


Research and Application of GPS Trajectory Data Visualization

Li Cai^{1,2}  · Yifan Zhou² · Yu Liang² · Jing He²

Received: 15 October 2017 / Revised: 16 October 2017 / Accepted: 21 October 2017 /
Published online: 17 November 2017
© Springer-Verlag GmbH Germany 2017

Abstract Taxi trajectory data is a kind of massive traffic data with spatial–temporal dimensions, and plays a key role in traffic management, travel analysis and route recommendation for residents. Analyzing trajectory data with traditional methods is complicated, but visualization techniques can intuitively reflect the change trend of spatial–temporal data and facilitate the mining of knowledge and laws in the data. A novel taxi trajectory data visualization and analysis system, TaxiVis, has been designed and developed in this study. This system not only displays the traveling routes of every taxi on the map at the micro-level, dynamically analyzing every taxi’s operating indicators with varying time, but also displays the operating statistics of every taxi company at the macro-level. In addition, the TaxiVis provides route inquiry recommendation functions for users by GLTC algorithm. Implementation of front-end functions of this system are based on Node.js, D3.js and Baidu map, and the trajectory data has been stored in MySQL database. We evaluate TaxiVis with the trajectory dataset collected from 6599 taxis in Kunming. Experimental results show that the system can effectively process and analyze trajectory data, and provide precise data supporting and presentation for the comprehensive evaluation of taxi operation efficiency and mining the drivers’ intelligence.

✉ Li Cai
lcail@fudan.edu.cn
Yifan Zhou
efancn@ynu.edu.cn
Yu Liang
yuliang@ynu.edu.cn
Jing He
hejing@ynu.edu.cn

¹ School of Computer and Science, Fudan University, Shanghai 200433, China

² School of Software, Yunnan University, Kunming 650091, China

Keywords Visualization · GPS trajectory data · Taxi operation efficiency · Route recommendation

1 Introduction

Nowadays many taxis have installed GPS equipment, producing massive trajectory data. The analysis and mining of trajectory data can be widely used in urban planning, intelligent transportation and other fields [1,2]. Trajectory visualization is a hot topic in trajectory data research, and it provides a new method for analyzing and verifying the trajectory data [3]. Using visualization techniques to visually show the spatial–temporal characteristics of vehicle trajectory, we can find the multidimensional relationship existing in the trajectory data, and explore the knowledge and law hidden behind data [4]; as well as discovering the wrong data or outlier data.

Yu et al. [5] mined the historical trajectory data of Beijing taxis and discovered drivers' knowledge and experience on urban roads, implementing a driving routes recommendation system. Zhang et al. [6] established a system for predicting passengers' appeared probability using one-year taxis data in Hangzhou, and recommend a possible position for a taxi driver carrying passenger combining spatio-temporal characteristics of the hot spots. Andris et al. [7] taking the taxis in Shenzhen as an example, built a display model of the urban residents' travel by analyzing the operation behavior of the drivers and revealed the behavior law of taxi drivers. Liu et al. [8] studied the flow feature of the residents based on trajectory data of buses, subway and taxis, and depicted the people's daily flow patterns and change trends.

In summary, the present visualization systems for trajectory analysis mainly aim at high-level applications, but existing the following problems: (1) these visualization systems are mainly high-level application oriented, namely using trajectory data to analyse or mine hidden laws and knowledge, and show results at the macro level. However the system of analyzing every taxi's route and operating efficiency at the micro level is less. (2) Due to the lack of personalized taxi operation analysis system, it is unable to grasp the operational efficiency of every taxi in every day, resulting in taxi companies can only charge fixed contract costs to the drivers, unable to achieve dynamic charges. (3) The existing taxi operating systems generally only provide customers with monitoring and management functions, system interaction is not strong, and the functions are simple.

Therefore, we use data mining and visualization techniques to design and develop a novel taxi trajectory visualization system called TaxiVis, which not only provides accurate data analysis and visual display for every taxi's operation, but also provides route recommendation function based on data mining.

2 The System Architecture of TaxiVis

TaxiVis include five functions: vehicle trajectory generation, multi window switch, vehicle operation statistics, time-axis variation, travel route retrieval. The system architecture is shown in Fig. 1.

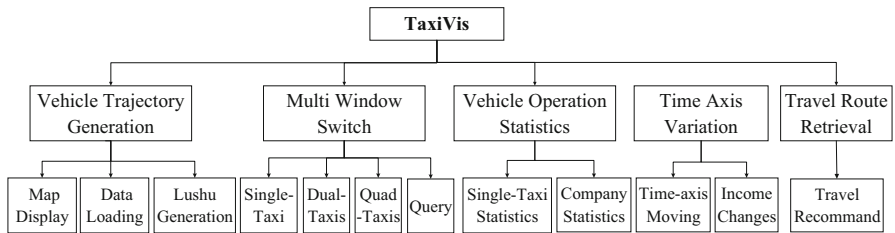


Fig. 1 System architecture of TaxiVis

2.1 Vehicle Trajectory Generation

The vehicle trajectory generation is divided into three sub-modules: Baidu Maps Display, Trajectory Data Loading and Lushu Generation. By setting the central coordinate, TaxiVis can load the area which is provided for user to browse, and the map will be able to arbitrarily enlarge, narrow, or move after loading. The map zoom level is set to 15 to help users to observe the trajectory carefully. Once the map loading has completed, the trajectory data of taxi will start to be loaded, which means the system will transmit the trajectory-point data from database to the map, then connect the trajectory points with a polyline to create the travel route of a taxi.

2.2 Multi-window Switch

In order to allow users to better observe the operation situations of taxis, TaxiVis provides four different modes: single-taxi mode, dual-taxi mode, quadruple-taxi mode and query mode. According to the user's selection, the system can display the operating situations of one, two, or four taxis, meanwhile, showing the retrieval results of travel route in the search mode.

2.3 Vehicle Operation Statistics

In general, the basic operating indicators of taxis consist of the number of moving taxis, the number of passengers, traveling time, unloaded rate, traveling mileage of every day, operating income and so on [9]. The analysis of taxis operating indicators is an important module of TaxiVis, and it is divided into two parts: single-taxi statistics and company statistics. The single-taxi statistics includes current operating income, one-day operating income, current mileage, one-day mileage, one-day traveling time and unloaded rate. The company statistics includes the number of current moving taxis, current operating mileage and gross income of operating taxis. These statistical modules can dynamically analyse the vital indicators of every taxi or every taxi company in an operating time (24 h or longer).

2.4 Time-Axis Variation

The variation of gross income with time is dynamically reflected by this module, it consists of two sub-modules: time-axis moving and income variation. Time-axis moving is used to draw a two-dimensional dynamic axis, in this axis, X-axis showing time while Y-axis showing gross income. When a user select the trajectory data of a taxi to visualize, the time axis will move along X-axis; the value of Y-axis can be changed by the sub-module income variation, meanwhile, the diagram would be drawn by the system.

2.5 Travel Route Retrieval

In query mode, users can apply this function to retrieve the optimized travel route by designating origin and destination in query dialog, then the system will employ the trajectory clustering algorithm named GLTC to query the all travel routes through the origin and the destination and recommend the best travel route from these retrieved routes to users.

3 The Key Techniques of TaxiVis

The following describes the key techniques for data processing and visualization

3.1 Techniques of Data Processing

3.1.1 Map Matching

As the most key procedure of floating car data processing, map matching fixes a sequence of GPS points, which represent the locations of a vehicle, to the correct link of an urban road network and thus determine a vehicle's travelling trajectory [10, 11]. For improving the accuracy of map matching, the closest point with the heading and pathfinding map matching algorithm [12] has been adopted in this paper, this algorithm is a comprehensive algorithm concerning with projection distance, heading and road network accessibility. Its basic principle is: to take the best matching result of previous GPS point as the start, and every point from the matching result sets of the current point as the end, then to perform single-source shortest path between the start and the end. Pathfinding not only can calculate the specific traffic routes, but also can calculate whether a path is reachable within the time interval. After the completion of pathfinding, we use three kinds of weighted information to calculate the optimal matching point from the matching set of current GPS point, the three weights can be described as the following formulas:

Distance weight is expressed as:

$$W_D = Dw * f(D); \quad f(D) = 1 - D/D_{TH} \quad (1)$$

where D_w means distance weight coefficient, D_{TH} is set to 25 meters, and D represents the distance from positioning point to the candidate road segments.

Heading weight is expressed as:

$$W_H = H_w * f(\Delta\theta) \quad (2)$$

where H_w means heading weight coefficient, $f(\Delta\theta) = \cos(\Delta\theta)$, $\Delta\theta$ represents the included angle between vehicle heading and road segment direction.

Accessibility weight is expressed as:

$$W_R = R_w * X \quad (3)$$

where R_w means accessibility weight coefficient, X represents whether the path is accessible. If accessible, X is 1, otherwise X is -1

The weight coefficients of D_w , H_w and R_w can be set to $1/3$ or other ratio, the total weight is:

$$W = W_D + W_H + W_R \quad (4)$$

Finally, according to the sum of the weights, the point with the maximum value is taken as the optimal matching result of the current point.

3.1.2 Data Access and Data Forwarding

The trajectory data and statistical information displayed in the map are stored in the MySQL database. Express, the framework for Node.js, is employed to read the data, the main functions of Express include responding to HTTP requests, defining a routing table for performing different HTTP requests; dynamically rendering HTML pages by passing parameters to the template. In practice, TaxiVis can store the data which can be displayed in the browser in variables, after this, a Json formatted file which is able to be read by `D3.json()` method will be generated and transferred to the browser. While the Json formatted file has been read in the browser successfully, the information containing the plate number, driving time and current income will be displayed in the command line of Node.js.

3.1.3 Trajectory Clustering

Because the trajectory data contains a lot of information, researchers can obtain the moving pattern, identify the hot paths and mine the spatio-temporal characteristics of the residents' travel by clustering analysis. To mine the popular routes from massive trajectory dataset, we propose a novel trajectory clustering algorithm, named GLTC. The GLTC algorithm consists of two stages, the first stage: based on global characteristics, performing a rough clustering for the whole trajectory; the second stage: based on the local characteristics, performing clustering for the trajectory segments; then integrating the results of two stages to complete the entire clustering process. In the GLTC algorithm, global characteristics indicates the distance of two trajectories and

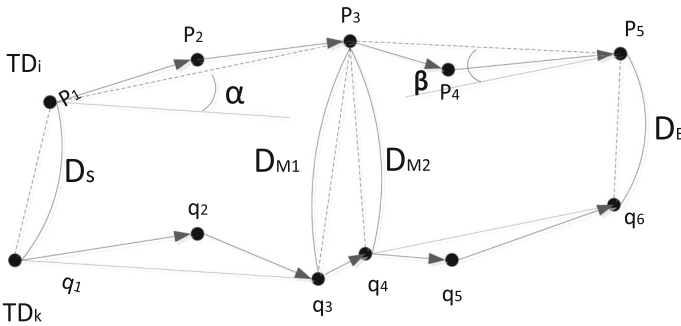


Fig. 2 The relationship between two trajectories

their angle, and local characteristics include two trajectories’ corner, direction and position information.

- Trajectory clustering

In the first stage of GLTC algorithm, we calculate the distance between the two trajectories, and this distance D consist of start-points distance (D_S), middle-points distance (D_M) and end-points distance (D_E). Suppose TD_i and TD_k represent two trajectories, the relationship between them is shown in Fig. 2.

In Fig. 1, the D_S means the distance of point p_1 and point q_1 , the D_E represents the distance of point p_5 and point q_6 . The calculation of the D_M is divided into two cases, as shown in Eq. 5. The distance D can be calculated by Eq. 6.

$$D_M = \begin{cases} D_{M1} & TD_i \text{ and } TD_k \text{ both have odd number of points} \\ \frac{1}{2}D_{M1} + \frac{1}{2}D_{M2} & TD_i \text{ or } TD_k \text{ has even number of points} \end{cases} \quad (5)$$

$$D(TD_i, TD_k) = W_S D_S + W_E D_E + W_M d(TD_{ic}, TD_{kc}) \quad (6)$$

In Eq. 6, W_S , W_M and W_E represent the corresponding weight of three distances, the values of W_S and W_E can be calculated with parameters α and β . The parameter α is the angle between two start-midpoint dashed line segments of TD_i and TD_k , and β is the angle between two midpoint-endpoint dashed line segments. The values of W_M is set to 1/3.

These two distances D and D_S are used as the standard to measure the similarity between the trajectories to achieve trajectory clustering. Two parameters S_{Thred} and D_{Thred} are represented respectively the thresholds of D_S and D .

- Trajectory segments clustering

The second stage of GLTC algorithm is based on the results of first stage, adding the local characteristics (corner, direction and position), which are shown in Fig. 3, using DBSCAN clustering method to complete clustering for trajectory segments. We first calculate the trajectory corner, which is used to divide the trajectory segments, and combined with the local features of the trajectory to calculate the structural distance between the trajectory segments, then this distance is used as the standard to measure the similarity between the trajectory segments to achieve trajectory clustering.

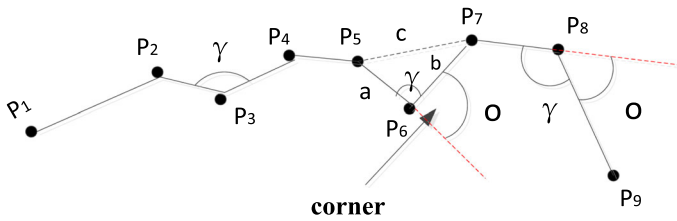


Fig. 3 The corner and included angle of trajectories

Trajectory corner (o) between adjacent points in the trajectory reflects the movement of the trajectory. We can calculate the corner o by the angle γ . Direction (Dir) represents the degree of deviation of the trajectory segments L_i and L_j belonging to the different trajectories in the moving direction. Corner Angle ($Angle$) represents the change in direction and the degree of fluctuation within the trajectory segments. Position (Loc) indicates the positional distance of the trajectory segments L_i and L_j , which uses the Hausdorff distance.

The notation $Ssim$ represents the similarity between the trajectory segments, and it is measured by calculating the structural distance ($SDist(L_i, L_j)$). $Ssim$ can be calculated with Eqs. 7 and 8, where W_D , W_A and W_L correspond to the weights of the local characteristics, and each weight value is greater than or equal to zero and the sum is equal to one. Finally, we do the normalized ($Norm()$) of the structural distance to convenience comparison.

$$SDist(L_i, L_j) = W_D Dir + W_A Angle + W_L Loc \tag{7}$$

$$Ssim = 1 - Norm(SDist(L_i, L_j)) \tag{8}$$

With the above description, GLTC algorithm takes into account the global and local characteristics of the trajectory, so it improves the reliability and credibility of the clustering results.

3.2 Visualization Techniques

3.2.1 Vehicle Travel Route Generation

For TaxiVis, it is an important visual function to visually show the taxi traveling route in the road. Due to the straight line between two GPS points cannot truly reflect the vehicle’s trajectory, it is necessary to find the path between the two points, and obtain the points array which can reflect the road information. We design one method for acquisition of points array by providing several single-source shortest path algorithms (e.g. Dijkstra algorithm, A* algorithm and branch and bound algorithm). Figure 4 shows the execution result by this method combined with the Lushu function [13]. Point A indicates the origin point while point B is the destination point, and the taxi travels from point A to point B.



Fig. 4 A taxi travels along the route

3.2.2 Window Split

TaxiVis uses window split technique to provide users with greater information. Except for the single-taxi mode, TaxiVis has increased the dual-taxi mode and quad-taxi mode. Using these two modes, the manager can compare different taxi's operating efficiency belonging to the same taxi company, or different taxi companies's operating situations. The window split is based on an iframe tag in HTML that implements the creation of multiple inline frames in a main page, and multiple subpages are contained in the main page.

3.2.3 Time Axis Variation

The time axis is used to dynamically reflect the relationship between time and income. By using the Canvas tab in HTML5, animations can be flexibly implemented in Web pages. Using the Canvas tab to draw the X and Y-axis in the visual interface, the Canvas will generate the corresponding graphs based on the different X and Y values. The movement of the time axis is achieved by changing the value of "margin-left" in real time. As shown in Fig. 5, the distance between the time axis and the left edge

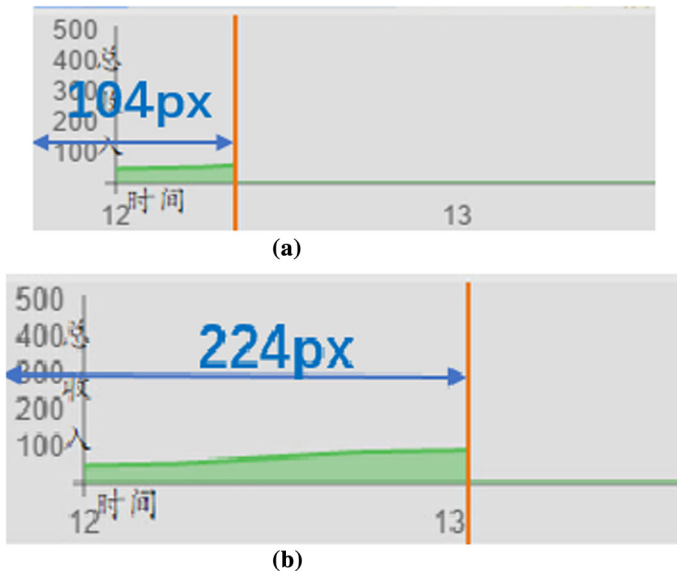


Fig. 5 Distance of time axis. **a** Distance 104 px. **b** Distance 224 px

of page is 104 px, and the distance value is increased by the situation that time-axis changes with X-axis, judging from this, in (b), when the time axis is about $X = 13$, the distance is increased to 224 px.

4 TaxiVis System Implementation

To develop TaxiVis, we use HTML5 and JavaScript language, and Node.js [14] is used for web server while Mysql is used for database, meanwhile, using Baidu Map JavaScript API v2.0 [15] to display the map. D3.js is a JavaScript class library based on the web environment that uses data-driven methods to produce data visualization. TaxiVis uses Node.js to build the server, provides the support of database by the JavaScript, and controls the operation with Express framework in Node.js. Due to Node.js has only provided data forwarding function, unable to read data for the browser, therefore, we use a data access mode named Node.js and D3.js [16].

In order to verify the visualization analysis of TaxiVis, the GPS trajectory data provided by Kunming Vehicle Management Office is used to be the experimental dataset in this paper. The dataset contains the trajectory data generated by 6599 taxis in a week of August 2012 (August 13th–August 19th).

5 System Presentation

TaxiVis provides three different operating modes: single-taxi display mode, dual-taxis display mode and quadruple-taxis display mode, meanwhile, five widgets are contained in every mode: map, data/date, function button, time axis and statistical



Fig. 6 Single-taxi display mode

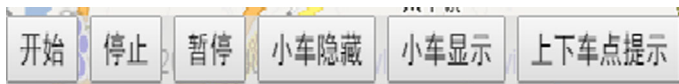


Fig. 7 Button area of TaxiVis

information. In addition, the query display mode is used to show the results of trajectory clustering, and provides the optimal route for user. The following describes the difference between these display modes.

5.1 Single-Taxi Display Mode

The single-taxi display mode is used to show the operation of a taxi, as shown in Fig. 6.

In Fig. 6, the attributes (License_id, date, etc.) can be selected and dynamically displayed at the area “Data/Date”. “Button area A” is used to control a taxi’s running menu, such as move, stop, pause, hide, display and so on, as shown in Fig. 7. “Button area B” provides the switch function to other three modes, and clicking the “Magnifier” button can switch to the query mode while clicking the “LCD display” button can switch to the other two display modes. The time axis widget shows the current changes with the gross income, and the X-axis represents data time and the Y-axis represents gross income. The widget “Statistics” can display the current income, the number of passengers and mileages. Once the taxi has arrived at the destination points, this system will compute and display the result with cumulative calculation. At the same time, after 24 h, the system will calculate the unloaded ratio and analyse operational efficiency of every taxi.



Fig. 8 Dual-taxis mode

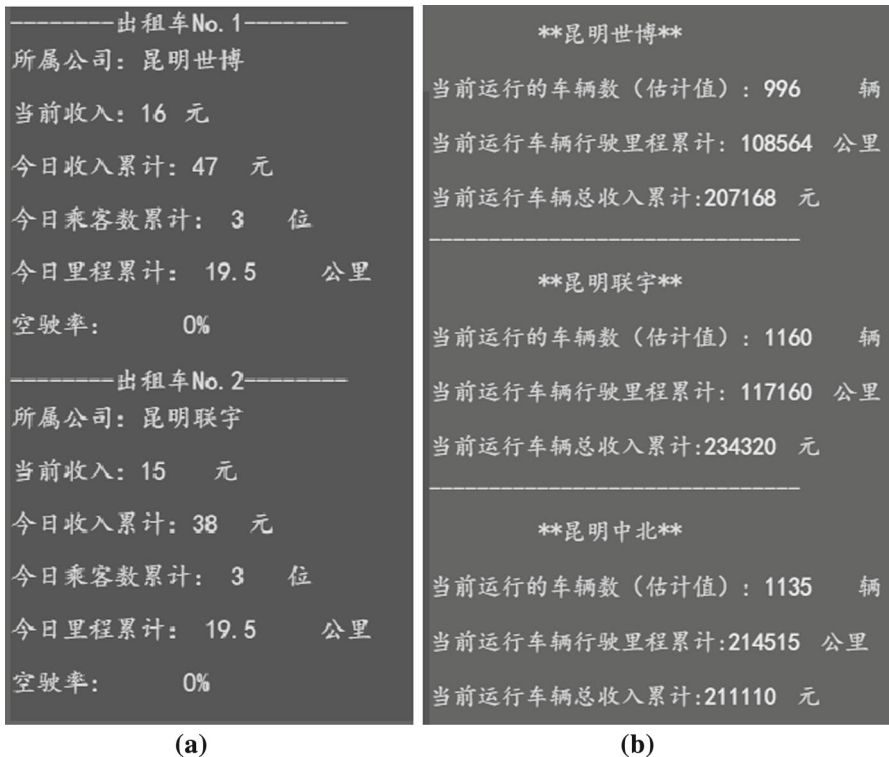
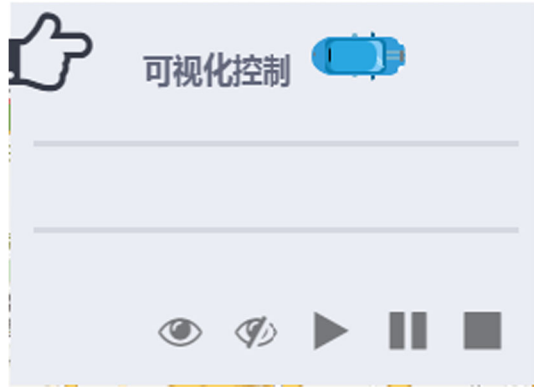


Fig. 9 Company information and operating statistics

Fig. 10 Visual control menu**Fig. 11** Data/Date

5.2 Dual-Taxis Mode

Dual-taxi mode shows the operation of two taxis, as shown in Fig. 8

The function of dual-taxi mode is similar to the single-taxi mode, and the difference is that it uses left and right split-screen to show two moving taxis. Moreover, Clicking on the “Menu” button in the upper left corner shows the statistics of the two taxis on the left side of the screen, as shown in Fig. 9a, at the same time, we can calculate the gross income of different taxi companies, as shown in Fig. 9b. By clicking the “Control” button, we could see the “visual control menu”, as shown in Fig. 8. The function of this menu is similar to the button area in Fig. 10. When the user click on the play button, the two taxis will move at the same time and the statistical information will be updated dynamically.

“Data/Date” is displayed in the bottom left corner of every interface, before every taxi runs, you can select the date, and by clicking the clock button to set and view the time, as shown in Fig. 11.



Fig. 12 Quadruple-taxi mode



Fig. 13 Query mode

5.3 Quadruple-Taxis Mode

The function of quadruple-taxi mode is similar to the single-taxi mode and dual-taxi mode, but there are four interfaces in this mode, the different taxi can be displayed in every interface while the statistics of companies and four taxis can be shown in Fig. 12.

5.4 Query Mode

The results of the trajectory clustering algorithm are used in query mode. When one user inputs origin and destination, GLTC algorithm has been invoked to execute trajectory clustering mining, then an optimal travel route (the shortest distance) selected from the clustering results is recommended to the user. An example of travel recommendation is shown in the Fig. 13. Users who in the single-taxi, dual-taxis, or quadruple-taxi mode are access to query mode by click the “Magnifier” button.

6 Conclusions

The TaxiVis system developed in this study can effectively solve the problem of lack of personalized analysis and fine-granularity display for vehicle operation information in the existing taxi trajectory visualization systems, and display the mined trajectory data. This system can intuitively display the traveling trajectories of one or multiple taxis on the map, and obtain the association and differences among operation data such as traveling time, space, traveling distance, expense, passenger numbers and companies’ gross income. Furthermore, this system can also apply the drivers’ intelligence in planning routes by the trajectory clustering algorithm to the functions of traveling route retrieval and recommendation, which providing quicker and more efficient trip service for users or novice drivers. The TaxiVis’s user interface is friendly and easy to operation, and provides data support and decision support for operation management professionals through visual analysis.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant No. 61663047 and No. 61363084.

References

1. Vandenbergh W, Vanhauwaert E, Verbrugge S et al (2012) Feasibility of expanding traffic monitoring systems with floating car data technology. *Intell Transp Syst* 6(4):347–354
2. Tseng P-J, Hung C-C, Chang T-H et al (2013) Real-time urban traffic sensing with GPS equipped Probe Vehicles. In: *Proceedings of the 2011 international conference on ITS telecommunications*, Taiwan, pp 306–310
3. Lei R, Yi D, Shuai M et al (2014) Visual analytics towards big data. *J Softw* 25(9):1909–1936
4. Jiansu P, Huamin Q, Mingxuan N (2012) Visualization of motion trail. *J Comput Aided Des Comput Graph* 24(10):1273–1282
5. Yu Z (2015) Introduction to urban computing. *Geomat Inf Sci Wuhan Univ* 40(1):1–12
6. Li B, Zhang D, Sun L et al (2011) Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: *2011 IEEE international conference on pervasive computing and communications workshops (PERCOM workshops)*. IEEE, pp 63–68
7. Liu L, Andris C, Biderman A et al (2009) Revealing taxi driver’s mobility intelligence through his trace. *IEEE Pervasive Comput* 160:1–17
8. Liu L, Biderman A, Ratti C (2009) Urban mobility landscape: real time monitoring of urban mobility patterns. In: *11th International conference on computers in urban planning and urban management (Landscape and Urban Planning)*, vol 106, no 1. Hong Kong, data in Shanghai, 16–18 June, pp 73–87
9. Lin Q (2013) The research on characteristic of taxi-traffic operation based on GPS data. *Harbin Inst Technol* 7/8:68–71

10. Quddus MA (2006) High integrity map matching algorithms for advanced transport telematics applications. Imperial College, London
11. Cai L, Zhu B (2013) Research on map matching algorithm based on nine-rectangle grid. In: Proceedings of the 2013 international symposium on computer architecture, pp 363–369
12. Cai L, Zhu B, Zhou S, Luo Y (2016) Interactive map matching and its visualization: techniques and system. In: Proceedings of the 4th international conference on CCF big data, Lanzhou, Oct 11–13
13. The JavaScript Open Source Library of Baidu-Map. The Example of LuShu[EB/OL]. (2015-11-19)[2016-04-20]. <http://api.map.baidu.com/library/LuShu/1.2/examples/LuShu.html>
14. Yu Z (2013) A distributed data-stream service system built by Node.js. *Comput Syst Appl* 7/8:68–71
15. Baidu-Map Open Platform [EB/OL]. (2010-04-23) [2016-04-20]. <http://developer.baidu.com/map/reference/>
16. Yunliang Z, Zhaofeng Z, Xiaodan Z (2013) The dynamic interaction and visualized function realization of knowledge organization systems by D3js. *New Technol Libr Inf Serv* 22(2):127–131