How to Tour Shanghai Disneyland Park

Yan Liu[1,2], Chi Guo[1], Manli Zhang[2], Xiaoqun Wu[3], Ye Zhou[4]

[1] Nation Engineering Research Center for Satellite Positioning System, Wuhan University, China
[2] School of Computer, Wuhan University, China
[3] School of Mathematics, Wuhan University, China
[4] Wuhan Xuezhilv International Travel Service Co., Ltd, China
{2012301500058, guochi, 2012301500040, xqwu}@whu.edu.cn, zhouye@163.com

Abstract

We prensent a tourism management system, based on indoor and outerdoor collaborative precise positioning (Xihe positioning technology) in this paper. The system consists of a kind of tourists card of high precision positioning and a background data service system. By using the system, we can get tourists' location data including real-time data and historical data. We can also provide a more convenient mobile management service for tourists. Then, we put forward the concept of coverage ratio and degree of fatigue, and design algorithms for tour quality analyses and a route planning algorithm. These algorithms involve collected track data and waiting time in the travel spots and aim to analyze whether tourists have an effective tour and recommend reasonable routes. Moreover, we apply the tourism data of Shanghai Disneyland Park as a sample set to validate the algorithms. According to the above algorithms, we get two routes to satisfy tourists. They can travel most tourist spots with minimum time.

Keywords: Xihe high precision positioning, Tourism data analysis, Coverage ratio analysis, Degree of fatigue analysis, Route plan

1 Introduction

As the tourism consumption market changes and the new generation of communication technology applies in the tourism industry, the idea of collective smart tourism is also becoming popular, focusing on "the new generation of information technology". The collective smart tourism that people concerned mainly involves two aspects: one is how to use the Internet, cloud computing, sensors and other technologies comprehensively to design an information system to serve tourism; the other is how to use data analysis to support the actual decision.

The traditional group travel, the primary and middle school students in particular, are facing the following problems when they are traveling:

(1) Management convenience and tourists' safety. The traditional way of group travel is that a number of tourists follow one tour guide. Therefore, it is difficult for the tour guide to manage all tourists effectively.

(2) Analysis of tour behaviors. In the traditional way, it is not common for researchers to conduct an analysis of the positioning data of tourists in traveling. They put more attention to the navigation and real-time service. Thus, tourists cannot be fully satisfied without analysis of tour behaviors.

In compared to the traditional smart travel system, our system uses Xihe precise positioning based on LBS. GPS is popular in our life. However, our system can provide more accurate positioning data than that obtained by GPS. Based on this system, we can use the real-time positioning data to provide tourists navigation. Moreover, we utilize several algorithms to conduct a more detailed analysis about tourists' information such as their behaviors and travel habits via the historical tracking data with high accuracy.

The contributions of this paper are as follows:

(1) We design a smart tourism system based on Xihe high precision positioning. It can supply accurate positioning data by terminal (a kind of tourist card);

(2) We have provided a series of tourism services after getting the precise location of the tourists through the system and put forward tourism quality analysis algorithm and route planning algorithm;

(3) Combine this system with the summer camp of elementary and middle school students of Wuhan, central China's Hubei province and taking the location data of tourists in Shanghai Disneyland Park from June 23, 2016 to July 25, 2016 as an example, we analyze the parameters such as the tourist coverage ratio, the degree of fatigue to measure the tour quality and design an optimal route for tourists with the minimum waiting time.

2 Related Work

Under the background of "tourism + Internet", the

*Corresponding Author: Chi Guo; E-mail:guochi@whu.edu.cn
DOI: 10.3966/160792642019012001003

traditional tourism informatization has been difficult to meet the changing needs of tourists and to process the massive data, thus the smart tourism came into being. In this field, Li Yingfei [1], taking eastern China's Jiangsu province as an example, uses comprehensive evaluation method, spatial econometric model and other empirical research methods to study its smart tourism development and tourism public service system quantitatively. There have been plentiful researches about Beidou high precision positioning. For example, Zhu Feng et al put forward a system scheme [2], consisting of three parts: ground-based augmentation system (GBAS), FM subcarrier broadcast system and terminal modules, which broadcast pseudorange differential data produced by GBAS through FM subcarrier, and discussed the application prospects of the technology scheme in intelligent transportation system (ITS). At the same time, Lin Chao et al also proposed the Beidou high precision positioning market enjoys a bright prospect [3]. They analyzed the application advantages of the Beidou high precision measurement receiver in the field of displacement monitoring, and put forward that the relevant special security areas in China should get rid of the situation that GPS is the primary position system.

About smart travel, Hung gave a Smart-Travel System (STS) to aid traveler with personal requirements to tour by smart phone [4]. The STS provides personalised travelling real-time information, and automatically tells traveler when they show up around the task. Also, Hung et al introduced a new ubiquitous tourism system based on SNS, IoT, and UGC [5], which is designed to develop the tourism industries, and designed a new way to look up the travel information according to the cloud-based service's needs. Sabatucci et al reported an experience of developing a smart travel system by adopting MUSA, a Middleware for User-driven Service Adaptation [6]. The key characteristics of such a system are the ability of self-configuring a set of heterogeneous services and self-adapting to unexpected circumstances.

As smart tourism is based on big data, there are many researchers conducted related researches and gained important useful information in the field of historical track big data. For example, Wang Xiaoming, Zhang Hong et al. have conducted researches in regard to the application of big data of taxi GPS track in ITS [7]. They firstly collected large-scale and all-weather driving data through vehicles installed GPS. Then combining with the fastness, real-time and predictability quality of big data and cloud computing, they conducted real-time analysis and processing, statistical learning and deep mining for GPS tracking data. Finally, they analyzed the development situation of ITS driven by GPS tracking data from three aspects: traffic state analysis, operation management & support and path planning and forecasting. Based on the data analysis of the historical track of vehicle GPS, Zhang Qing proposed a data processing framework system in ITS in order to provide a reference to construct big data processing system in ITS [8]. He established a distributed data warehouse by Hadoop to store and manage GPS data of each vehicle, and to eliminate the redundancy and error data in the process of data processing. He also proposed four kinds of GPS historical track for optimizing vehicle dispatch. Based on the vehicle track data mining technology, Xu Ming conducted in-depth study upon a series of issues such as map matching of the low sampling rate tracking data [9], road network key node identification, road network abnormal events detection and regional connectivity evaluation. All of the above are the analyses on vehicle track. However, Kimm presented an improved algorithm that traces a track path in real time using GPS and Accelerometer data [10]. He collect accurate GPS position data and obtains acceptable GPS track paths by applying 3-degree accelerometer data. S Karagiorgou et al. proposed an automatic road network generation algorithm [11], which took the vehicle track data as input and generated a road network. After analyzing the historical track data of vehicle, JA Downs et al. extended chronological and spatial applications [12], and mined a potential path tree, namely, the driving behavior of vehicle. In the course of the study, they used TGDE (time-geographic density estimation) method to map and quantify the most possible route, starting point, intermediate stop and termination of the vehicle. Based on GPS and Intelligent Transportation Systems, Lo S C et al. investigate different routing schemes [13], and propose a series of data forwarding strategies by predicting the positions of neighboring and destination nodes.

We know from the above analyses that the location mode is simple and the accuracy is not high enough in the field of smart tourism. The researches in the field of high precision positioning only relate to high precision positioning but no combination with smart tourism. The system used in this paper is based on Beidou high precision positioning. Therefore, the accuracy of the track data collected through this system is very high. Combining with the idea of smart tourism, we conducted related data analysis and mining.

3 System Framework

The whole system mainly consists of the Xihe System, the Terminal and the Service, as shown in Figure 1.

3.1 Xihe System

"Xihe", an extension of the BDS navigation, is a positioning service system can provide full spatial and temporal domain seamless positioning service and

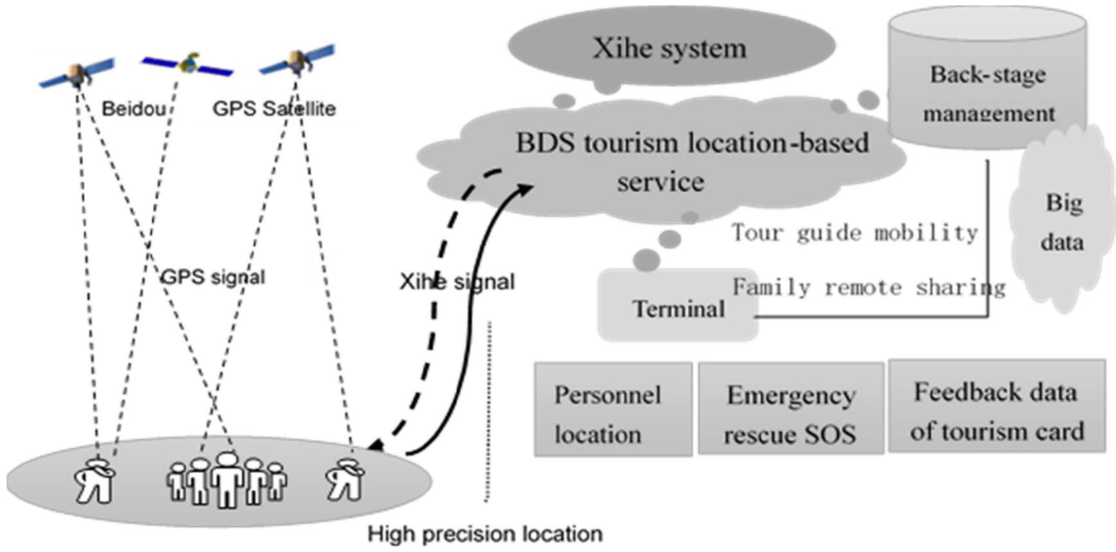


Figure 1. System architecture diagram terminal, combining GPS with Xihe signal form a high precision position

mainly consists of "Beidou ground-based augmentation system". Beidou ground-based augmentation system, dominated by the Beidou satellite navigation system under the unified planning and construction of China, is compatible with other ground-based augmentation system of GNSS system. Its ground reference station distance is 50-300km. It broadcasts navigation signal correction and assistant orientation signals through the ground communication system to provide sub-meter even centimeter precision navigation and positioning service for users and assisted enhanced service for common user terminals. Based on the wide area differential and precise point positioning technology, Beidou wide area and real-time precise positioning technology, integrated with advanced technologies such as real-time data processing, Internet and satellite communication can realize satellite navigation augmentation with the positioning accuracy less than 1 meter in the land, sea and air of China. Wide area and real-time precision positioning system focuses on improving the BDS/GPS positioning accuracy of users in China. It becomes the high precision navigation and integrated information service system in China by upgrading BDS satellite navigation augmentation system.

3.2 Terminal

Taking the smart tourists card as a carrier, the terminal uses the Beidou satellite navigation system (BDS), GPS of the United States, GLONASS of Russia and GALILEO of the European Union. It mainly focuses on Beidou/GPS dual-mode positioning, supplemented by base station/WiFi positioning to achieve seamless positioning. Its appearance shows in Figure 2. Each student card has a unique imei number to identify.

Tourists card terminal, a carrier to obtain location data, can upload its real-time location to the server through the built-in positioning function, making the mobile phone terminal and web terminal management system can share the location thus to view the location and historical track records of each terminal in real-time. Its button function description shows in Table 1.

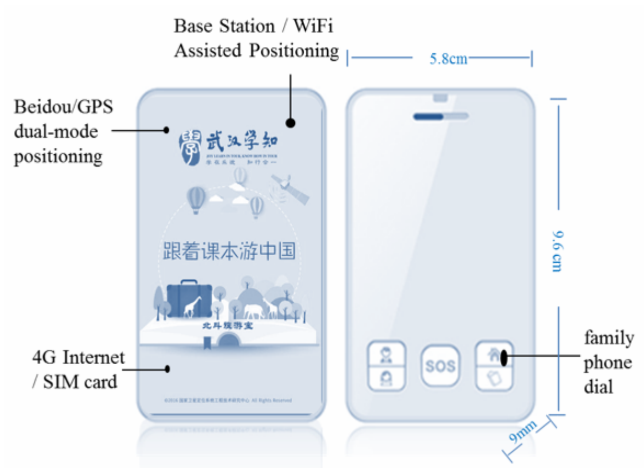


Figure 2. Introduction of student ID card terminal appearance

Table 1. Button function description

Button	Function description
Familiarity number button	Long press this button for more than 3 seconds can dial the corresponding familiarity number
SOS button	Long press this button for more than 3 seconds to switch on in shutdown mode Press this button twice continuously in power-on mode can switch silent mode or ring mode Long press this button for more than 6 seconds can trigger the emergency alarm, sending a message to the family number and calling the family number circularly
Reset button	Press the reset button, reset the phone and the phone will be power off, then press the power button to restart

3.3 Service Description

Service supported by the terminal is listed in Table 2.

Table 2. Service supported

Function serial number	Introduction
①Real-time positioning	Select a tourist and click. You can view the location of the tourist's tourists card, its battery consumption and other information
②Historical track	Select a tourist. You can view the tourist's treading track of any period
③Electronic fence	Set scope for the tourists who have student ID card. When the tourist who holds student ID card walks out of the specified scope, the background will give a warning
④Telephone call	Managerial personnel can call the tourist and speak to him/her directly
⑤Remote monitoring	Use network and SMS to monitor, you can monitor the situation nearby
⑥administrator	Operate the tourists card such as setup basic information for the tourists card

4 Analysis and Algorithm of Tourism Data

Based on the data provided by the tourism management system, this chapter presents the analysis and algorithms of the data, that is, the algorithm to analyze the tour quality and the algorithm to design the optimal route. The tour quality is a quantitative description of the relationship between the number of tourist attractions one tours and the amount of time one takes. That is to say, high tour quality refers to a tourist tours more attractions in a relatively short time and enjoys himself/herself. On the contrary, the tour quality is low. This paper analyzes the tour quality from two aspects: the tour coverage ratio and degree of fatigue.

4.1 Coverage Ratio Analysis

4.1.1 Definition

Definition 1. Circle an area: to circle a closed quadrilateral on the map which contains all the attractions to be studied.

Definition 2. Grid: the vertex of the irregular quadrilateral are A, B, C and D. Regularize the quadrilateral into a rectangular A'B'C'D'. Divide the opposite sides of the regular rectangular A'B'C'D' into n equal parts and link the points of the opposite sides together, thus the regular rectangular is divided into equal n*n small rectangular grids.

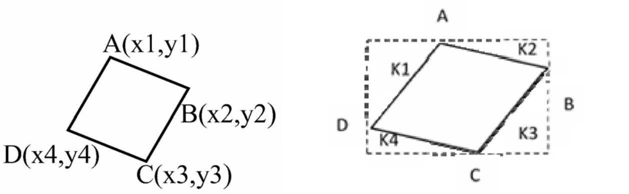
Definition 3. Coverage ratio: it refers to the percentage that the number of toured grid filled by the real-time uploaded location of the tourists accounts for the whole grids of all the scenic spots. If a tourist's position data in a travel fills grids of s, and there are n grids, the coverage ratio is s/n.

Definition 4. The section ratio: the coverage ratio is divided into ten equal intervals and the span of each interval is 10%. For example, 0% to 10% is the first interval, and 10 to 20% is the next one and so on. The section ratio refers to the percentage of tourists in each interval divided by the total number of people counted.

Definition 5. Cumulative coverage ratio: divide the coverage ratio into ten sections: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, that every percentage means the minimum coverage ratio of each section. Then count the tourist percentage that exceeds the coverage ratio of each section.

4.1.2 Coverage Ratio Algorithm

Landscape division model as shown in Figure 3, algorithm one and algorithm two:



(a) Irregular quadrangle area of the park

(b) Regular rectangular area

Figure 3. Landscape division model

- Circle an area:

Circle all the scenic spots to be studied on the map manually, then get four geographic coordinates, as shown in subgraph (a) of Figure 3, x says longitude, y says latitude. Four coordinates gather the scenic area into an irregular quadrangle ABCD. It is difficult to calculate the data of the irregular quadrangle ABCD, thus it is processed to a regular rectangular A'B'C'D'. Take the maximum longitude value minlng and the maximum latitude value minlat of the four vertex of the irregular quadrangle ABCD as the upper-lift corner coordinates of A’ of the regular rectangular A’B’C’D’, minlng=min{x1,x2,x3,x4},minlat=min{y1,y2,y3,y4}.

Take the maximum longitude value minlng and the maximum latitude value minlat of the four vertex of the irregular quadrangle ABCD as the lower right corner coordinates of the C’ of the regular rectangular A’B’C’D’,

maxlng=max{x1,x2,x3,x4},maxlat=max{y1,y2,y3,y4}.

Thus to get, B (maxlng, minlat), D (minlng, maxlat).

- Divide the area into several grids and remove excess blank

In order to divide the attractions more precisely, divide them into grid and remove excess blank. As the rectangle contains four blank areas outside the area, it will lower the accuracy if use the rectangle directly. Thus, it is necessary to remove the blank area. As shown in the algorithm one, the remaining grids has covered the whole attractions to the greatest extent after removing the blank areas.

Algorithm one: The algorithm to remove extra blank of regular rectangular A’B’C’D’

```
Input:grid(a,b,c,d),
A(x1,y1),B(x2,y2),C(x3,y3),D(x4,y4)
Output:grid(a,b,c,d)
k1=(y1-y4)/(x1-x4),b1=y1-k1*x1
k2=(y1-y2)/(x1-x2),b2=y2-k2*x2
k3=(y2-y3)/(x2-x3),b3=y3-k3*x3
k4=(y3-y4)/(x3-x4),b4=y4-k4*x4
FOR() 400
ay' = k1*ax + b1;
by' = k2*bx + b2;
cy' = k3*cx + b3;
dy' = k4*dx + b4;
IF (ay<ay'&&ay<by'&&ay<cy'&&ay<dy')||
   (by<ay'&&by<by'&&by<cy'&&by<dy')||
   (cy<ay'&&cy<by'&&cy<cy'&&cy<dy')||
   (dy<ay'&&dy<by'&&dy<cy'&&dy<dy')
grid is in the scenic area, store in mongodb
ELSE grid is not in the scenic area, give up the grid
END IF
END FOR
```

grid = {a, b, c, d} is a grid before calculation. a,b,c,d are the four vertex coordinates of the grid. grid(a,b,c,d) is a grid in scenic spot after calculation. k1,k2,k3,k4 are slopes of the four sides of the irregular quadrangle ABCD

- To give weight to the grid in order to increase the coverage ratio of attractions with high tour frequency

The popularity of each attraction varies thus the number of tours to each attractions varies, too. If the weight of each grid is equal, it will lower the accuracy of the data. Method description as described in algorithm two:

Algorithm two: Grid weight algorithm

```
Input: HistoryLocation
Output: GridWeight
FOR (,) n*n
FOR(,) HistoryLocation.size()
IF(HistoryLocation[i] in Grid[j])
GridWeight[j] += 1
END for
FOR(,) n*n
Grid Weight[k] /= HistoryLocation. size()
Store Grid Weight[k] in mongodb
END FOR
```

Grid Weight= {gridWeight1, gridWeight2, ..., gridWeightn*n}, weight of n*n grids

HistoryLocation = {{lng1, lat1}, {lng2, lat2}... {lngn, latn}} the position coordinates of all the tourists cards

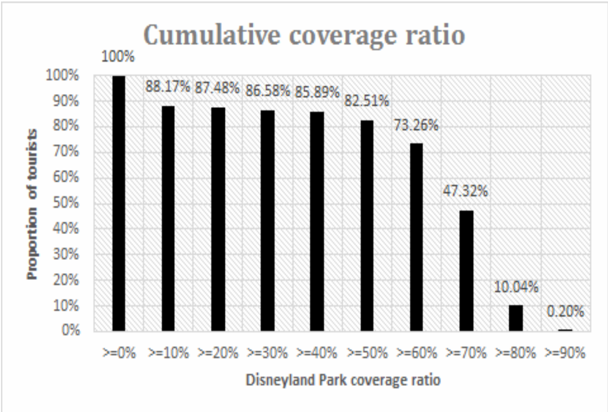
- Tour coverage ratio

Let the location of historical tracking data of each tourist in a day to traverse all the grids, if the location of a tourist falls within the grid, plus 1 in count, if the location of that tourist falls within the grid once again, plus 0 in count. We get the coverage ratio of that tourist = count/number of grid *100%. We calculate the coverage ratio according to the time point, thus we can obtain the coverage ratio of each tourist at each time point. We calculate the coverage ratio in accordance with this method for all the tourists and we can obtain the time-ordered coverage ratio of all the tourists. The coverage ratio at the last time point is the maximum coverage ratio of the tourist in a day. We conduct analysis according to tour data from Shanghai Disneyland Park based on AutoNavi Maps.

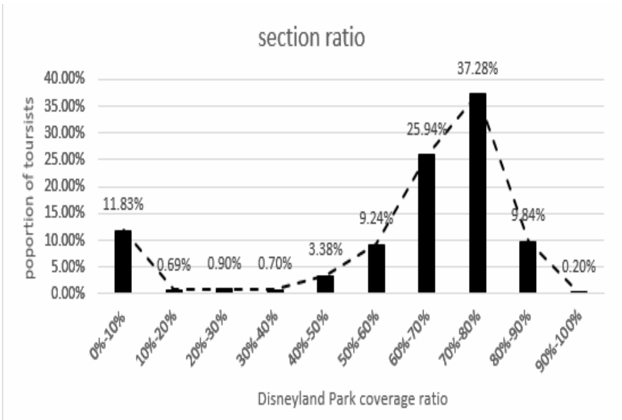
4.1.3 Verification of Coverage Ratio Algorithm

In order to verify the validity of the algorithms, we conducted analysis by taking Shanghai Disneyland Park as an example based on the AutoNavi Maps:

According to the method of circling, we circled the park into irregular quadrangle on AutoNavi Maps, and then processed them into regular rectangle. Next, we divided the regular rectangle into grids of 20*20. There were 252 grids after removing the blank according to algorithm one. Finally, we gave weight to these 252 grids according to algorithm two, and calculated the cumulative coverage ratio and interval section ratio based on the collection of the 1006 tourists’ tracking as shown in Figure 4.



(a) The cumulative coverage ratio of Disneyland Park

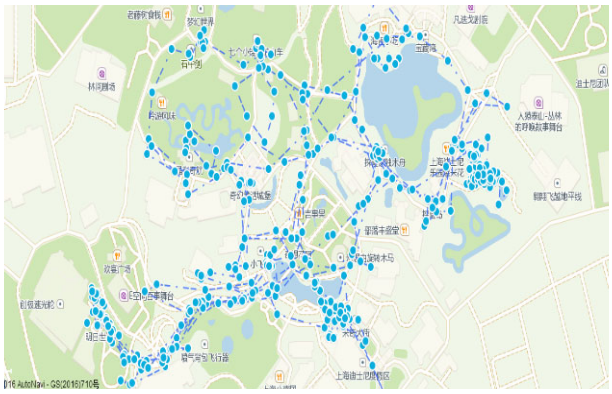


(b) The section ratio of Disneyland Park

Figure 4. Disneyland Park coverage ratio

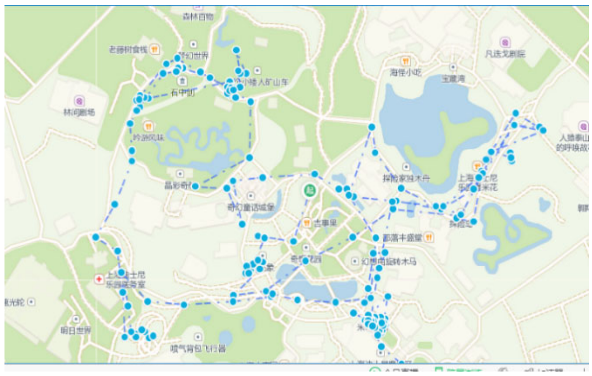
As shown in the subgraph(a) of Figure 4, the coverage ratio of 82.51% of the whole tourists in the park could reach more than 50%, but the coverage ratio of only 10.04% of the tourists could reach more than 80%. As shown in the subgraph(b), the coverage ratio of 37.28% of the tourists was between 70%-80%; the coverage ratio of 9.84% of the was between 80%-90%; the coverage ratio of only 0.20% of the tourists could reach 90%-100%. That is to say, the coverage ratio of only two tourists among the 1006 tourists reached more than 90%.

In order to give a more simple description for the coverage ratio of different tourists' track, we showed the historical track of the tourists among the 1006 tourists whose coverage ratio reached 90%, 80%, 70%, 60%, as shown in Figure 5. As shown in the subgraph(d), the tourist whose coverage ratio was only 60.19% did not tour many scenic spots such as Enchanted Storybook Castle, Dumbo the Flying Elephant, Voyage to the Crystal Grotto, Seven Dwarfs Mine Train, Treasure Cove and Fantasyland, so the tour quality of that tourist was not good enough. As shown in the subgraph(c), the tourists whose coverage ratio was 70% just had a rough look in the park and did not tour many scenic spots so there was few historical tracks for them.



(a) Coverage ratio of 91.97%

(b) Coverage ratio of 80.01%



(c) Coverage ratio of 70%

(d) Coverage ratio of 60.19%

Figure 5. Historical track of student cards

4.2 Degree of Fatigue Analysis

4.2.1 Definition

Definition 1. KL distance namely the relative entropy, refers to the distance between the two probability distributions to measure the difference of the two probability distributions in the same event and space. The calculation formula is as follows:

$$D(P \| Q) = -\sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad \textbf{(1)}$$

($D(P \| Q)$) shows the KL distance of probability distribution P(x) and Q(x). P(x) and Q(x) are two different distribution)

Definition 2. Degree of fatigue k: refers to the kl distance at each time point. When k tends to 0 or to a small value shows the coverage ratio does not change basically, that is to say a tourist does not move more or reaches a new place thus the tour is saturated. We define this situation as the degree of fatigue.

4.2.2 Algorithm Description

This paper took the Shanghai Disneyland park tourists as examples to analyze their degree of fatigue by using the KL distance algorithm.

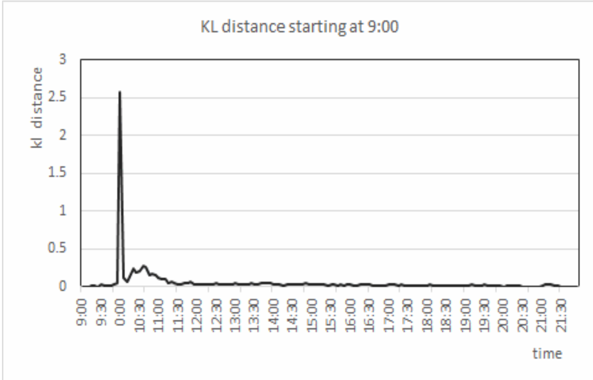
Count the tracking of all the tourists every 5 minutes speaking, the coverage ratio is the highest at 11:05 during this period of time. If there is no coverage ratio for a tourist at 11:05, we will take the coverage ratio of the time point before 11:05 as the coverage ratio. If there is still no coverage ratio, we need to continue looking for until we find one. Then count the coverage ratio of this time point (the maximum coverage ratio is over 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%, respectively). We got a total of 144 time points' cumulative coverage ratio. We can calculate the kl distance according to the cumulative coverage of the adjacent moments. The calculation formula is as follows:

$$(D(P \| Q)) = \sum \log((\frac{P(i)}{Q(i)})) \quad \textbf{(2)}$$

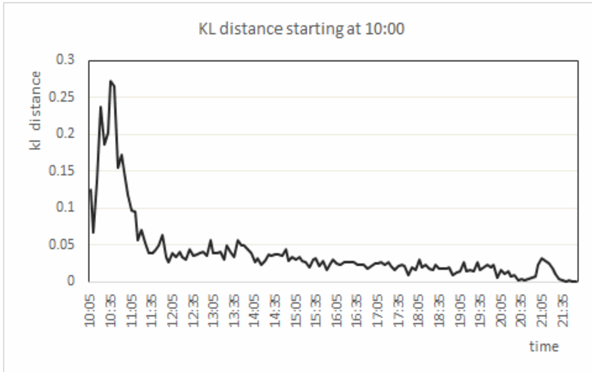
(i from 0 to 10 shows the cumulative coverage ratio from 0% to 90%. P (i) and Q (i) show the cumulative coverage ratio of the corresponding i respectively. D (P||Q) shows the kl distance of the two adjacent time points, namely shows the change of cumulative coverage ratio between two time points)

4.2.3 Algorithm Validation and Results

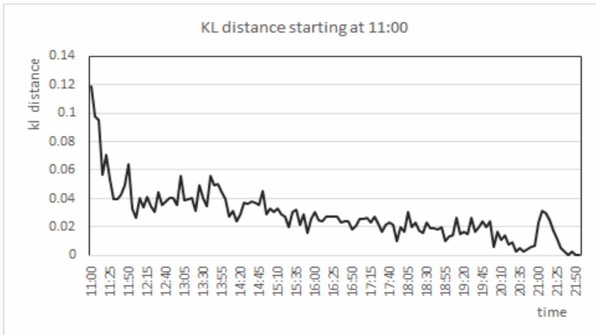
Figure 6 Summary of the tourists' behavior, as shown in Table 3.



(a) KL distance starting at 9:00



(b) KL distance starting at 10:00



(c) KL distance starting at 11:00

Figure 6. KL distance change

Table 3. Tourist behavior analysis

Time	Tourist behavior
10:00	In the subgraph (a), kl appears a peak, also the maximum distance. It shows the largest number of tourists start to enter Shanghai Disneyland Park in a day at this time point and the coverage ratio becomes high rapidly.
10:30	In the subgraph (b), kl appears the second peak. Tourists are in the highest spirit at this time point, then they quickly spread out and tour around, thus the coverage ratio becomes high rapidly once again.
11:30	In the subgraph (c), it appears another peak at this time point. It shows the tourists are in high spirit again and the coverage ratio obviously becomes high again.

Table 3. Tourist behavior analysis (continue)

12:00	In the near time points, kl appears the minimum value. It shows the tourists begin to feel tired and no longer tour around.
13:30	KL appears another peak in the near time points. It shows the tourists begin to relieve fatigue and start their tour in the afternoon.
14:45	KL appears the last peak in the near time points. It shows the tourists are in high spirit at 14:45 in the afternoon.
15:00-20:00	KL has been showing a downward trend. It shows the tourists' excitement has decreased and they begin to feel tired at 20:00.
21:00	KL appears a peak at 21:00. It shows the tourists are moving toward the exit and leaving.
21:30	KL value is 0 at 21:30. It shows the tourists have already left the park at this time.

Therefore, Shanghai Disneyland Park can adjust its business hours and entertainment program according to tourist behavior and degree of fatigue in different time.

For example, most of the entertainment projects in Disneyland Park are open at 9:00. However, the tourists enter the park at 10:00 in general. Therefore, we recommend these projects open at 9:30 but not 10:00 as there will be some tourists arrive earlier.

Mickey's Film Festival and film projection booth open at 9:30, but 10:00 tourists reach the first exciting point till 10:00. It is recommended to open at 11:30, the third exciting point so that it can meet the requirements of some tourists who want to see the movie, but also meet the needs of some tourists who are tired at 12:00 and want to see the movie and relax. Alice in Wonderland maze and Disney Jungle Characters at Happy Circle closed too early, at 17:30 and 18:30 respectively when the tourists are still tour around thus it cannot meet the needs of tourists. It is too late to close the park at 22:00 because most of the tourists begin to leave the park at 21:00 and all the tourists have already left at 21:30. Therefore, we advise to close the park at 21:30 to shorten the working hours of staff.

4.3 Best Tour Route Planning

The results of coverage ratio analysis of section 4.1 showed that only 10.04% of tourists could tour most of the attractions. The tour quality is unsatisfactory. This section aims to design a tour route that the tourists can tour more attractions within a limited period of time to improve tour coverage ratio.

4.3.1 Definition

Definition 1. define the Disneyland attraction number as i, the definition of i as shown in Figure 7.

Number of scenic spot	Name of scenic spot
0	Alice in Wonderland Maze (Fantasyland)
1	Buzz Lightyear Planet Rescue (Tomorrowland)
2	Siren's Revenge (Treasure Cove)
3	Become Iron Man (Gardens of Imagination)
4	Camp Discovery (Adventure Isle)
5	Fantansia Carousel (Gardens of Imagination)
6	Disney Jungle Characters at Happy Circle (Adventure Isle)
7	Pirates of the Caribbean: Battle for the Sunken Treasure (Treasure Cove)
8	Voyage to the Crystal Grotto (Fantasyland)
9	Roaring Rapids (Adventure Isle)
10	Marvel Universe (Gardens of Imagination)
11	"Once Upon a Time" Adventure (Fantasyland)
12	Minnie Mouse and Friends (Mickey Avenue)
13	Mickey's Film Festival (Mickey Avenue)
14	Jet Packs (Tomorrowland)
15	Seven Dwarfs Mine Train (Fantasyland)
16	Enchanted Storybook Castle (Fantasyland)
17	Meet Mickey at the Gardens of Imagination (Gardens of Imagination)
18	Millennium Falcon (Tomorrowland)
19	Garden of the Twelve Friends (Gardens of Imagination)
20	Stitch Encounter (Tomorrowland)
21	Explorer Canoes (Treasure Cove)
22	Dumbo the Flying Elephant (Gardens of Imagination)
23	The Many Adventures of Winnie the Pooh (Fantasyland)
24	Star Wars Launch Bay (Tomorrowland)
25	Hunny pot Spin (Fantasyland)

Figure 7. Scenic spot number

Definition 2. define the major scenic spot m that the tourist attraction of i belongs to, named in the brackets as shown in Figure 7. A major scenic spot m includes many attraction spots.

Definition 3. the parameters defined in the top10 algorithm are as follows: ①Key--label of student card (namely the tourists), 15 digits of imei number for identification of student ID card terminal and 8 digits of date and time when the student ID card terminal uploads data form the label. ②center (i)--the center coordinates of tourist attractions of i; center(i).lat--latitude coordinates; center (i).lng--longitude coordinates ③j--the labels of top10, 0-9 represents top0, top1......top9 respectively; j.lat--latitude; j.lng--longitude; j.time--time; ④standardLat--to pinpoint the latitude of the tourist attraction region; standardLng--to pinpoint the longitude of the tourist attraction region, together with the standardLat to form one tourist attraction region, as a judgment for tourists whether to tour this tourist attraction.

Definition 4. define the label of top10 tour quality as key.

Definition 5. define the process of obtaining the time of a tourist to arrive at the scenic spot as identification, i identification refers to obtain the tourist reaches the scenic spot i.

4.3.2 Algorithm and Verification Process

Algorithm three: cognition algorithm recorded as RECO

```
FOR(center(i)){
IF((Math.abs(j.lat-center(i).lat)<=standardLat))&&(Math.abs(j.lng-center(i).lng)<=standardLng)){
RETURN j.time;}
```

Input: center(i),j.lat,j.lng

Output: j.time's list is a timeline for tourist j to arrive at the scenic spot i.

This process is RECO (i)

Algorithm four: lustering algorithm: Choose 10 tourists who have top tour quality and get their key. Next, use the recognition algorithm to get the time of reaching each scenic spot of each key, and then get the possible time to reach each attraction of the top10 tourists through the k-means clustering.

```
FOR (key) {// conduct for each key
FOR(center(i)){
RECO (i); // get the sequence N*1 matrix X of j.time, that is get the whole time to reach scenic spot i of key
[Idx,Ctrs,SumD,D]=kmeans(X,3,'Replicates',3,'Options',opts,'Start','uniform');
plot(X(Idx==1,1),'r'','MarkerSize',14)
hold on
plot(X(Idx==2,1),'b'','MarkerSize',14)
hold on
plot(X(Idx==3,1),'c.','MarkerSize',14)
plot(Ctrs(:,1),'yp','MarkerSize',14,'LineWidth',4)
plot(Ctrs(:,1),'yp','MarkerSize',14,'LineWidth',4)
plot(Ctrs(:,1),'yp','MarkerSize',14,'LineWidth',4)}}
```

Input: center(i)(i=0,1….9)

Output: the cluster center of each center that is the clustering of time for each scenic spot

X data matrix of N*P, the storage of N vectorP

Idx vector of N*1, the storage of the cluster label of each point

Ctrs matrix of K*P, the storage of K cluster centroid location, combined with the actual situation, the use of k=3 clustering

SumD the sum vector of 1*K, the storage of the sum of the distances between all points among classes and the centroid point of the class

D matrix of N*K, the storage of the distance between each point and all the centroids;

Algorithm five: route planning algorithm

Assuming an arbitrary starting major scenic spot is m, ①the query range: the scenic spot belongs to m and the scenic spot included in the first node of Chain table for adjacent scenic spots of m, select the scenic spot i_0 with the minimum time t_0, which t_0-t>=10; ②after removing the scenic spots have been chosen in ①, repeat step ① until all the scenic spots of m have been selected or 80% of the whole of m have been selected, and mark the last selected scenic spot for i_k, time for t_k, which t_k-t_{k-1}>=10&&<=60; there are no repetition from i_k to i_0; add m into mlist; ③ confirm m_k of i_k, and regard m_k as the initial scenic spot, then repeat steps① ② to get n1 scenic spots and mark the last selected scenic spot for i_{k+n1}, time for t_{k+n1}; add m_k into the mlist; ④mark m_n that i_n located in as the initial scenic spot; repeat steps①② to get the last scenic spot for i_{n+1}, time for t_{n+1}, if m_{n+1} that i_{n+1} located in is not in the mlist, continue steps①②, if it is in the mlist, then consider to add the next node of m_n that not included in the mlist into query range; and repeat steps ①②④ to get the last scenic spot for $i_{n+1,}$ time for t_{n+1}, if m_{n+1} that i_{n+1} located in is not in the mlist, continue steps①②, if it is in the mlist, continue to add the next node of m_n that not included in the mlist into query range until all nodes included in the mlist that the algorithm is over. Therefore, the route is (i_k, t_k) sequence.

```
for (mi) {// for the major scenic spot mi
p[n][1]=query(mi[n]) // n is the number of spots that mi conclude, the function query is to get the data of mi and storage it to two-dimensional array p.
re[0]=null;re[1]=null; // re[0] storages the number of the latest spot, re[1] storages the time of the latest spot.
do{
r[]=findLeastTime(p) // find the minimum time from p and return the number and time of the spot. Array r storage them.
if(r[1]-re[1]>=10&&(r[1]-re[1]<=60)){
result.i.add(r[0])
result.time.add(r[1])// the two linked lists i and time in structure result add the result of array r into the list.
re[0]=r[0];re[1]=r[1];
reset(r[0]) // r[0] refers to the spot has added into the result linked list; reset the data of r[0] to null in p.
}else{
n=n-1 // n is number of major scenic spot mi
}
}while(isempty(p)||(n=0))
mi=mlist.next // search for the next major scenic spot and set it to mi, then repeat the for loop.
```

input：the original major scenic spot mi

output：a structure includes two linked list i and time which shows the number and time of the spots in the planned route.

Verification process:

(1) Confirm the top10

We get the top10 who have the best tour quality based on the analysis of 4.1. The labels of top10 are as follows:

867478A2AA112A120160710,867478A2AA1123520160705,867478A2AA1124320160712,867478A2AA1129220160704,867478A2AA1131820160703,8674

78A2AA1135920160705,867478A2AA114172016070 3,867478A2AA1141720160711,867478A2AA114582 0160629,867478A2AA1163120160701. (these labels are index numbers for just searching in the database for data and mean nothing else)

We get the coordinates of the top10 uploaded at random time.

(2) To get information of attractions of Shanghai Disneyland park

To get the information of all the attractions of Shanghai Disneyland Park, we get six major scenic spots from the official app of Shanghai Disneyland Park: Tomorrowland, Fantasyland, Mickey Avenue, Gardens of Imagination, Treasure Cove, Adventure Isle. The detailed information of the attractions as shown in Figure 8, marking the number, the area and the central coordinates of the scenic spots.

i	center position	clustered time
0	31.144433,121.659484	11:35/15:00/18:00
1	31.140956,121.657608	14:11
2	31.145145,121.662918	12:31/15:29/18:00
3	31.142747,121.660517	11:06/17:00/19:37
4	31.143768,121.663795	11:35/15:00/18:00
5	31.142559,121.661315	10:18/15:51/19:24
6	31.144318,121.664157	10:40/14:59/18:26
7	31.145526,121.661697	11:39/15:34/17:24
8	31.143721,121.658477	12:14/16:00/19:41
9	31.142843,121.663154	12:11/13:52/14:23
10	31.142454,121.658683	12:38/15:00/19:11
11	31.145448,121.658691	12:00/13:51/18:58
12	31.141836,121.661803	10:28/20:00
13	31.141819,121.66161	10:23/20:00
14	31.141548,121.65876	17:34
15	31.145103,121.659729	12:00/17:00
16	31.143428,121.659623	10:41/13:12/16:33
17	31.142375,121.659263	12:48/16:00/19:00
18	31.14176,121.656781	11:00/14:01/18:09
19	31.142971,121.660949	14:00/15:16/17:03
20	31.141193,121.658344	13:52/18:38
21	31.143813,121.662156	12:51/15:00/19:26
22	31.142635,121.659812	12:03/16:18/19:26
23	31.145607,121.658645	12:00/13:47/19:15
24	31.141373,121.657636	13:00/16:00/19:41
25	31.145306,121.658712	12:00/14:00/19:00

Figure 8. Location information of the scenic spots and the clustering time of the top10 in each of the scenic spot

(3) We used recognition algorithm to calculate the time that the top 10 arrived at each scenic spot and got the time they reached the scenic spot i (i=0,1......9). Then we used clustering algorithm to conduct k-means clustering of time to obtain the time clustering of each attraction, as shown in Figure 8.

(4) Route planning: For the Disneyland area, 6 major scenic spots m are: Fantasyland, Tomorrowland, Adventure Isle, Treasure Cove, Gardens of Imagination, Mickey Avenue. Construct a table of adjacent scenic spots of m according to the geographical location and the actual situation, as shown in Figure 9. According to the route planning algorithm, after inputing the name of starting major scenic spot respectively: Treasure Cove, Adventure Isle, Gardens of Imagination, Tomorrowland, we got two reasonable routes, which have toured more than 75% of the whole scenic spots in a day.

m	Chain table for adjacent scenic spots of m			
Treasure Cove	Adventure Isle	Gardens of Imagination	Fantasyland	
Adventure Isle	Adventure Isle	Gardens of Imagination	Fantasyland	
Gardens of Imagination	Fantasyland	Tomorrowland	Treasure Cove	Adventure Isle
Tomorrowland	Fantasyland	Gardens of Imagination		
Fantasyland	Gardens of Imagination	Tomorrowland	Treasure Cove	

Figure 9. Chain table for adjacent scenic spots of m

4.3.3 Route Planning Results

Route planning results, as shown in Figure 10. From the number of toured scenic spots, route 1 is better; compared the average time after calculating the average time of each route, route 2 is better. Therefore, there are two optimal routes.

route1			route2		
number on the map	arrival time	number of scenic spot	number on the map	arrival time	number of scenic spot
0	10.28	12	0	10.18	5
1	10.4	6	1	10.28	12
2	11.35	4	2	11.06	3
3	12.11	9	3	11.35	4
4	12.31	2	4	12.11	9
5	12.51	21	5	12.31	2
6	13.12	16	6	12.51	21
7	13.47	23	7	13.12	16
8	14	11	8	13.47	23
9	14.11	1	9	14	25
10	15	0	10	14.59	0
11	15.16	19	11	16	8
12	15.51	5	12	16.18	22
13	16.18	22	13	17	15
14	17	15	14	17.34	14
15	17.34	14	15	18	18
16	18	18	16	18.38	20
17	18.38	20	17	19	17
18	19	25	18	19.11	10
19	19.11	10	19	20	13
20	19.37	3			
21	19.41	24			
22	20	13			

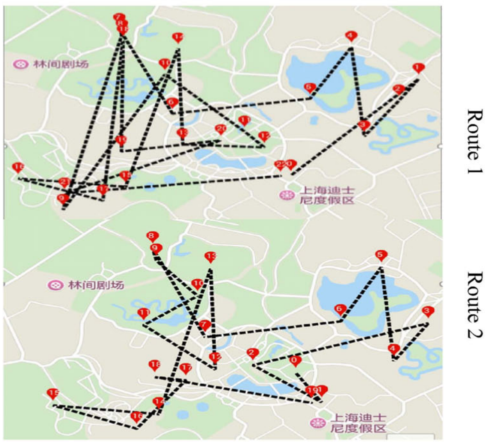


Figure 10. Route map. It shows the name of the scenic spots that two routes go through and the scheduled arrival time. The routes were marked on the map

4.3.4 Route Planning Results

Definition 1. actual waiting time: The waiting time of reaching a certain scenic spot according to planning routes corresponding to the waiting time in Figure 11;

time	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00
0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	0			
1	5	5	15	20	10	5	5	5	10	5	5	25	15	20	10	10	5	5	5
2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
3	5	5	5	5	15	15	10	10	20	5	10	10	10	10	10	10	10	10	10
4	20	20	20	20	20	10	10	10	10	10	30	20	30	30	20	20	10	5	0
5	40	30	40	30	30	30	30	20	10	10	20	20	30	30	30	30	50	40	10
6	5	5	5	10	10	10	5	5	5	5	5	5	5	10	10	20	10	0	
7	25	30	30	30	15	15	15	20	10	10	15	10	5	10	10	5	5	0	5
8	10	20	40	40	30	20	40	40	30	20	30	40	30	30	30	30	20	20	10
9	40	30	60	60	50	50	60	60	75	50	75	60	60	75	60	35	35	25	5
10	3.3	3.3	3.3	3.3	3.3	5	3.3	3.3	3.3	3.3	5	5	5	3.3	5	1.6	5	5	3.3
11	5	10	20	30	30	10	30	50	10	5	30	20	30	30	60	30	30	10	5
14	30	40	40	30	40	30	30	40	30	30	50	40	40	30	30	30	30	30	5
15	40	40	50	60	65	70	70	80	90	90	80	100	100			50	75	60	5
16	30	40	40	30	30	30	30	30	30	20	40	40	40	30	50	0			
17	5	10	25	15	10	10	10	25	10	10	20	10	10	30	20	10	5	20	5
20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20		
21	40	40	40	50	40	40	40	40	40	30	50	30	0						
22	50	60	50	60	50	50	50	40	30	10	50	40	40	30	40	0			
23	10	20	50	30	30	40	40	50	50	20	40	40	50	50	60	0	30	40	5
25	5	5	10	10	10	10	20	20	10	10	5	10	10	10	10	20	5	5	5

Figure 11. A table for different waiting time of scenic spots in different time

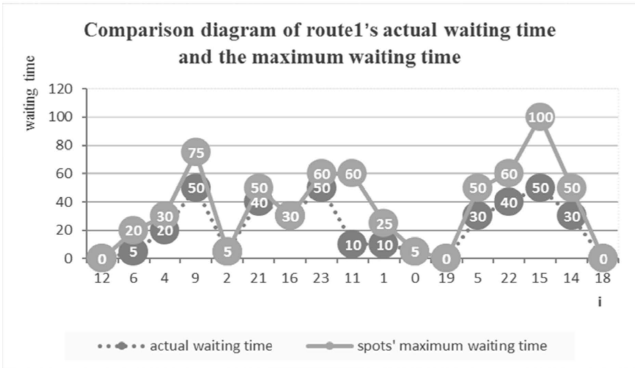
Definition 2. maximum waiting time for a scenic spot: The maximum waiting time in a day of a scenic spot;
Definition 3. average waiting time for a scenic spot: The arithmetic average of the whole time waiting of a scenic spot in a day;
Definition 4. major scenic spot: A scenic attraction whose maximum waiting time is more than 60 minutes. Under the premise of the above definitions, this paper has verified the rationality of the route from three aspects.

① Whether it is the peak waiting time when tourists arrive at a certain scenic spot according to the planning routes

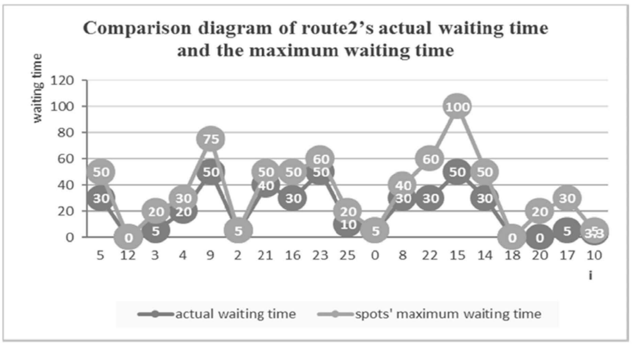
First of all, we collected the waiting time of all the scenic spots every half hour after 10:00 and got the table of different waiting time for each scenic spot at different time as shown in Figure 11, thus we knew the maximum waiting time of each scenic spot.

The scenic spot was temporarily closed during our recording time or the scenic spot has closed at that time thus there were several blank grids in the table.

We compared the waiting time when tourists arrived at a certain scenic spot with its maximum waiting time of the route1 and route2 respectively, and got the comparison diagram as shown in Figure 12. Where the graph's horizontal axis shows the scenic spot number i, and the vertical axis shows the corresponding waiting time. The number of major scenic spots of route1 was larger than that of route2 and route1 saved more time than route2, showing that comparing with route2, route1 was time-saving and tourists could tour more scenic spots in the long waiting time.



(a)



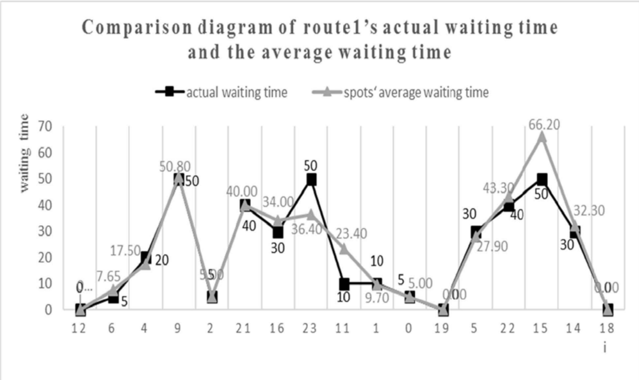
(b)

Figure 12. Comparison diagram of two routes and the maximum waiting time. Route1 included the number of scenic spots were 9, 23, 11, 22, 15, and it saved 25+10+50+20+50=155 minutes in total when toured these scenic spots. Route2 included the number of scenic spots were 9, 23, 22,15, and it saved 25+10+30+50=115 minutes in total when toured these scenic spots.

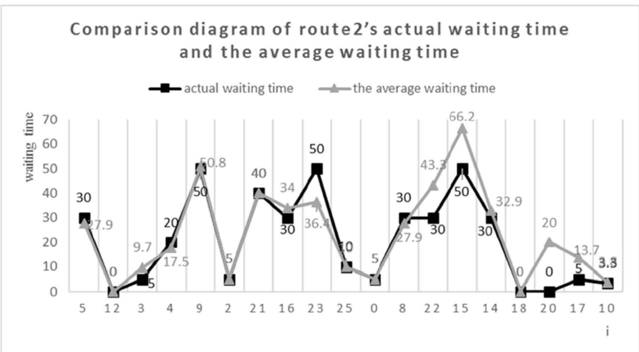
②Took the average waiting time of each route into consideration

The horizontal axis shows scenic spot number i and the vertical axis shows the waiting time of scenic spot i of the two routes, as shown in Figure 13. We concluded that compared with the average waiting time of each scenic spot, route2 saved more time than route1 and proved to be better.

③ Extracted three random routes of tourists to compare route1 and route2



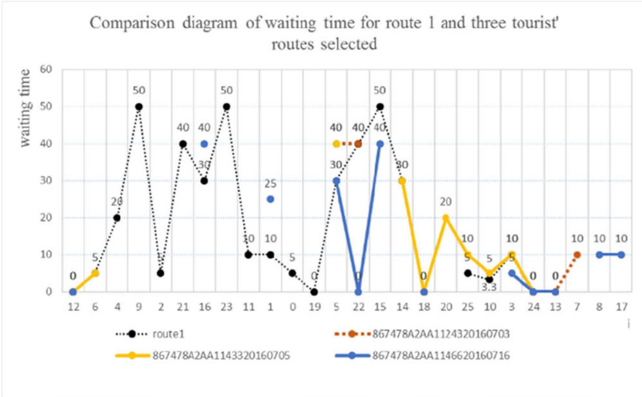
(a) Comparison diagram for average waiting time of route1. The waiting time for most of the scenic spots of route1 was less than or equal to the average waiting time of a certain scenic spot, but there were a scenic spot number 23 whose waiting time exceeded its average waiting time. Compared with the average waiting time, route1 saved 24.15 minutes



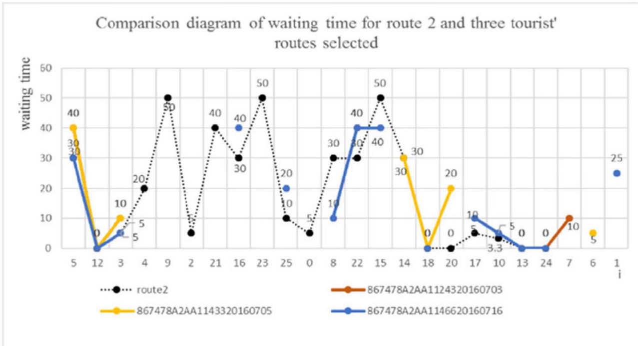
(b) Comparison diagram of average waiting time of route2. Compared with the average waiting time, route2 saved 50.8 minutes

Figure 13. The comparison of actual waiting time and average waiting time

As shown in Figure 14, the number of toured scenic spots of route1 was greater than that of the three random routes, and the waiting time of route1 is less than that of the other three routes for the same scenic spots. In comparison with the maximum waiting time, route1 saved 271.7 minutes in Figure (1); the waiting time of the tourists whose key tail number were 03, 05 and 16 saved 60 minutes, 50 minutes and 195 minutes respectively compared with the maximum waiting time. The waiting time of the tourist with key tail number 03 increased 4.38 minutes, the tourist with key tail number 05 increased 6.85 minutes, the tourist with key tail number 16 saved 21.24 minutes compared with the average waiting time as shown in Figure(2). Route2 saved 50.8 minutes compared with the average waiting time so route2 was better.



(a)



(b)

Figure 14. Comparison diagram of waiting time for route 1, route 2 and three random routes

5 Conclusion

In this paper, we mainly focused on the BDS/GPS dual mode positioning, supplemented by base station/WiFi positioning in the experiment, integrating indoor and outdoor positioning to achieve high precision seamless positioning. The hybrid application of a variety of positioning modes and Internet support, have guaranteed the positioning accuracy and the coverage. We proposed idea of circling the area and divided the area into grids based on the high precision position data and used weight algorithm for them. We also proposed the cumulative coverage ratio based on kl, to analyze the tour quality and tourist behavior from the cumulative coverage ratio and put forward the route planning algorithm based on K-means clustering to plan the optimal route for tourists in this park. This experiment in this paper took Shanghai Disneyland Park as an example, using the proposed algorithms to carry on experiment. We collected a large number of tourists' tracking data and analyzed them to give some suggestions to Shanghai Disneyland Park based on the tourist behaviors. What's more, we panned an optimal route for touring the scenic spots in this park. We tested the planning route through model test and it proved to be optimal.

Based on the idea of big data analysis, we analyzed and mined the collected track data, and made useful conclusions to find more potential value, which could provide related services and advice to smart tourism industry.

In future work, on the one hand, we are ready to collect more tracking data to know more potential behaviors and habits of tourists. On the other hand, we will optimize the tour route and plan optimal routes for different situations so that tourists can choose different routes according to different circumstances, making the tour route suitable for different situations and more intelligent.

Acknowledgements

The research was supported by a grant from the National Key Research and Development Program of China under grant No.2016YFB0501805, and by a grant from the National Natural Science Foundation of China under grant No. 61573262.

References

[1] Y. F. Li, *Study on Smart Tourism Development and Tourism Public Service System Construction*, M.A. Thesis, Lanzhou University, Gansu, China, 2015.

[2] F. Zhu, C. H. Sun, Beidou High Precision Positioning Technology and Intelligent Transportation System, *Information Technology & Standardization*, Vol. 2014, No. 10, pp. 24-27, December, 2014.

[3] C. Lin, J. Zou, J. M. Huang, Research on Application of High-precision Measurement Receiver Based on Bei Dou in the Field of Displacement Monitoring, *Bulletin on Surveying and Mapping*, Vol. 2015, No. 2, pp. 64-66, February, 2015.

[4] J. C. Hung, The Smart-travel System: Utilizing Cloud Services to Aid Traveler with Personalized Requirement, *International Journal of Web & Grid Services*, Vol. 8, No. 3, pp. 279-303, September, 2012.

[5] J. C. Hung, V. Hsu, Y. B. Wang, A Smart-Travel System Based on Social Network Service for Cloud Environment. *Third International Conference on Intelligent NETWORKING and Collaborative Systems IEEE*, Fukuoka, Japan, 2011, pp. 514-519.

[6] L. Sabatucci, A. Cavaleri, M. Cossentino, Adopting a Middleware for Self-adaptation in the Development of a Smart Travel System, *Intelligent Interactive Multimedia Systems and Services 2016*, Puerto de la Cruz, Tenerife, Spain, 2016, pp. 671-681.

[7] H. Zhang, X. M. Wang, X. C. Guo, J. Cao, C. S. Zhu, Y. R. Guo, Application of Taxi GPS Track Big Data in Intelligent Transportation System, *Journal of Lanzhou University of Technology*, Vol. 42, No. 1, pp. 109-114, May, 2016.

[8] Q. Zhang, Data processing of Bus GPS Historical Track Based on Big Data, *Internet of Things Technologies*, Vol. 6, No. 4, pp. 52-53, April, 2016.

[9] M. Xu, *Study on Key Problems of Urban Road Network Analysis Based on Vehicle Track Mining*, Ph.D. Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2015.

[10] H. Kimm, H. Kimm, Improved Track Path Method in Real Time by Using GPS and Accelerometer, *International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security IEEE*, Granada, Spain, 2016, pp. 77-84.

[11] S. Karagiorgou, D. Pfoser, On Vehicle Tracking Data-based Road Network Generation, *International Conference on Advances in Geographic Information Systems*, Redondo Beach, CA, pp. 89-98, 2012.

[12] J. A. Downs, M. W. Horner, Probabilistic Potential Path Trees for Visualizing and Analyzing Vehicle Tracking Data, *Journal of Transport Geography*, Vol. 23, No. 3, pp. 72-80, July, 2012.

[13] S. C. Lo, W. K. Lu, Packet Routing Schemes in Vehicular Ad Hoc Networks, *Journal of Internet Technology*, Vol. 12, No.6, pp.931-940, November, 2011.

Biographies

Yan Liu was born in 1995. She received the bachelor's degree of engineering from Wuhan University, Wuhan, China, in 2016. She is currently a postgraduate student and pursuing for master degree. Her research interests is information security

Chi Guo was born in 1983. He received his Ph.D. in computer science from Wuhan University, Wuhan, China, in 2010. Currently, he is an associate professor of satellite navigation and positioning technology research center of Wuhan University. He researches on navigation and location services, ubiquitous mapping technology and methods.

Manli Zhang is a graduate student of School of Computer Science in Wuhan University. She got the bachelor degree of engineering from Wuhan University. And she won the freshman scholarship in 2016.

Xiaoqun Wu received the B.Sc. degree in applied mathematics and the Ph.D. degree in computational mathematics from Wuhan University, Wuhan, China, in 2000 and 2005, respectively. She is currently a Professor with the School of Mathematics and Statistics, Wuhan University.

Ye Zhou was born in 1983. She is a master of sociology science, and her major research interests are intelligent tourism, the analysis and processing of tourism big data.