuring urban efficiency and is widely applied in transportation research.

Accessibility studies can utilize Geographic Information Systems for prediction, with GIS operating on the premise of searching for shortest paths and weighting the resulting minimum times. Novak et al. used personal information data to calculate accessibility for different time periods, providing more realistic reflections of the relationship between travel routes and times, though this approach requires massive amounts of personal data to compute representative population travel distributions.

With the 普及 of taxi GPS systems, this data has been widely applied in urban planning, geographic research, tourism demand modeling, and travel time estimation due to its large volume and ability to explain inter-regional communication. Laha et al. demonstrated that processing and mining taxi GPS data can yield information on road segment travel times, average speeds, congestion levels, driver route preferences, and passenger pickup/drop-off density, thereby reflecting urban traffic flow patterns. This data helps passengers understand travel conditions and assists drivers in optimizing navigation routes. Additionally, taxi GPS data has been used to design night bus routes by analyzing unidirectional or bidirectional flow patterns.

GPS-based analysis methods are considered more suitable for predicting transportation plan feasibility and evaluating regional accessibility. The core of these approaches involves mining and learning travel times. Accessibility can be quantified using average inter-regional travel times or through constant-speed, distance-weighted methods that calculate travel times and accessible hotspot counts. However, these methods only support single-dimensional GPS coordinate analysis, producing statistically estimated rather than data-mined results, and cannot avoid errors caused by external factors such as time periods and weather. Furthermore, due to the massive scale of raw GPS data and the presence of significant noise, data cleaning is crucial. Most approaches, such as that of Cui J X et al., directly delete erroneous data, which, while filtering noise, destroys data continuity and may discard useful information due to improper filtering thresholds.

2 Regional Accessibility Evaluation Model

This paper constructs a multivariate data-supported urban area accessibility evaluation model using taxi GPS trajectory data, time periods, weather, and other multidimensional data as assessment criteria. We design a multivariate data regional accessibility ratio calculation method based on multidimensional OD matrices, using accessibility ratio as the quantitative standard for regional accessibility. The accessibility ratio represents the proportion of trips from region Pi to other regions that can be completed within a specified travel time.

This metric more objectively and realistically reflects regional accessibility quality. The overall architecture of the accessibility ratio prediction model is shown in [Figure 1: see original paper].

The model consists of three main components: (a) Data preprocessing, including GPS sequence cleaning, complete trip extraction, and integration of multivariate data to generate travel conditions; (b) Regional accessibility ratio calculation, including hotspot generation through clustering, learning region determination, OD matrix generation, and accessibility ratio computation. For irregular travel start and end points, we employ DBSCAN density clustering to form hotspots, connecting seemingly unrelated trips by giving them common attributes in start and end locations, thereby reflecting road accessibility between learning regions. Regions containing at least one hotspot are designated as learning regions. Complete trips provide start and end points, which are mapped to hotspots to determine origin and destination learning regions, enabling the construction of multidimensional OD matrices. Finally, the OD matrices are used to calculate regional accessibility ratios from a trip-based perspective; (c) Regional accessibility evaluation, including computing average accessibility ratios across regions and assessing regional accessibility.

2.1 Taxi GPS Data Filtering and Cleaning

Taxi GPS data is massive but contains errors that can interfere with accessibility ratio prediction. Such erroneous data must be removed or corrected. This paper performs data filtering and cleaning in three aspects:

a) Taxi GPS data provides geographic coordinates. When a GPS data point is erroneous, its coordinates are marked as 0, indicating the point cannot be used or corrected and should be filtered out directly.

b) From GPS data, we can calculate the velocity Vm between adjacent GPS points Gm(Cm,Tm,Sm) and Gm+1(Cm+1,Tm+1,Sm+1) as:

$$V_m = \frac{\sqrt{(x_{m+1} - x_m)^2 + (y_{m+1} - y_m)^2}}{T_{m+1} - T_m}$$

where Cm and Cm+1 represent coordinates. When Vm exceeds the threshold Vmax, the data is considered erroneous and should be corrected. The acceleration am between adjacent temporal sequences can also be derived:

$$a_m = \frac{V_{m+1} - V_m}{T_{m+1} - T_m}$$

When acceleration exceeds threshold M_a, the data is deemed erroneous but can be corrected using statistical methods.

c) Not all drivers choose the shortest optimal route; some select longer routes for profit, directly increasing passenger travel time. To avoid artificially low accessibility ratios caused by detours and multi-destination trips, we use REDUN_Djour as a filtering condition:

$$REDUN\_D_{jour} = \frac{DIS_{real}}{DIS_{ec}}$$

where DIS_ec is the straight-line distance of the trip:

$$DIS_{ec} = \sqrt{(x_n - x_1)^2 + (y_n - y_1)^2}$$

and DIS_real is the actual travel distance:

$$DIS_{real} = \sum_{k=1}^{n-1} \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}$$

Trips with REDUN_Djour exceeding the threshold (3.5) are considered unrepresentative of actual traffic conditions and are filtered out entirely.

2.2 Multivariate Travel Data Structure

Each taxi generates numerous GPS coordinate points daily, containing many trips.

Definition 1. A trip refers to the journey experienced by each taxi passenger, including start and end coordinates, total travel duration, trip date, and travel distance. Such trips can be viewed as sequences of GPS points ordered by time. We define a GPS point structure as G(C,T,S), where a GPS trajectory can be represented as G1(C1,T1,S1) ···Gm(Cm,Tm,Sm) ···Gn(Cn,Tn,Sn). Here, C represents GPS coordinates, T represents time, and S indicates taxi status (0 for empty, 1 for occupied). Taxis generate GPS data at intervals, with a single trip producing n GPS records (subscripts range from 1 to n).

To extract valuable information from each trip, we construct a 4-dimensional OD matrix based on the OD matrix concept. The 4D OD matrix structure is illustrated in [Figure 2: see original paper]. The first and second dimensions represent origin and destination regions, respectively. The city is divided into W×Z regions, with Pi denoting origin region (i 1···W×Z) and Pj denoting destination region (j 1···W×Z). The third dimension represents travel conditions: since road accessibility varies between weekdays and weekends, across different times of day, and under different weather conditions, this dimension includes multivariate travel information such as DAY_T=0 for non-workdays and DAY_T=1 for workdays; TIME_T represents different time periods (e.g., 7:00-9:00, 9:00-12:00, 12:00-16:00, 16:00-19:00, 19:00-next day 7:00); WEATHER=0 indicates rain, WEATHER=1 indicates clear weather. These three conditions (DAY_T, TIME_T, WEATHER) can combine into 20 different travel conditions (e.g., weekday morning with clear weather, weekend evening with rain), denoted as type. The types and their meanings are shown in . The fourth dimension contains specific trip information including travel time, actual distance, and trip date.

TIMEjour represents the total trip duration:

$$TIME_{jour} = T2 - T1$$

where T1 and T2 are the start and end times of a trip, respectively. DAY represents the trip date, serving as a primary key to ensure each trip's uniqueness. The OD matrix can thus be expressed as (Pi, Pj, type, trip).

Algorithm 1 presents the trip information extraction algorithm. It processes GPS data from a single taxi, identifying trip start points when the status bit S transitions from 0 to 1, recording start time, location, weather, and workday information. The algorithm continues scanning subsequent GPS points to determine trip end time, location, and time period when S transitions from 1 to 0.

2.3 Determining Learning Regions

To facilitate urban regional accessibility research, the city is first divided into W×Z grid regions. Complete trips are extracted from GPS data, with their start and end points scattered across various regions. However, not all regions contain valuable travel information—areas like rivers, buildings, and parks lack trip data and are excluded from study. Some regions may contain trips but with insufficient frequency to be meaningful. Therefore, we must identify valuable learning regions from all regions. A region is considered valuable and selectable as a learning region if it contains sufficient trip origins or destinations. With enough pickup/dropoff points, adjacent locations can be considered a hotspot. Any region containing at least one hotspot qualifies as a learning region.

We generate hotspots using clustering methods, where hotspots represent dense travel occurrence locations. While actual trips are defined by start and end coordinates that can appear anywhere, when trip origins and destinations within a small area reach a certain threshold, the area is clustered as a hotspot. Learning regions are determined by checking whether predefined regions contain hotspots. **Algorithm 2** details this process, using DBSCAN density clustering on trip start and end coordinates with search radius E and minimum sample count MINp to identify hotspots. Regions containing non-zero hotspot counts are designated as learning regions and renumbered.

2.4 Establishing Accessibility Ratio Calculation Model

The regional road accessibility ratio measures the completion degree of trips from region Pi to other regions within a specified time range—the proportion of trips that can be completed on time relative to all trips originating from the region. More completed trips indicate higher accessibility ratio and better accessibility. The calculation involves two main parts: constructing the 4D OD matrix and using it to compute regional accessibility ratios.

Algorithm 3 generates the OD matrix by storing extracted trips appropriately for efficient computation. Using trip information and conditions from Algorithm 1 as the third and fourth dimensions, the matrix correctly records trips with origin and destination regions (first and second dimensions) determined by Algorithm 2.

The traditional silhouette method calculates accessible hotspot counts to quantify regional accessibility. It first computes the total number of hotspots in region Pj that can be reached from region Pi under specific travel condition t, denoted as countODij. The total number of trips from Pi to Pj is countODij, obtainable from the OD matrix. The method then calculates the total accessible hotspots from Pi to all regions under all travel conditions:

$$Ac_i = \sum_{t=1}^{20} \sum_{j=1}^{n} a_{ij}^t$$

where a_{ij}^t represents the number of hotspots in region Pj involved under travel condition t. The average travel time from Pi to Pj under condition t is:

$$\overline{TIME}_{ij}^t = \frac{\sum_{k=1}^{countOD_{ij}} TIME_{jour,k}}{countOD_{ij}}$$

The threshold TM_max is defined as the upper limit of acceptable travel time. The traditional silhouette method evaluates regional accessibility using Ac_i, where higher values indicate better accessibility.

However, this approach has limitations. A low average travel time does not guarantee all trips meet the threshold, nor does a high average time mean all trips exceed it. The method' s reliance on hotspot counts can overestimate or underestimate accessibility by ignoring travel distribution characteristics. Some regions may have good accessibility but few involved hotspots due to concentrated travel patterns, leading to underestimation.

Our proposed method replaces hotspot counts with accessibility ratios. For region Pi under travel condition t, the number of trips completed within the specified time range TM_max is:

$$Acp_i^t = \sum_{j=1}^{n} trip_{java}^t$$

where $trip_{java}^t$ counts trips from Pi to Pj under condition t with travel time TM_max. The comprehensive accessibility ratio for region Pi across all conditions is:

$$Ac_i = \frac{\sum_{t=1}^{20} Acp_i^t}{\sum_{t=1}^{20} \sum_{j=1}^{n} countOD_{ij}}$$

This trip-based approach reduces erroneous estimates caused by uneven travel distributions and more objectively represents actual accessibility levels.

For accessibility evaluation, we first compute the citywide average accessibility ratio as a threshold. Regions with ratios below this threshold are deemed poorly accessible, while those above are well accessible. Since our model integrates traffic conditions, travel distance, and weather, it can not only assess accessibility quality but also analyze underlying causes.

The average travel speed Vk for each trip is:

$$V_k = \frac{DIS_{real,k}}{TIME_{jour,k}}$$

The average inter-regional speed Vij is:

$$V_{ij} = \frac{\sum_{k=1}^{countOD_{ij}} V_k}{countOD_{ij}}$$

If average speeds are normal but accessibility is low, long travel distances are likely the cause. If speeds are low, poor road conditions are the primary factor. By comparing accessibility ratios across travel conditions (e.g., clear vs. rainy weather from), we can identify specific causes of congestion and provide targeted improvement recommendations.

3 Case Study

To illustrate the application of our regional accessibility assessment method, we present a case study of Shenyang city using taxi GPS trajectory data and meteorological data from March 1 to May 31, 2016. Precipitation information from weather stations was used to determine clear and rainy conditions.

3.1 Taxi GPS Data Cleaning

The raw dataset contained 1,925,867,139 GPS records, including invalid data requiring cleaning through deletion and correction steps. The data volume after each cleaning step is shown in [Figure 3: see original paper].

The first cleaning step removed coordinates obviously outside Shenyang (e.g., (0,0)). The second step corrected or removed points with unrealistic speeds and accelerations, using Vmax = 150 km/h and M_a = 5 m/s^2. While the first step removed substantial data (mostly from nighttime when taxis are inactive and the database was filled with garbage), the second step removed fewer records but corrected many more due to inherent GPS system errors.

3.2 Determining Learning Regions

Shenyang was divided into a 46×46 grid (2,116 regions). Using DBSCAN clustering on trip start and end coordinates from the OD matrix with search radius E = 0.1 km and minimum samples MINp = 15, we identified 4,369 hotspots ([Figure 4: see original paper]). This yielded 521 learning regions containing at least one hotspot.

3.3 Predicting Learning Region Accessibility Ratios and Analyzing Low-Accessibility Areas

For accessibility ratio calculation, we first determined the upper limit of average travel time TM_max = 28 minutes. We then computed Acp_i^t for each region under various travel conditions. For example, [Figure 5: see original paper] shows region Pi' s accessibility ratios to all other regions under weekday morning clear weather (WMF) conditions.

[Figure 6: see original paper] displays Pi' s average accessibility ratios across all conditions, separated into clear and rainy weather scenarios. From these, we computed comprehensive accessibility ratios for all 521 learning regions ([Figure 7: see original paper]). The citywide average comprehensive accessibility ratio (58.6%) serves as the evaluation benchmark. Regions above this threshold have good accessibility; those below require further analysis.

Most regions exceed the city average, but those below can be analyzed by examining average speeds across the 20 travel conditions grouped into four categories: weekday clear, weekend clear, weekday rainy, and weekend rainy. Using Equations (12) and (13), we compute Vij. Normal speeds with low accessibility suggest excessive distances, while low speeds indicate congestion. Comparing ratios across conditions (e.g., clear vs. rainy) reveals specific causes, such as weather-related road deterioration.

4 Comparative Experiments

To validate the accuracy of our multivariate data-based regional accessibility evaluation model, we compared its results against the widely used baseline model and traditional silhouette method. Experimental data consisted of training data (March–May 2016) for model assessment and test data (June 1–15, 2016) for generating ground truth.

[Figure 8: see original paper] compares average travel times computed by the baseline method and our approach against actual values for trips from one region to 520 others. Our algorithm' s results (black points) closely match actual values (red points), while the baseline method (blue points) consistently underestimates travel times due to its reliance on straight-line distances, leading to poor accessibility assessment accuracy.

The silhouette method uses Ac_i to count accessible hotspots. To enable comparison, we converted hotspot counts to relative accessibility ratios by dividing by total city hotspots. [Figure 9: see original paper] compares the comprehensive accessibility ratios from both methods against ground truth for Shenyang' s 521 learning regions. Our method (black points) closely tracks actual values (red points), while the silhouette method (blue points) shows more dispersed results due to its statistical nature and tendency to overlook travel characteristics.

**** compares the number of regions assessed as having good accessibility by each method. Our method' s assessment most closely matches the true results, with the highest number of correctly identified regions, demonstrating superior accuracy. The baseline method overestimates good-accessibility regions the most while having the lowest correct match count, resulting in the lowest accuracy due to underestimated travel times. The silhouette method underestimates good-accessibility regions and has fewer correct matches because its binary hotspot count approach cannot account for uneven travel distributions.

5 Conclusion

This paper addresses the limitations of traditional urban area accessibility evaluation methods—namely, their inability to support multivariate data analysis and their imprecise quantification approaches. By using taxi GPS trajectory data, travel periods, weather, and other multidimensional data as assessment criteria, we construct a multivariate data-supported urban area accessibility evaluation model. We design a multivariate data regional accessibility ratio calculation method based on multidimensional OD matrices, using accessibility ratio as the quantitative standard to reduce errors from uneven travel distributions and improve accuracy. Additionally, we develop a data cleaning and effective travel information extraction method for massive taxi GPS datasets, providing a more effective data foundation for accessibility assessment.

Comparative experiments demonstrate that our multivariate data-based accessibility evaluation model improves average travel time calculation accuracy by

18.5%-31.6%, regional comprehensive accessibility ratio accuracy by 12.6%-35.5%, and overall regional accessibility assessment accuracy by 9.1%-37.8% compared to existing methods.

Future research can incorporate additional factors affecting traffic conditions, such as road traffic volume, pedestrian flow, and driver behavior patterns, to extract more comprehensive accessibility information. While our accessibility ratio quantification approach represents an improvement, it is not absolutely accurate. Subsequent research can optimize the calculation method under multivariate data conditions to further enhance regional accessibility assessment accuracy.

References

[1] Coutts C J, Horner M, Chapin T. Using GIS to model the effects of green space accessibility on mortality in Florida [R]. Department of Urban & Regional Planning Faculty Publications, 2017.

[2] Cui J X, Liu F, Janssens D, et al. Detecting urban road network accessibility problems using taxi GPS data [J]. Journal of Transport Geography, 2016, 51: [3] Zhang A, Song S, Wang J. Sequential data cleaning: a statistical approach [C]// Proc of International Conference on Management of Data. New York: ACM Press, 2016: 909-924.

[4] Shen Q. Spatial technologies, accessibility, and the social construction of urban space [J]. Computers Environment & Urban Systems, 1998, 22 (5): [5] Stelder D. Regional accessibility trends in europe: road infrastructure, 1957-2012 [J]. Regional Studies, 2016 (6): 1-13.

[6] 潘竟虎, 从忆波. 中国民用机场可达性与服务范围测度 [J]. 经济地理, 2015, 35 (2): 46-53.

[7] 陈艳艳, 魏攀一, 赖见辉, 等. 基于 GIS 的区域公交可达性计算方法 [J]. 交通运输系统工程与信息, 2015, (2): 61-67.

[8] 周爱华, 张景秋, 张远索, 等. GIS 下的北京城区应急避难场所空间布局与可达性研究 [J]. 测绘通报, 2016 (1): 111-114.

[9] Novak D C, Sullivan J L. A link-focused methodology for evaluating accessibility to emergency services [M]. [S. l.]: Elsevier Science Publishers, [10] Qian X, Ukkusuri S V. Spatial variation of the urban taxi ridership using GPS data [J]. Applied Geography, 2015, 59: 31-42.

[11] Phiboonbanakit T, Horanont T. Who will get benefit from the new taxi fare rate?Discerning the real driving from Taxi GPS data [C]// Information and Communication Technology for Embedded Systems. 2016: 73-78.

[12] Lu Y, Li S. An empirical study of with-in day od prediction using taxi GPS data in Singapore [J]. Langmuir the ACS Journal of Surfaces & Colloids, 2014, 30 (31): 9567-9576.

[13] 徐先瑞, 彭仲仁. 基于案例的城市道路行程时间预测 [J]. 交通运输系统工程与信息, 2016, 16 (4): 199-205.

[14] Mustary N R, Chander R P, Baig M N A. A performance evaluation of VANET for intelligent transportation system [J]. World Journal of Science & Technology, 2012. 2 (10) , 89–93.

[15] Laha, A. K, Putatunda. Travel time prediction for taxi-GPS data streams [J]. Iima Working Papers, 2017.

[16] Chen C, Zhang D, Zhou Z H, et al. B-planner: night bus route planning using large-scale taxi GPS traces [C]// Proc of IEEE International Conference on Pervasive Computing and Communications. 2013: 225-233.

[17] Chen C, Zhang D, Li N, et al. B-planner: planning bidirectional night bus routes using large-scale taxi GPS traces [J]. IEEE Trans on Intelligent Transportation Systems, 2014, 15 (4): 1451-1465.

[18] 陈广胜, 程逸群, 景维鹏. 基于 KD 树划分的云计算 DBSCAN 优化算法 [J]. 计算机工程, 2017, 43 (4): 21-27.

Note: Figure translations are in progress. See original paper for figures.