

Paper

by Wenhao Kang

FILE	17875_WENHAO_KANG_PAPER_260366_1146237429.DOCX (1.85M)		
TIME SUBMITTED	19-NOV-2020 11:56AM (UTC+0000)	WORD COUNT	13052
SUBMISSION ID	136773488	CHARACTER COUNT	70097



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

**Resident travel survey system based on
smart phone**

by

Wenhao Kang

Supervisor: **Gaynor Thomas**

Project submitted as part of the requirements for
the award of MSc ... (fill in as applicable)

November 2020

DECLARATION

I, ...Wenhao Kang..... declare that I am the sole author of this Project; that all references cited have been consulted; that I have conducted all work of which this is a record, and that the finished work lies within the prescribed word limits.

This work has not previously been accepted as part of any other degree submission.

Signed : Wenhao Kang.....

Date : 14.11.2020.....

FORM OF CONSENT

I Wenhao Kang..... hereby consent that my Project, submitted in candidature for the M.Sc. Applied Computing degree, if successful, may be made available for inter-library loan or photocopying (subject to the law of copyright), and that the title and abstract may be made available to outside organisations.

Signed : Wenhao Kang.....

Date : 14.11.2020.....

Abstract

Urban residents' travel behavior survey is the core content of urban planning traffic dispatch. By collecting residents' travel characteristics and rationally analyzing travel sample data, discovering the potential travel needs of urban residents and controlling the key hubs of urban transportation can analyze and plan the current situation of urban traffic. Management provides scientific basis. Due to the high cost of traditional travel questionnaires, low efficiency, low data validity, and difficulty in post-processing and analysis, it cannot meet the growing demand for urban growth.

As a necessary device for everyone in the Internet age, smartphones, coupled with the deployment of 5G today, make data sampling more accurate and faster. In today's era of explosive growth of data traffic every day, it is very meaningful to design a comprehensive information system that can collect, manage and analyze residents' travel data efficiently and at low cost. At present, most of the domestic and foreign countries use traditional survey forms or pure GPS data for data collection and analysis. It does not use the multi-sensor advantages of the smartphone itself. The main research contents of this paper are:

- (1) Combining the technical principles of the various sensors of the mobile phone and the advantages and disadvantages of each sensor parameter, a method of stopping point recognition based on mobile phone multi-motion parameters is proposed, and its effectiveness is analyzed empirically.
- (2) Based on a large amount of data collected on the spot by the mobile phone travel survey collection system, a travel feature extraction method based on personal travel records is proposed to identify the comprehensive characteristics and laws of travel survey data.
- (3) Based on the JAVA WEB platform, combined with key technologies such as the mobile phone APP platform and map services, the design and implementation of a mobile phone-based travel survey system has been completed.

This paper studies and designs a real-time stay recognition algorithm based on mobile phone multi-motion parameters, which is implemented on the iOS mobile phone platform, and by uploading the data collected on the mobile phone to the server background, the data is semantically performed on the Web platform Feature extraction, and through cluster analysis, can actively identify urban hotspots, and realize a smart phone-based resident travel survey system. The entire system can be successfully implemented.

Keywords: travel survey, smart phone, stop point recognition, travel characteristics

Content

1 Introduction.....	3
1.1 Subject source	3
1.2 Background and significance of the project.....	3
2 literature review	4
3 research methodology	11
4 design and implementation	14
4.1 Recognition algorithm of stay point based on mobile phone accelerometer .	14
4.1.1 Design of stop point identification process.....	14
4.1.2 Algorithm implementation.....	16
4.2 Feature extraction for travel survey	19
4.2.1 Personal travel characteristics model.....	19
4.2.2 Judgment of travel effectiveness.....	20
4.2.3 Travel distance	21
4.2.4 Travel time and time distribution	21
4.2.5 Travel speed	22
4.3 Hot spot discovery	23
4.4 System Requirements Analysis.....	25
4.5 System Design	26
4.6 Overall system architecture.....	28
4.7 Design Modeling of System Realization Process Based on UML	29
4.8 Implementation of the recognition method based on mobile phone multi-motion parameters	30
4.9 Reality of the Web-based travel survey process	32
5 Results	34
5.1 Stop point recognition result	34
5.2 Analysis of travel feature extraction results.....	40
5.3 Hot spot results	42
6 conclusion	43
7 Reflection	45
8 Reference	46

1 Introduction

1.1 Subject source

The research work of this topic comes from the traffic travel survey project based on smart phones.

1.2 Background and significance of the project

In recent years, with the urbanization of the world's population and the rapid development of the market economy, the people's economic level and material and cultural needs continue to increase. With the urbanization of rural areas and the acceleration of the process of urban-rural integration, urban traffic conditions have increasingly become urban governance. The difficulty. Due to traffic congestion, commuting distance, and long commuting time in domestic big cities, the cost of urban transportation construction has increased sharply, resulting in a decrease in the happiness of urban residents, hindering the development of the city, damaging the competitiveness of the city, and becoming one of the main urban governance problems. For example, due to high transportation costs and other reasons, Beijing has become the city with the highest operating cost in China. Therefore, urban traffic management and planning are facing tremendous pressure and challenges now and in the future [1].

Resident travel surveys have always been a popular research direction at home and abroad. Traditional traffic travel surveys are manual face-to-face questionnaire surveys, which are inefficient and difficult to obtain a large amount of effective data. For the interviewees, when filling out the questionnaire, they always use memory to recall where they went and what they did. Such sampling is not accurate enough. Therefore, this kind of reminiscence sampling not only causes a certain living burden on the respondents, but also the recovered data will vary depending on the cognitive ability and memory of different people. It can be seen that traditional travel survey methods are not enough to support people's needs for travel behavior surveys in the new era.

This paper designs and implements a mobile phone-based travel behavior survey application and a Web-based travel survey process management system. This paper designs a mobile-based stop-point recognition algorithm based on mobile phone multi-sensor parameters, which avoids the problem of data missing and positioning data drift of GPS equipment. The mobile phone survey application is used to track the user's location in real time and generate user stay points, and can upload the generated stay points to the Web travel survey management system. In the Web system, follow-up management and investigation of the travel trajectories and stay points of a large number of users can be performed. A travel feature extraction model for users' personal

trips is proposed. Based on this, cluster analysis can be used to discover urban hotspots for investigating comprehensive travel characteristics.

2 literature review

Urban residents' travel surveys can be roughly divided into two categories, namely, active surveys and passive surveys. The active investigation method mainly includes

home visit investigation, telephone investigation, leaflet investigation, and website visit investigation. The specific content of the survey mainly depends on the questionnaire designed by the investigator. The passive survey mainly includes GPS-based trajectory survey and smart phone data collection.

Since 1996, scholars have applied GPS trajectory to travel survey research. In the past two decades, there have been many documents and research results at domestic and abroad. Most of the research content is to extract feature information from the original GPS track data for semantic analysis. For example, Jianhe et al [2] built a model that can propose travel segments from a large amount of GPS trajectory data, thereby improving the accuracy of travel rate calculation results. Tsui and Shalaby [3] (2006) developed an analysis system that can automatically process personal travel surveys based on GPS trajectories. Wendy Boht et al [4] designed and developed a travel survey system that combines GPS equipment to obtain data, GIS analysis data and Web interactive verification data. The main purpose is to obtain and verify the user's travel purpose and travel mode. Finally, it was carried out in the Netherlands. A lot of field trials. Hongmian Gong et al [5] designed a system that can extract five travel modes from raw GPS data, namely walking, car, bus, subway, and railway. Neng Wan et al [6] proposed a GPS data processing method to calculate the user's living space, and launched a tracking survey that lasted four months to infer the user's activity information and living space.

With the rapid development and popularization of smart phones, coupled with the limitations of GPS active surveys, passive travel surveys based on smart phones have become a research hotspot in current travel survey data collection. In foreign countries, Singapore's FMS (Future Mobility Survey) has conducted in-depth research. FMS is a survey of residents' travel conducted by the Singapore Future City Mobility Group. The basic facilities in the FMS project are mobile phones and computer terminals. The collected data is uploaded to the server for analysis, and the data collection automation is realized together. Caitlin D. Cottrill et al [7] (2013) have designed and implemented a smartphone-based travel survey system, which mainly includes a reminiscence web diary Investigation, based on smart phone trajectory collection, and subsequent data analysis. Sari, Assemi et al [8] designed a personal travel data collection system based on smartphones, and applied it to residents' travel surveys in New Zealand.

On the stop point recognition algorithm, the main methods currently used are the exploratory rule-based method and the clustering method based on data mining. The exploratory-based empirical rule method is based on the periodic acquisition of data by GPS equipment, and comprehensively uses various parameters to set rules to determine the stay point. The parameter thresholds are based on actual experience or system sensitivity requirements. Commonly used basic parameters include Time, speed, distance, etc. Gong H is set within 200 seconds of the specified time. If the distance between all tracks is less than 50m, it is judged as staying. The disadvantage of this

method is that because the duration is used as the threshold, the data containing noisy points or floating points will be missed.

Wang Zhongmin proposed a method of behavior recognition and classification based on the daily human behavior data collected by the built-in three-dimensional acceleration sensor of the mobile phone. This method extracts a variety of statistical features in the horizontal and vertical directions, including standard deviation, interquartile range, signal amplitude, Declination, kurtosis and correlation coefficients, etc., using sliding window technology to use SVM classifier to classify and recognize, identify the daily behaviors of mobile phone carriers such as standing, walking, and running.

Zhao F et al. proposed a rule-based stop point identification method based on mobile phones. The method has 6 steps:

Step1. Filter the noise points in the original trajectory data;

Step2. Use the K-median method to generate undetermined stop points for the positioning track data;

Step3. For the inaccurate location of the pending stop caused by the GSM handover problem, the GSM signature is used for association matching;

Step4. According to empirical rules, merge the continuous fine-grained pending stay points;

Step5. Detect the static motion mode to make up for the stop point identification problem under the lack of positioning data;

Step6. Eliminate the misjudged stay points.

This method has greatly improved the accuracy of stop point recognition, but there are still some problems. For example, the K-median method is still used to generate the remaining stay points, and the stay point samples containing noise cannot be identified. The clustering method is also used for low-precision GSM positioning data. As a result, there will be several staying points far away from each other because the mobile phone is in the handover area of several GSM signal towers.

In summary, the current researches on mobile phone-based stop point recognition at home and abroad are still based on GPS positioning data, and the data source is relatively single. When the GPS signal is interfered with, the recognition effect is not ideal, and the smart phone is not fully utilized. Other sensor advantages. Based on this, this paper fully combines mobile phone hybrid positioning technology and acceleration sensor, and proposes a multi-motion parameter model based on mobile phone stay point recognition, and proposes different classification methods for multiple stay modes, which better solves the single GPS Many limitations of staying point recognition under the data source improve the accuracy of staying point recognition.

As we all know, the ultimate goal of urban residents' travel survey is to extract the macro-level travel characteristics of the city from the travel survey data, and use this

feature to make guiding suggestions for later urban construction. Therefore, the personal feature extraction method based on mobile phone travel survey also has very important research significance.

Ramaswamy Hariharan et al [9] (2004) analyzed the distribution of the number of user's stay points in a week by using all the user's staying point data in a week. In order to use mobile phone data in transportation research, Francesco et al [10] (2013) proposed a method for extracting useful travel information from a large amount of mobile phone raw data to study the travel patterns of urban residents. This method mainly extracts the travel frequency, travel distance and the user's resident location, and conducts related analysis based on this. Vangelis et al [11] (2013) proposed a framework for extracting travel feature information from mobile phone network positioning data, and based on this, predicted and analyzed the travel demand in a specific area. Hu Zhongwei et al [16] (2013) proposed a technical route for the extraction and analysis of residents' travel characteristics based on mobile phone positioning data, and conducted a comprehensive analysis of the residents' travel OD distribution, population distribution and travel characteristics parameters, and carried out relevant analysis using example data. Youngsung et al [12] (2014) proposed a learning model to identify the type of travel activity of the user based on the user's stay point and economic attributes.

In order to obtain user stay points for results, the first thing we need to do is to accurately locate users through mobile phones. As we all know, mobile phone signaling data is generated by the GSM (Global System for Mobile Communications) communication network. Mobile phone positioning technology can be divided into three categories: positioning technology based on mobile phone GPS sensors, network positioning technology based on base station communication calculations, and hybrid Positioning technology (according to different scenarios, the first two technologies are applicable).

Based on the GPS sensor positioning technology of the mobile phone, the chip embedded in the mobile phone is used to obtain the geographic coordinates in the WGS (World Geodetic System) 84 geographic coordinate system, and the base station signal is used to quickly download the online satellite information to complete the rapid satellite search and positioning functions. GPS positioning higher accuracy.

The network positioning technology based on base station communication calculation uses the mobile communication network of the mobile phone to calculate the position of the mobile phone using certain algorithms, including our common single cell source (COO: Cell of Origin).

Hybrid positioning uses two positioning methods at the same time, so that location information can be obtained more quickly and accurately. There are currently two mainstream mobile phone operating systems, Android and iOS. Among them, iOS provides hybrid positioning combining Wi-Fi positioning, network positioning and

GPS positioning. The positioning operation is performed in the background, and the front-end application can call the packaged SDK to directly obtain the positioning data, which is very convenient and fast.

The acceleration sensor can measure the three-dimensional acceleration of the carrier through the sensor coefficient. In the current era of rapid technological development, many mobile electronic devices are equipped with this sensor. For example, smart phones, smart watches, VR, etc. [13]. The measurement principle of the acceleration sensor is that the sensitive element converts the acceleration signal of the measuring point into the corresponding electrical signal, enters the pre-amplification circuit, improves the signal-to-noise ratio of the signal through the signal conditioning circuit, and then performs analog-to-digital conversion to obtain the digital signal, and finally sends it to the computer, The computer then performs data storage and display.

The low power consumption characteristics of the acceleration sensor ensure the battery life of the mobile phone and will not occupy too much mobile phone power. The X, Y, and Z axes of this smart phone are all based on the mobile phone screen, and the X axis is along the mobile phone screen. From left to right, the Y axis is from bottom to top along the phone screen, and the Z axis is from the inside to the outside along the phone screen. The data measured by the three-axis accelerometer is the component of each axis in the three-axis coordinate system of the mobile phone. The specific calculation formula is shown in formula (2-1):

$$\vec{a}_x + \vec{a}_y + \vec{a}_z = \vec{a}_o + G \quad (2-1)$$

Where G is the acceleration of gravity, which is the actual acceleration of the sensor when the device is moving.

Because acceleration can integrate the acceleration in the X, Y, and Z directions of the device, it can more accurately reflect the speed change of the carrier, and more accurately reflect the movement state of the carrier, so it is based on the travel survey data measured by the acceleration sensor of the mobile phone, Has a strong reference value for judging the current user's exercise status.

The following is the definition of trajectory and stay point.

Definition 1. Trajectory: A description of the position change of a moving object, a collection of various attribute combinations in a continuous time and space, the expression is as follows (2-2):

$$T_p = \{p_i \mid p_i = (l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i), i \in I\} \quad (2-2)$$

l_i : the latitude and longitude coordinates of the track point;

t_i : time stamp of track point;

alt_i : altitude of track point;

v_i : track point velocity

$Accu_i$: horizontal position accuracy of track points

a_i : the modulus of the acceleration of the track point

s_i : motion state of track point

$type_i$: Track point positioning type

I : The set of points that make up the track

Definition 2. Staying point: A staying point refers to a user staying at a certain position during a trip. At this time, this track point can be defined as a staying point. The expression is as follows (2-3):

$$s_i = \{l_i, t_i^{start}, t_i^{end}\} \quad (2-3)$$

l_i : stop point coordinates

t_i^{start} : stop at the start moment

t_i^{end} : end of stay

Since positioning data is prone to drift in different environments, it is necessary to clean and filter the data before storing the data, so that the data stored in the database are all valid data.

(1) According to experience, the wrong location point will also produce the wrong altitude, so the altitude value at this time will have a very obvious error [14], by comparing $pi.alt$ with the local average altitude value, the noise can be judged by setting the threshold. Then, according to the average speed of the current point, the noise points can be filtered out through threshold comparison. For network positioning points, because most of the overall data is inaccurate, they can only be filtered according to the accuracy value $Accu$. The threshold is also set. Here we set the threshold to 50. If $Accu$ is greater than 50, the current positioning point will be filtered out. The specific data cleaning and filtering pseudo code is as follows:

Table 2-1 Pseudo code of mobile phone trajectory data cleaning

```

Input: raw location  $p_i = \{l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i\}$ 
if  $type_i == GPS$ 
    if  $(alt_i > 1.5 * ALT^{max} \mid Accu_i > 30)$ 
        remove  $T_i$ ;
    if  $(Distance(T_i, T_{i-1}) / (t_i - t_{i-1}) > v^{max})$ 
        remove  $T_i$ ;

if  $(type_i == Network)$ 
    if  $(Accu_i > 50)$ 
        remove  $T_i$ ;
end

```

(2) The same coordinates displayed on different maps will have different results. This is because the coordinate systems used by map service providers are different. Therefore, it is very necessary to understand the difference and conversion relationship between coordinate systems before data processing and analysis. The current map coordinate system is roughly divided into several categories:

1. The WGS-84 coordinate system (World Geodetic System-1984 Coordinate System), also the original GPS coordinate system, is an internationally recognized geographic coordinate system. It takes the center of mass of the earth as its origin, and the Z axis points to BIH (International Time Service Agency) The defined protocol Earth pole (CTP) direction, the X axis points to the intersection of the zero meridian of BIH 1984.0 and the Earth pole (CTP) equator, and the Y axis, X axis and Z axis constitute a right-handed coordinate system. This coordinate system is commonly used in international map services.

2. GCJ-02 coordinate system (G stands for Guojia country, C stands for Cehui surveying and mapping, J stands for Ju Bureau), also known as "Mars Coordinate System", is the coordinate system of the geographic information system formulated by the State Bureau of Surveying and Mapping of China. An encryption algorithm for latitude and longitude data, that is, adding random deviations. Most domestic map service providers use this coordinate system, such as AutoNavi Map and Tencent Map.

3. Other coordinate systems are generally calculated from Mars coordinates through a certain offset algorithm, such as the coordinate system BD-09 used by Baidu Maps.

In order to facilitate the later data processing, the coordinate system used by the positioning data source must be consistent with the map coordinate system, otherwise it will cause a large deviation in the later data analysis results. The conversion formulas and methods of each coordinate system have been referenced in many documents and materials. I will not repeat them here.

3 research methodology

This article uses Java Web's SSH (Spring+Struts+Hibernate) technology to implement a Web-based travel survey management system as a server-side data persistent storage

and travel survey process management and result analysis. SSH is three frameworks in the development of Java WEB, namely Spring, Struts and Hibernate. SSH is a Java Web application framework based on the MVC (Model-View-Controller) model.

The MVC framework is mainly divided into three major components, Model, View, and Controller modules [17]. Using these three components and the interfaces provided can separate business logic to a high degree, which reduces The degree of coupling between the functional modules of the system improves the portability and maintainability of the system.

As the main body of the system, Model mainly includes the encapsulation of business logic and data model to realize the storage and operation of business data. Since the code of the application model only needs to be written once, it can be reused by multiple views, which greatly improves the reusability of the system.

View is mainly used to provide a friendly user interaction interface, analyze and display the data obtained by the model, and expose the system access and operation interface to the user, and pass the user's operation access request to the control layer.

Controller is mainly used to manage the flow of the control system and process user requests. When a user's request is received, the request data is mapped to the model, and the corresponding view is selected to make a corresponding response according to the updated model data.

Struts in SSH provides corresponding components for Model, View and Controller. As a lightweight inversion of control and aspect-oriented container and framework, Spring can better solve the complexity of large-scale software development. Hibernate is an open source The object-relational mapping framework of code, as a universal data persistence framework, greatly reduces the tedious workload of adding, deleting, checking, modifying and coding.

At the mobile phone application level, the development platform of the iOS-based Apple mobile travel survey data collection system is used for application software APP development. And because the travel data collected by the application software APP needs to be transmitted online to the back-end server, this system uses the HTTPS protocol as the data communication protocol because HTTPS is more secure than HTTP. Data transmission is uniformly transmitted in JSON (JavaScript Object Notation) format.

JSON data mainly has the following characteristics. JSON is a lightweight text data exchange format, which is self-descriptive and easier to understand; JSON uses a completely language-independent text format and uses JavaScript syntax to describe data objects, but JSON is still independent Based on language and platform. The JSON parser and JSON library support many different programming languages. Because it

does not require strict closing tags like XML, the effective data volume and the total data packet ratio are greatly improved, and the pressure of network data transmission is reduced [15].

For system map display, this article adopts the map and location services of AutoNavi open platform services, such as AutoNavi Location API, AutoNavi Map API, etc.; in terms of system integration, iOS location API is used as the data source, and Web maps and mobile terminals are used. Map API is used as a map display, and real-time communication and inheritance between systems are realized by HTTP protocol and JSON as data communication methods.

As an iOS application development, the main implementation technical architecture of the system is as follows.

UI layer: Provide a friendly user interaction interface and mode.

Third-party library interfaces: such as AFNetworking library, SDWebImage library, AutoNavi map API, AutoNavi search service API, etc.

Multi-threaded HTTPS communication: Multi-threaded communication is realized for communication with back-end servers, data exchange and transmission, etc.

Controller layer: handles the core business logic of APP and is the bridge between the Model layer and the View layer.

View layer: displays the pages of mobile applications, including image rendering.

Model layer: Convert the Json data received from the server into a Model, and obtain the attributes of the data by parsing the Model data in the View layer to display it.

SQLite data local persistence: Persist data in the local storage of the mobile phone, reducing the communication pressure between the mobile phone and the server.

As a Web-side travel investigation management and result analysis system, it uses Java Web as the development platform, SQL Server2008R2 as the system database, and EasyUI and JQuery as the front-end interface display framework.

4 design and implementation

4.1 Recognition algorithm of stay point based on mobile phone accelerometer

4.1.1 Design of stop point identification process

Regardless of any activity, people always go from indoor to outdoor, or from outdoor to indoor. The positioning mode is also switched accordingly, and the stay points are also divided into indoor stays and outdoor stays. Indoor stays such as office work, rest at home, dining at a restaurant, etc.; outdoor stays such as transfer waiting, picking up people, and scenic spots Wait. This paper takes the above characteristics as the core idea of constructing the stop point recognition algorithm. The algorithm flow chart is shown in Figure 4-1. The method of classifying trajectory points is adopted, and the trajectory is divided into two states of staying and moving, which are the subsequent travel characteristics. Extraction provides basic semantics.

Since human activity in the real world is a continuous state, and the data collected in this article is a discrete sampling of human travel activities, the analysis of each data point cannot be divided into consideration, and the continuity of activities should be characterized to the greatest extent. In other words, the collected discrete positioning data can be restored into a continuous travel trajectory. Therefore, this paper adopts a sliding window algorithm, the reference point of the window is the current point, the size of the window is all the data points from the current point to Δt before the current point, and Δt is the time threshold of the stay point judgment above. At the same time define a set of global parameters to identify the current state, *state* is the state the current user is in (stay stop or move move), $t^{laststop}$ is the end time of the last stay point;

traceList is the collection of all collection points in the window, *gpsList* is the collection of all collection points in the window whose positioning type is GPS, and *stopList* is the set of stay points, which is the output of the algorithm.

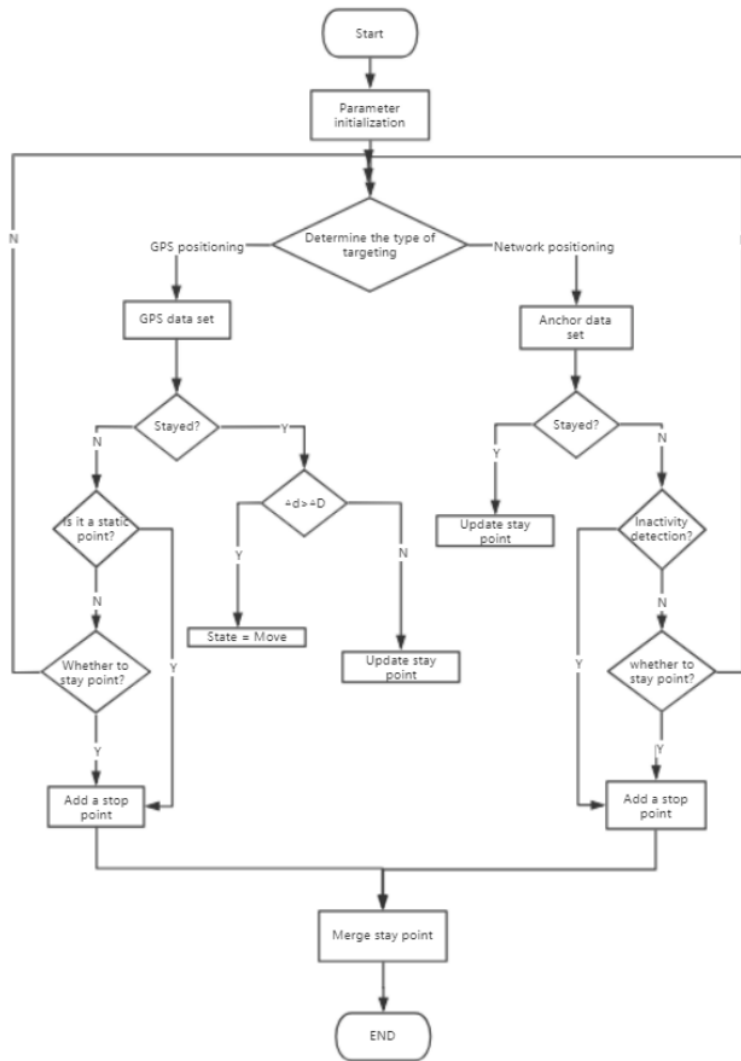


Figure 4-1 Flow chart of stop point recognition algorithm

The algorithm flow of the entire stay point is as follows:

Step1. Cold start, parameter initialization, *traceList* is the collection of all anchor point data, *gpsList* is the collection of GPS data points, state is the state, and *lastStop* is the end time of the last stay;

Step2. Put the current point into the *traceList* to determine the current positioning type, if it is GPS, go to step 3, if it is network positioning, go to step 6;

Step3. Put the current point into the *gpsList*, judge the current state *state*, if it is stop, judge whether the distance between the current point and the last stay point is greater than Δd , if it is, the current state is from stop->move, otherwise update the *lastStop* time, repeat 2 ; If the current status is move, go to step 4;

Step4. Run acceleration static detection on *traceList*, if it is judged to be static, execute 9; otherwise, execute 5;

Step5. Run GPS outdoor stay detection on *gpsList*, if it is judged to be staying, execute 9; otherwise, repeat 2;

Step6. Determine the current *state*, if it is stop, update the *lastStop* time; if the current state is move, execute 7;

Step7. Run acceleration static detection on *traceList*. If it is judged to be static, execute 9; otherwise, execute 8;

Step8. Run Network indoor stay detection on *traceList*. If it is judged to be staying, State will be moved->stop, update the *lastStop* time, and execute 9, otherwise repeat 2.

Step9. Add new stay points, merge stay points, repeat 2.

4.1.2 Algorithm implementation

When the APP system is initialized and started, since there is no data, it needs to wait for Δt time for cold start. In this process, stop point detection is not performed, only the data is filtered and saved. After Δt time is started, the stop point recognition operation based on the mobile phone's multi-motion parameters is performed on the data set in a sliding window. In the whole process, the definition of global variables used in the implementation of this algorithm is declared. *TraceList* represents the collection of all collection points in the current algorithm operation time window; *gpsList* represents the collection of all GPS positioning types of collection points in the current algorithm operation time window; *stopList* Represents the set of stay points, that is, the final output result of the entire method.

Table 4-1 Pseudo code of acceleration static detection algorithm

Input: the current point location, $p_i = \{l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i\}$		Output: true or false
Initialize: <i>traceList</i> = { <i>p_i</i> }		
For(int i=1; i<=n; i++)	//Divide the trajectory into n segments, the number of	
	//points in each segment is m	
For(int j=0; j<m; j++)		
Sum+=a _j ;		
end		
a=sum/j;	//Calculate the average acceleration of each data set	

```

    accAver.add(a);           //accAver is the aggregate variable of the acceleration
                             //average value in the segment
End
Sum=0;
For(int i=1; i<=n;i++)      //Calculate the acceleration standard deviation of each
                             //data set
    For(int j=0;j<m;j++)
        Sum+=Math.pow(aj-a, 2);
    end
    std=Math.sqrt(sum/m);
    accStd.add(std);         //accStd is the set variable of the standard deviation of
                             //each acceleration
End
Array.sorts(accStd);        //Array.sort sorts the entire collection
if accStd.get(0.75*n)<0.057
    Return true
Else
    Return false;

```

The core part of the whole method realization process is mainly divided into three parts, which are the realization of the static detection algorithm based on accelerometer, the realization of the improved Cluster stay point recognition algorithm and the realization of the network positioning stay point algorithm based on empirical rules.

As shown in Table 4-1, it is the pseudo code of the specific realization process of the accelerometer-based static detection algorithm, where *accAver* is the set variable of the acceleration average value in the segment, and *accStd* is the set variable of the acceleration standard deviation of each segment, *Array.sort()* The method is to sort the set. The third quartile is the sample point at the 75% position in the ordered set. If it is judged to be static, the program returns "true", otherwise it returns "false".

In real life, users are often not in a static state when staying in certain situations, such as pacing while waiting for a bus, wandering in parks or scenic spots, etc. In these cases, the static detection algorithm based on accelerometers often fails, but in these cases The real stay of the user can be detected on the characteristic attributes such as time and space. Based on the two types of positioning data, the idea of "divide and conquer" is adopted to identify staying points. As shown in Table 4-2, under GPS positioning, a pseudo code based on the improved Cluster stop point recognition algorithm is adopted.

Table 4-2 Pseudo code based on the improved Cluster stop point recognition algorithm

Input: the current point location , $p_i = \{l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i\}$ **Output:** stopList
Initialize: gpsList= $\{p|p.type=0\}$,currentState ,lastStopTime, i;

```

if(currentState==move)
begin
    //When the status is "moving", do stay detection
    if  $t_i - t_{laststop} < \Delta t$  //Cold start to wait
        exit;
    else
        if( $Count(gpsList, \Delta d, \Delta t) > m \ \&\& \ P_{80}(v_i) < v_{walk}$ )
            curentState=stop;
            stopList.add( $T_i$ );
        else
            exit;
    else
        //When the state is "stationary", perform //motion
        //detection
        begin
            if( $Distance(l_i, stopList.lastone) > \Delta d$ ) //Judge whether the current point is greater than
                //  $\Delta d$  from the last stay point
                currentState=move;
            else
                lastStopTime =  $t_i$ ;
        end
end

```

In Table 4-2, *currentState* represents the current state of motion, and the method *Count* (*gpsList*, Δd , Δt) represents the current point as the center, and counts the number of points within Δt in the GPS data set *gpsList* that are less than Δd from the current point. $P_{80}(v_i)$ represents the 80th percentile based on the speed value within the Δt time in the calculation data set *gpsList*.

Distance($l_i, stopList.lastone$) means to calculate the spatial distance between the current point l_i and the last stop point.

Table 4-3 shows the pseudo-code of the stay point recognition algorithm based on empirical rules proposed in this paper under network positioning. The variable *c* represents the degree of accuracy change defined above, and *Math.abs()* is the absolute value function.

Table 4-3 Pseudo-code of the recognition algorithm of network positioning staying point based on empirical rules

Input: the current point location , $T_i=\{l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i\}$ **Output:** stopList={ S_i }
Initialize: traceList={ p_i } currentState,lastStopTime

```

if(currentState==stop)
begin
    lastStopTime = ti;
    exit;
end
else
    For(i to j)
        if(type==NETWORK)
            count++;
            c+=Math.abs(Accui-Accui-1)
        if(count/n>0.8 && c<15)
            currenState=stop;
            stopList.add(Ti);
        else
            exit;
    
```

4.2 Feature extraction for travel survey

After the processing of the stay point recognition algorithm in the previous section, the stay point data set generated by the user and the corresponding stay point start and end time can be obtained in this system, and the basic semantic fields of the user's movement state during travel. On this basis, a user's personal travel characteristics model can be established. By extracting a variety of travel characteristic indicators, it is possible to control the travel characteristics of urban residents at a macro level, and provide data support for subsequent urban traffic travel analysis and urban hotspot analysis.

4.2.1 Personal travel characteristics model

In the tens of thousands of years of historical evolution of human beings, a clear day and night schedule has been formed, and travel data also reflects the typical characteristics of taking the sky as the basic unit. This article establishes a personal travel model with the granularity of the day, and analyzes the characteristics of travel.

Table 4-4 Personal travel record parameters

ID	Name	Abbreviation	Field	Description
1	Record number	I	i	Characterize the travel information of a

				certain user on a certain day
2	User ID	User	u	Characterize the travel user to which the record belongs
3	Plan ID	Plan	p	Characterize the travel survey to which the record belongs
4	Record date	Date	d	Characterize the date the trip occurred
5	Track points	Number	n	Characterize all track point data of the travel record
6	Trip mileage	Distance	dis	Indicates the total travel distance of the user on a certain day
7	Trip time	Duration	dur	Indicates the total travel time of the user on a certain day

Therefore, the personal travel model in this article is a semantic representation of all the original travel data of a person on a certain day, which is the most simplified form of feature extraction. The basic travel features include travel volume, travel duration, time distribution, spatial distribution and travel. Way etc.

Travel record: The basic element that can characterize the travel characteristics of an individual in a day is a series of parameters calculated from the space-time point data of a certain user on a certain day. This article uses R to represent the travel record (Record). For specific parameter definitions, see Table 4-4, the mathematical model is shown in formula (4-10).

$$R_i = (u_i, p_i, d_i, n_i, Dis_i, Dur_i) \quad (4-10)$$

R_i corresponds to the travel record of the user u_i on a day with the date d_i , and corresponds to the unique set of track points $T_p \cdot R_i$, whose mathematical expression is shown in formula (4-11).

$$T_p \cdot R_i = \{p_j \mid p_j \cdot u_j = r_i \cdot u_i, r_i \cdot d_i < p_j \cdot t_j < r_i \cdot d_i + 1, j \in I\} \quad (4-11)$$

Here the track point set $T_p \cdot R_i$ is all track points where the user u_j of the track point is equal to $r_i \cdot u_i$ and the track point time $p_j \cdot t_j$ is between the time $r_i \cdot d_i$ and $r_i \cdot d_i + 1$.

4.2.2 Judgment of travel effectiveness

Judging the effectiveness of travel is the most important part of judging travel characteristic indicators. For a completely stationary user, all the track points at this time are all stay points, this kind of data is invalid, and the corresponding trip is also

invalid. This article defines the conditions for valid travel as the travel distance exceeds 500m and the number of moving points in the trajectory data is at least 60. When the user's travel distance is short and does not exceed the set threshold. Or when the travel trajectory point data is seriously missing, this article believes that this part of mobile phone users cannot reflect the user's true travel status, and will be deleted as invalid data in the subsequent travel feature extraction step.

4.2.3 Travel distance

First, for individual users, calculate the daily travel distance of individual users. Zhang Wei [1] analyzed the distance between users' stay points. Although a certain amount of calculation is reduced, the final calculation result has a large error. A user's personal travel record corresponds to a collection of the user's travel trajectory points on a certain day. On this basis, this article accumulates the adjacent trajectory points to obtain the final travel distance. First, for the point p_i , calculate the distance formula between it and the adjacent point, see formula (4-12).

$$D \cdot p_i = r * \pi / 180 * \sqrt{(p_i \cdot lat_i - p_{i+1} \cdot lat_{i+1})^2 + (p_i \cdot lon_i - p_{i+1} \cdot lon_{i+1})^2} \quad (4-12)$$

$D \cdot p_i$ represents the distance between the trajectory point p_i and p_{i+1} , r is the radius of the earth taking 6371km, and π is the p_i .

For any travel record, the travel distance in one day can be calculated, see formula (4-13).

$$D \cdot R_i = \sum D \cdot p_j, p_j \in T_p \cdot R_i \quad (4-13)$$

For all travel records, the average travel distance per person per day can be obtained, see formula (4-14).

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D \cdot R_i \quad (4-14)$$

Where n is the number of travel records.

4.2.4 Travel time and time distribution

Resident travel time can be used not only as one of the measurement methods of urban travel volume, but also as a reward indicator for participation in transportation surveys. In a sense, the travel time distribution can be used as a guide and support data for transportation planning and construction. Resident travel time is the total length of travel time of a resident on the day. Travel time distribution refers to the distribution of residents' daily travel volume in different time periods. Usually one hour is used as the

basic unit to calculate the travel volume during the time period, and the calculation is The travel occurrence rate of each period, that is, the travel volume within the segment divided by the travel volume during the whole day, to obtain the travel time distribution throughout the day.

This article uses the state field to segment the motion state of the track point, $p_i \cdot s_i = 1$ means the stay point, $p_i \cdot s_i = 0$ Indicates a moving point. At the same time, the trajectory point collection cycle T of the travel data collection system in this article is 5s, and the travel time can be calculated by the number of moving points, see formula (4-15).

$$duration = T * n \quad (4-15)$$

T is the sampling period of track points, and n is the number of motion track points.

The state field can be used to count the user's personal travel time distribution, the algorithm is as follows.

1. For the user's travel record R_i on a certain day, filter out the set of mobile points among them, denoted as C, then C can be expressed as formula (4-16).

$$C = \{p_j \mid p_j \cdot s_j = 0, p_j \in T_p \cdot R_i, 1 < j \leq T_p \cdot R_i\} \quad (4-16)$$

2. This article defines the one-hour period with each time as the starting point as q, $q=0$, that is, the period from 0:00 to 1: $0 < q < 24$, the travel movement points of each period are counted separately, then each The set of moving points in each period is E, see formula (4-17).

$$E_q = \{p_j \mid q < p_j \cdot t_i < q+1, p_j \in C\} \quad (4-17)$$

Then the travel distribution at any time of the day in a travel record can be obtained by formula (4-18).

$$H_q = \frac{|E_q|}{|C|}, 0 < q < 24 \quad (4-18)$$

Among them, H_q is the probability of trip occurrence in period q, |C| is the number of mobile points in a day, and $|E_q|$ is the number of mobile points generated in period q.

Through the travel time distribution of individual users, the overall travel time distribution of all users participating in the travel survey can also be calculated.

4.2.5 Travel speed

Resident travel speed is an important indicator to evaluate the service level of urban travel vehicles. Although the user can obtain the speed value measured by GPS during the trip, the GPS signal will be blocked during the city trip, and accurate data cannot be

obtained. Therefore, this article will compare the user's first movement point and The last movement point is used as the starting point and ending point of the trip to calculate the average travel speed of the user. For a certain travel record R_i , the travel speed calculation formula is as (4-19).

$$v_i = \frac{r * \pi / 180 * \sqrt{(Lat_e - Lat_s)^2 + (Lon_e - Lon_s)^2}}{t_e - t_s} \quad (4-19)$$

Among them, t_e , Lat_e , and Lon_e are the time, latitude and longitude of the last movement point in the travel record, and t_s , Lat_s , Lon_s are the time, latitude and longitude of the first movement point in the travel record. r is the radius of the earth. At the same time, the average travel speed of the travel records used can be obtained as formula (4-20):

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i \quad (4-20)$$

Among them, n is the number of travel records

4.3 Hot spot discovery

The algorithm that the hotspot area discovery mainly relies on is the centroid distance AGNES algorithm. The process is to first assume that there is a data set of n objects, treat the n objects as n initial clusters, and calculate the clusters between the n clusters. The distance d is to find the two clusters with the smallest distance. If the two clusters with the smallest distance comply with the cohesion rule, the two clusters are clustered into one cluster and become a new cluster. Then replace the merged two clusters with other original clusters with new clusters to recalculate the distance d between each cluster, update the distance between each cluster, and find the two clusters with the smallest distance between the clusters again. If the cohesion rule is met, the two clusters are merged. The new cluster replaces the merged two clusters and the original cluster is recalculated to update the distance d between each cluster. And so on, until the cluster stop condition is reached. The specific algorithm steps are as follows:

First, define each cluster in the cluster as Cluster c , which is a collection of points. Define the centroid of the cluster as the point with the smallest sum of distances to all other points in the cluster, and define the similar distance between two clusters. Is the spatial distance between the centroid points of two clusters, the radius of the cluster is defined as the maximum of the distances from the centroid point of the cluster to other points in the cluster, and ClusterSet is defined as the set C of clusters, that is, the

algorithm operation The object is also the final output result of the algorithm, as in formula (4-21).

$$C = \{c_i\} \quad (4-21)$$

Step1. Initialize each stay point as a cluster cluster c , and put it into the cluster set C , so the size of C during initialization is equal to the number of stay points;

Step2. Traverse the cluster set $ClusterSet$, find the closest two clusters $Cluster A$ and $Cluster B$, merge them in advance to get $Cluster AB$, calculate the radius of the merged cluster AB , and confirm to merge them if it is less than the set threshold Put $Cluster AB$ as a new cluster into the $ClusterSet$ and delete $ClusterA$ and $ClusterB$ to proceed to the next step. Otherwise, the algorithm ends, and the current cluster set is output as the final result.

Step3. Repeat step 2 until no new clusters can be synthesized;

Step4. After all the above steps are executed, the clustering process is complete. The final clustering result of the urban travel hotspot area is in the $ClusterSet$ object, and each $Cluster$ object represents the corresponding hotspot. The specific algorithm analysis steps are shown in Figure 4-2, and the implementation of the algorithm for urban travel hot spot discovery is shown in the pseudo code of Table 4-5.

Table 4-5 Hotspot discovery algorithm pseudocode

```

Input : Stop point set,  $S = \{p_i \mid p_i \cdot s_i = 1, p_i \in T_p\}$ 
Output : Hot class cluster,  $C = \{c_i\}$ 
Initialize :  $c_i \leftarrow p_i$ , for  $1 \leq i \leq S, C = \{c_i\}$ 
Loop
   $(c_i, c_j) \leftarrow \text{FindClosestPair}(C)$ ;
  If  $\text{Radius}(c_i, c_j) \leq d$ 
     $c_i \leftarrow \text{Merge}(c_i, c_j)$ ;
     $C \leftarrow C - c_j$ ;
  Else
    Exit
End

```

The FindClosestPair method is to find the closest two clusters c_i, c_j in the cluster set C ; $\text{Radius}(c_i, c_j)$ is to calculate the radius value of the cluster after c_i and c_j are merged; $\text{Merge}(c_i, c_j)$ merges the clusters c_i and c_j .

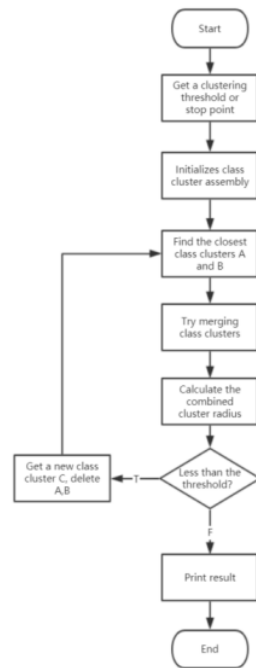


Figure 4-2 AGNES algorithm flow based on centroid distance

4.4 System Requirements Analysis

Urban residents travel survey is the core content of traffic survey, and its results can characterize the characteristics of residents' travel composition, flow and demand, comprehensively analyze the social and economic attributes of residents, find out the problems of urban residents' travel, and understand residents' travel needs, To control the crux of urban traffic travel, and provide scientific basis for urban traffic status analysis and planning management.

The economic cost, time cost, and labor cost required by traditional survey methods are too high. The survey results are easily affected by the content of the questionnaire and the subjective cognitive ability and memory of the respondents. At the same time, it will cause a greater burden on the life of the respondents. Many problems such as the difficulty of recycling and sorting cannot meet the needs of modern urban transportation development. Therefore, designing a comprehensive system that can collect, manage and analyze residents' travel data efficiently and at low cost is an urgent problem that needs to be solved.

Currently, the two mainstream mobile phone operating system platforms are the iOS system developed by Apple and the open source Android system of Google. Third-party developers develop specific system applications based on the C/S architecture model on the basis of both. It is characterized by high user viscosity,

convenient use, and low construction cost. Therefore, for data collection, this article will use C/S architecture with other key technologies to implement a travel data collection system based on the mobile phone of the surveyed person.

With the popularization of office automation, for travel investigators, the pre-travel survey management and the post-travel data sorting and mining are no longer cumbersome. Based on the B/S (Browser/Server) architecture model, compared to C/S Higher functional scalability and cross-platform, so this article will use the B/S-based architecture to realize the management function of the travel survey system.

To sum up, the requirements of the mobile phone-based travel survey stop point identification and feature extraction system mainly include:

(1) Move the traditional offline survey process online, combine the advantages of smart phones and mobile Internet to complete a mobile phone APP-based travel survey data collection system, and use the convenience of the mobile Internet to realize mobile travel data collection and The cleaning and sorting work will transform the traditional active survey work method into a passive survey with mobile phones as the core. While reducing the burden on the interviewees, the accuracy of later data and management analysis can also be better controlled.

(2) For investigators, the traditional survey method requires the recovery of the questionnaire, and the later manual sorting of the telephone interview. This article uses the form of online office to conduct online management of travel surveys. Based on the survey, the survey initiator and the respondent are associated, and the data management of different survey initiators in different regions is isolated. At the same time, the entire investigation process is based on "The thought management of "workflow", such as data review, data validity analysis, investigation reward management, etc., makes the entire travel investigation more convenient.

(3) When collecting data on the mobile phone, perform semantic analysis based on the user's travel trajectory data to extract the user's travel characteristics, such as staying point, travel distance, and travel mode.

(4) In terms of system implementation architecture platform, factors such as system practicability, construction cost and operational stability need to be considered. Therefore, it is necessary to adopt a hybrid architecture of B/S and C/S to realize an intuitive, stable, and cross-platform travel survey system.

4.5 System Design

The system is mainly divided into two major sub-systems, namely, a smart phone-based mass travel behavior information collection system (mainly for iOS operating system smart phones) and a WEB-based travel behavior investigation

management system (mainly manages and analyzes the information uploaded to the server data).

Among them, the public travel behavior survey app based on smart phones is mainly divided into the following functional modules.

(1) Basic user functions

User basic functions mainly include app general functions, such as user registration, user login, user logout, and user retrieve password. User basic functions are mainly used to identify the user's identity and facilitate specific analysis later.

(2) Data collection

The data collection function is mainly to collect the original trajectory data through the mobile phone sensor. Because there may be noisy track points, the data collection function cleans the travel data to ensure the reliability and validity of the travel data.

(3) Generation of stay point

Using the stay point recognition algorithm proposed above, the corresponding stay track points are generated through the mobile app, and the stay points are marked and uploaded to the server.

The Web-based travel survey process management and travel characteristic analysis system mainly includes the following functional modules:

(1) System management

System management mainly includes two aspects, user role authority management and system parameter configuration. User role authority management is used to control system user registration, login, access, operation and other authorities, and is a guarantee for system security and functional isolation. System configuration is the dynamic configuration of the necessary parameters for system operation, mainly the setting and modification of key parameters during system operation, ensuring stable operation of the system and global editing management.

(2) Process management of travel investigation

The management of travel survey process is mainly divided into survey plan management and survey result management.

Investigation plan management is mainly for investigation administrators, used to initiate investigations, review investigations, and maintain investigation functions; investigation result management is mainly used to view the data results of the investigation, and to review and analyze the data results.

(3) User travel feature extraction and travel area hot spot discovery

User feature extraction is to analyze the characteristics of travel data for all surveyed persons, and the results are used for later transportation planning and construction and urban travel data analysis. Hot spot discovery is a global analysis of the travel data of all travel respondents in a certain travel area, and a certain clustering algorithm is used to calculate the travel hot spots in the travel area.

4.6 Overall system architecture

From the mobile phone-based travel survey stop point identification and feature extraction system requirements and functional structure analysis, the constructed system mainly includes travel data acquisition based on mobile phone multi-sensor parameters, system operation and algorithm parameter configuration, data cleaning based on empirical rules, Real-time stop point generation method and travel feature extraction. The overall architecture of the mobile phone-based travel survey stop point recognition and feature extraction system is shown in Figure 4-2.

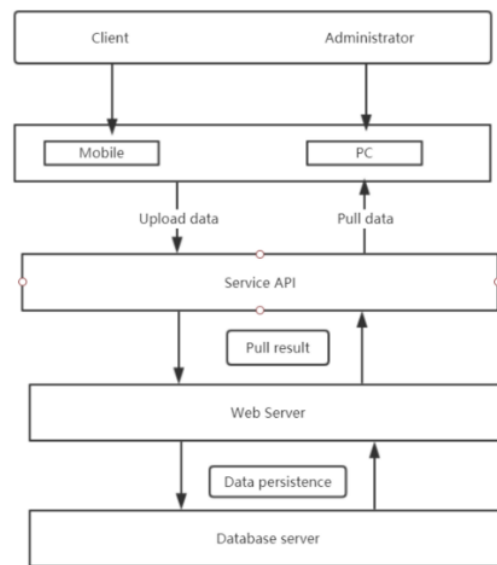


Figure 4-2 System overall architecture diagram

First of all, the key to the operation of the system is to obtain the positioning data and acceleration sensor data of the mobile phone user. When these two parts of data are available, it is guaranteed that we can obtain the mobile phone travel trajectory data stably. Secondly, we can make full use of the characteristics of each sensor of the smart phone to construct a mobile phone travel trajectory data model. Use the method based on empirical rules to clean the original trajectory data to obtain data with less noise,

which is more conducive to later data analysis. Through the method for identifying staying points based on mobile phone multi-motion parameters proposed in this article, staying points are automatically generated in the mobile app and stored in the iOS device sandbox. Users can upload the currently stored staying points to the server by themselves for subsequent follow-up Research and survey.

In addition, the server-side survey administrator can also initiate a survey plan. Mobile users can join the survey plan through the mobile app. This avoids the tedious steps of face-to-face communication in traditional survey plans, and users can join the plan faster , And more voluntary.

4.7 Design Modeling of System Realization Process

Based on UML

UML visualizes the association of various object models, so that software designers and developers can be more clearly familiar with software processes and product models, improving work efficiency and reducing development costs.

According to the needs and implementation process analysis of the mobile phone-based travel survey stop point identification and feature extraction system, this section mainly uses sequence diagrams, class diagrams and component diagrams to design and model the system.

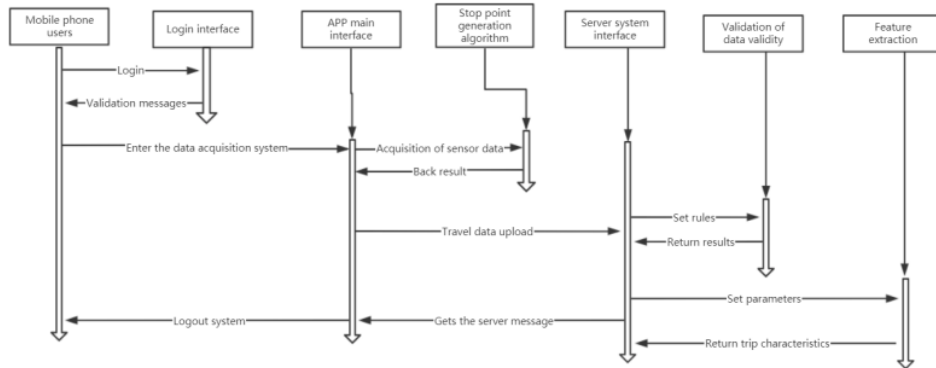


Figure 4-3 System timing diagram

The use of sequence diagrams to model the system process can effectively describe the functions of the various modules of the system, which is conducive to the grasp of the workflow of the entire travel survey system in this article. The specific timing diagram is shown in Figure 4-4. The user enters the user name and password through the mobile APP, and logs in to the system through background authentication. After the authentication is passed, the main interface of the APP will be entered. At this time, the APP actively obtains the sensor data, encapsulates the model and uploads it to the

server, and runs the stop point identification method to actively identify the user The stop point during the trip. The server-side system sets corresponding parameters and rules according to the received travel data, and then performs data validity review and feature extraction services.

The UML diagram of the system corresponds strictly to the design of the actual system database. User User and Trace have a "one-to-many" relationship. The survey plan Plan and User User have a "one-to-many" relationship. User User and travel record Record are "one-to-many". The relationship between User, Role, and Node is a "many-to-many" relationship.

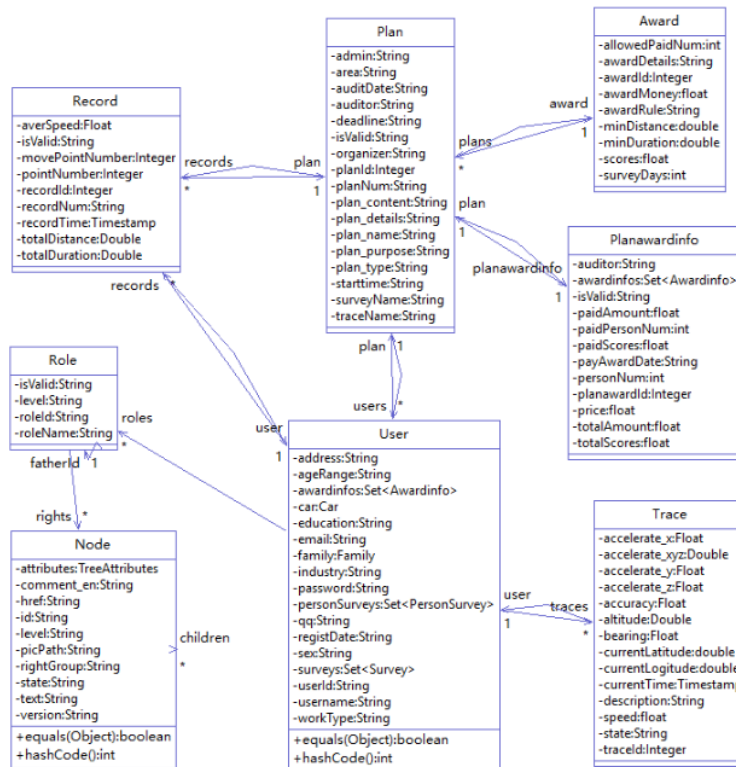
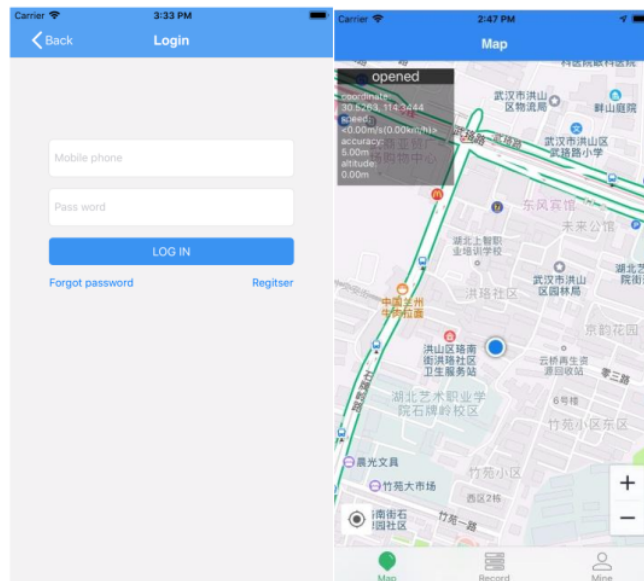


Figure 4-4 System UML diagram

4.8 Implementation of the recognition method based on mobile phone multi-motion parameters

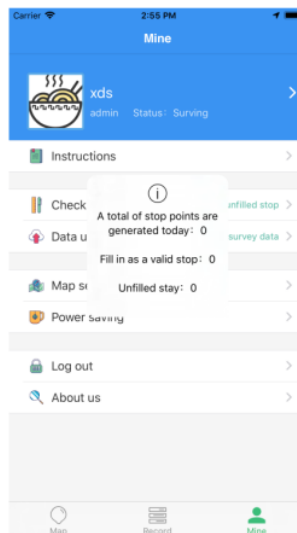
The process of stopping point identification is in the mobile app. The mobile app is used as a carrier to collect and manage user travel data, generate stopping points in real time and display them in the APP in a user-friendly interactive manner.

First, when using the app, the user needs to log in and enter the map location interface. At this time, the user's travel trajectory and stay points have been collected as shown in Figure 4-5.



4-5 User login and data collection

After the user enters the app, he can upload the data collected at that time on the personal information page. You can also select to check the collected data by pressing the button, and you can view the number of current stop points and upload status as shown in Figure 4-6.



4-6 Check the number of stay points collected

When the user runs the program in the background and exercises, the corresponding trajectory will be generated. The stop point recognition algorithm runs in the APP background in real time to determine the user's travel stop point. When the algorithm generates a stop point, it will pass on the map interface Marked out in the form of a red bubble icon. At the same time, users on the WEB side can also manage their travel records by logging in to their account.

4.9 Reality of the Web-based travel survey process

First, we need to log into the system management page, as shown in Figure 4-7 below. The Username and password are both 8888.

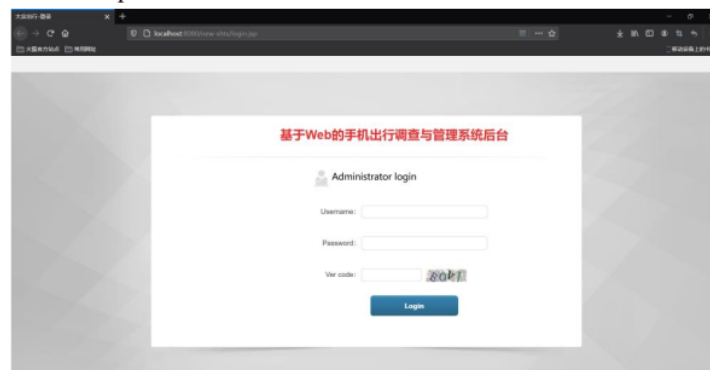


Figure 4-7 Administrator login interface

After the survey administrator logs into the system as an administrator, he enters the main work panel of travel survey management, and can view and manage the survey plan initiated by him, as shown in Figure 4-8.

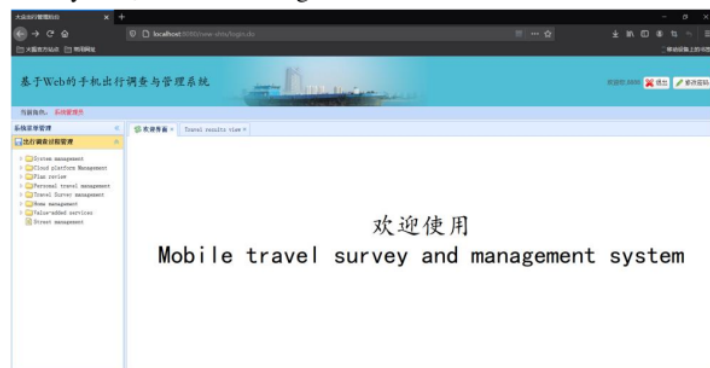


Figure 4-8 Web-based travel survey process management main panel

The entire system includes personal travel management, travel survey plan management, and travel process management. Respectively complete the inquiry of

personal travel information, the formulation and maintenance of travel plans, and the review and analysis of travel process information.

Regarding the personal travel characteristic model in the system, the administrator can view and manage it under the survey result management node, as shown in Figure 4-9. It is mainly used to view the travel records of all users in the corresponding survey plan, that is, the personal travel characteristic model. The travel records mainly record the daily travel characteristic information of all users under the survey, such as "account", "name", and "accumulated travel mileage on the day", "Accumulated travel incidents of the day" and "Results of preliminary screening of procedures", etc. The initial screening of the program is to judge the validity of travel. It is automatically judged based on the threshold parameters of "minimum accumulated travel mileage" and "minimum accumulated travel time" in the reward rules set by each survey. When the user's accumulated travel mileage and travel time on that day are greater than When the target value is reached, it is regarded as a valid travel record and passed, otherwise it is not passed.



Figure 4-9 Implementation of travel feature model

Furthermore, you can view the corresponding map trajectory diagram of each record to provide the survey administrator with guarantee for the validity and completeness of all travel records of all users under the survey plan, as shown in Figure 4-10, which shows a certain Travel trajectory diagram of travel records.

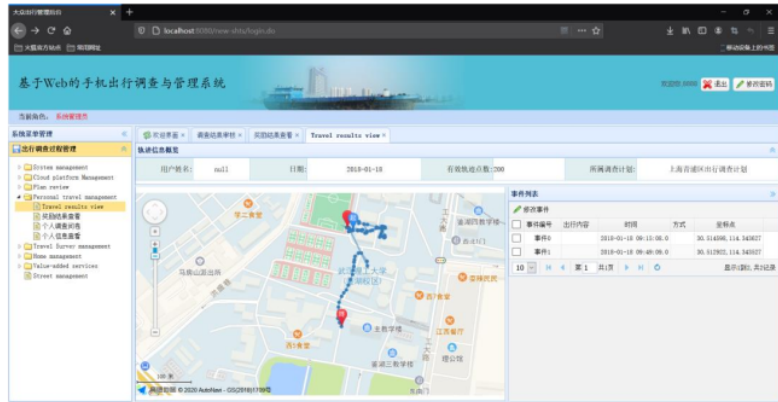


Figure 4-10 View the detailed trajectory of travel records

5 Results

5.1 Stop point recognition result

When the APP system is initialized and started, since there is no data, it needs to wait for Δt time for cold start. In this process, the stay point detection is not performed, only the data is filtered and saved, and after the Δt time is started, the stay point recognition operation based on the mobile phone's multi-motion parameters is performed on the data set in a sliding window. In the whole process, the definition of global variables used in the implementation of this algorithm is declared. TraceList represents the collection of all collection points in the current algorithm operation time window; gpsList represents the collection of all GPS positioning types of collection points in the current algorithm operation time window; stopList Represents the set of stay points, that is, the final output result of the entire method.

Table 5-1 Pseudo code of acceleration static detection algorithm

```

Input: the current point location,  $p_i = \{l_i, t_i, alt_i, v_i, Accu_i, a_i, s_i, type_i\}$  Output: true or false
Initialize: traceList= $\{p_i\}$ 

For(int i=1; i<=n;i++)           //Divide the trajectory into n segments, the number of
                                points in each segment is m

    For(int j=0;j<m;j++)
        Sum+=aj;
    end
    a=sum/j;                     //Calculate the average acceleration of each data set
    accAver.add(a);              //accAver is the aggregate variable of the acceleration
                                average value in the segment

End
Sum=0;
For(int i=1; i<=n;i++)           //Calculate the acceleration standard deviation of each data
                                set

    For(int j=0;j<m;j++)
        Sum+=Math.pow(aj-a, 2);
    end
    std=Math.sqrt(sum/m);
    accStd.add(std);             //accStd is the set variable of the acceleration standard
                                deviation of each segment

End
Array.sorts(accStd);            //Array.sort sorts the entire collection
if accStd.get(0.75*n)<0.057
    Return true
Else
    Return false;

```

As shown in Table 5-1, it is the pseudo code of the specific realization process of the accelerometer-based static detection algorithm, where accAver is the set variable of the acceleration average value in the segment, and accStd is the set variable of the acceleration standard deviation of each segment, Array.() The method is to sort the set. The third quartile is the sample point at the 75% position in the ordered set. If it is judged to be a static state, the program returns "true", otherwise it returns "false".

In order to use the GPS module of the smart phone and other sensor modules to collect experimental data to verify the algorithm designed in this paper, this paper developed a mass travel data collection system based on the iOS platform during the research process. The APP collection system collects and uploads sensor data once in a period of 5 seconds. The dwell point identification method performs sliding window operation on each current collection point, and the time window size is 100 sample points.

The software APP running interface is shown in Figure 5-1. The entire system has collected 129 volunteers and a total of more than 300,000 trajectory data. Figure 4-3 shows part of the data.



Figure 5-1 Travel survey data collection APP interface

#	A	B	C	D	E	F	G	H	I	J	K	L	M
1	traceId	Latitude	Longitude	currentTime	speed	accuracy	adcode	accelerate x	accelerate y	accelerate z	accelerate xyz	altitude	bearing
2	305875	30.513279	114.344772	19:52:41	0	25	420111	0.002961814	0.003419876	-0.004239082	0.006199816	0	0
3	305876	30.513279	114.344772	19:52:43	0	25	420111	-0.018413723	0.011844158	-0.023480415	0.032104193	0	0
4	305877	30.513279	114.344772	19:52:45	0	25	420111	0.009797394	-0.00329835	-0.004115919	0.011329461	0	0
5	305878	30.513279	114.344772	19:52:53	0	25	420111	0.024651825	-0.013289928	0.010437965	0.029887887	0	0
6	305879	30.513279	114.344772	19:53:02	0	25	420111	0.013593614	-0.003455639	0.008728981	0.016520378	0	0
7	305880	30.513279	114.344772	19:53:03	0	25	420111	-0.000561118	0.014900684	-0.000205994	0.014912669	0	0
8	305881	30.513279	114.344772	19:53:09	0	25	420111	0.007661104	0.004004955	-0.007184029	0.011240256	0	0
9	305882	30.513279	114.344772	19:53:13	0	25	420111	0.011244118	0.011073112	0.018395424	0.024237072	0	0
10	305883	30.51329882	114.3447428	19:53:14	0.94	131		0.002309054	0.006462124	-0.004617691	0.008289901	22	188.1999969
11	305884	30.5132211	114.3447179	19:53:20	0.27	117		0.054771751	-0.013864203	-0.014677048	0.058403066	23	39.5
12	305885	30.5132266	114.3447176	19:53:26	0.902	198		-0.00074599	0.012175063	-0.012482043	0.017439164	24	41.29999924
13	305886	30.51328993	114.3446696	19:53:32	0.831	192		-0.001870573	-0.006516933	0.014318466	0.015842599	23	43.29999924
14	305887	30.51327501	114.3446796	19:53:37	1.36	180		0.036964715	-0.018132687	0.00939640	0.042231689	23	44.90000153
15	305888	30.51322428	114.3448315	19:53:49	0.827	229		-0.051434577	0.007339954	-0.047262192	0.070236068	23	48.90000153
16	305889	30.51324029	114.3447076	19:53:49	1.201	168		0.012056708	-0.039475918	-0.021780968	0.046670363	23	46.5
17	305890	30.51329508	114.3448503	19:53:56	2.301	194		-0.01728186	0.027519703	-0.082968712	0.089105575	23	50
18	305891	30.51333984	114.3447062	19:54:02	1.142	187		-0.021134678	-0.006900311	0.017457008	0.028267223	23	51.90000153
19	305892	30.51327365	114.3447512	19:54:07	1.126	215		0.006519556	0.013077296	0.00845623	0.016879423	23	52.90000153
20	305893	30.513279	114.344772	19:54:20	0	25	420111	0.029788375	-0.001455784	-0.009991936	0.030418892	0	0
21	305894	30.5132137	114.3447708	19:54:20	0.52	154		-0.092361569	-0.000804983	0.004824038	0.09249154	24	55.70000076
22	305895	30.5132468	114.3447862	19:54:33	0.645	145		0.01013422	0.006015778	-0.002943039	0.012147159	25	60
23	305896	30.51323134	114.3447561	19:54:39	0.36	240		0.004502356	0.006421089	0.016470909	0.0182426	26	61.40000153
24	305897	30.513279	114.344772	19:54:54	0	25	420111	-0.028235316	-0.021855354	0.014462471	0.038523404	0	0
25	305898	30.513279	114.344772	19:54:54	0	25	420111	-0.003978014	0.024552822	-0.020555496	0.032267539	0	0
26	305899	30.513279	114.344772	19:55:00	0	25	420111	-0.028023899	-0.004102707	0.00153923	0.02836442	0	0
27	305900	30.51324816	114.344719	19:55:05	1.04	147		0.006132994	0.019912243	0.01895997	0.028108845	28	352.1000061
28	305901	30.51326989	114.3447423	19:55:10	1.284	246		0.029021502	-0.008618832	-0.019774437	0.036160201	28	69
29	305902	30.513279	114.344772	19:55:21	0	25	420111	0.028814256	-0.006354332	0.015677452	0.033412893	0	0
30	305903	30.513279	114.344772	19:55:26	0	25	420111	0.029759467	-0.024071693	0.022320747	0.044309007	0	0
31	305904	30.513279	114.344772	19:55:31	0	25	420111	-0.038610041	0.031173706	-0.021551132	0.054101631	0	0
32	305905	30.51329801	114.3445635	19:55:35	1.49	148		0.003600597	-0.005080223	-0.006804548	0.009267954	31	92.69999995

Figure 5-2 Part of the indoor data set recorded in the travel survey

Verification of static detection algorithm based on accelerometer: Figure 5-2 shows the data of the test user staying indoors. It can be found that when the user stays indoors, GPS and network positioning keep switching between each other, and the angle (bearing) in the figure is not 0 The data is GPS data. Although the user has been in a static state indoors, the GPS drifts and lacks due to the building's obscuration of the GPS, and the speed value measured by the GPS is far greater than the true value of the

actual situation. As shown in Figure 5-3, in this case, the distance value curve and speed curve between two adjacent points of the track point sample, this is a typical real indoor stay but GPS data is characterized as moving.

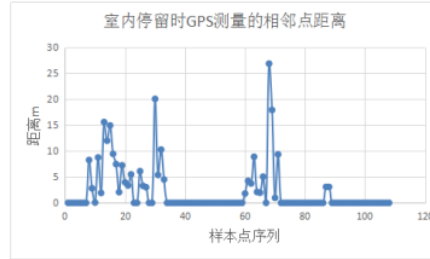


Figure 5-3 a graph of the distance between adjacent points of the sample when the positioning is unstable

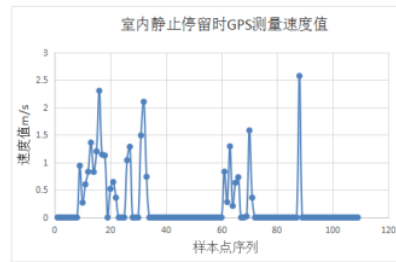


Figure 5-3 b GPS measurement speed value when the positioning is unstable

If only the characteristic parameters of time, space, distance and speed are used to identify the stay, once the parameter settings are unreasonable, it will lead to missed judgments. Some clustering algorithms use the distance between the two most distant points in the cluster to define a cluster. The special case is that if the distance threshold is set too small, it may cause missed judgment. This algorithm has poor anti-noise ability to judge the stay point. However, through the calculation results of the acceleration static detection method in this paper as shown in Figure 5-4, it can be seen that the detection of the stay point is successful.



Figure 5-4 a Sampled acceleration values of sample points in the window at rest

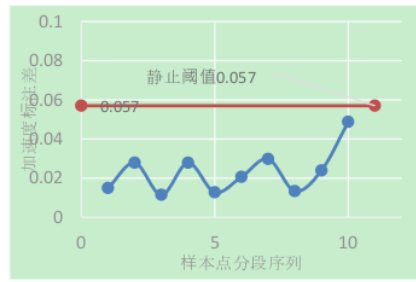


Figure 5-4 b Curves of the standard deviation of acceleration at each segment of the window at rest

Figure 5-4(a) is the acceleration sampling value of this sample. The horizontal axis is the sequence of sample points, the sampling interval is 5s, and the vertical axis is the acceleration modulus. It can be seen that the acceleration value is generally below 0.1 at rest, Figure 5-4(b) is the standard deviation curve after the sample segmentation (each 10 points is a segment). It can be seen that the standard deviations of all segments are less than the threshold value of 0.057. Through the staying point algorithm above, the sample can be determined to be in a static state. It can be seen that acceleration parameters, as reliable sensor data, can compensate for the instability of the recognition results caused by GPS positioning drift, and accurately identify the real user Stay point.

At the same time, by comparing the acceleration data in the moving state, there are significant characteristic differences between the two. Figure 5-5(a), (b) are the acceleration sampling values during walking and the standard deviation curve after sample segmentation. By comparing Figure 5-4(a) and Figure 5-4(b), we can find that walking The acceleration range at time is 1-5, and the standard deviation generally fluctuates between 0.5 and 2. Through the above stay point algorithm, it can be determined that the sample is in a non-stationary state.

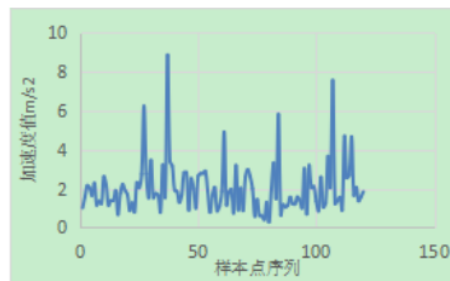


Figure 5-5 a sampled value of acceleration of window sample points when moving (walking)

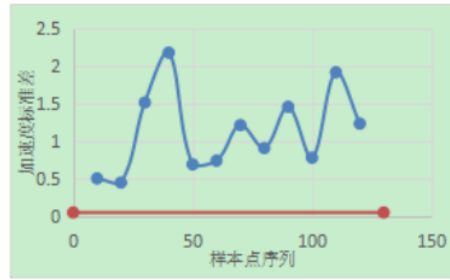


Figure 5-5 Curve of standard deviation of acceleration of each segment of the window during exercise (walking)

It can be seen that when the user stays in a static state, it can be effectively identified by the acceleration static detection method in this article. For staying points that cannot be identified by the acceleration threshold, the positioning data is used to determine whether to stay.

Validation of network positioning stay point algorithm based on empirical rules:

As shown in Figure 5-6, when staying indoors, the GPS signal is completely blocked, and the accuracy value curve in the network positioning environment is used.

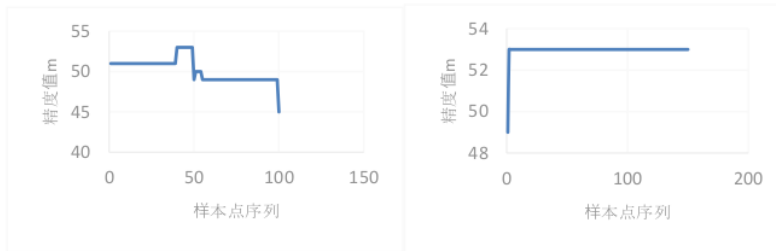


Figure 5-6 Sampling value of window sample point accuracy when stationary

The above figure is the accuracy curve during indoor stay, where the accuracy changes of the 100 sample point sequence are 12 and 4, respectively; Figure 4-8 is the accuracy curve during movement, where the accuracy changes of 100 sample points are 565 and 188, respectively. It can be seen that there is a big difference between the degree of accuracy change when staying and the degree of change when moving.

In summary, this article combines the hybrid positioning technology and acceleration sensor module under the smart phone, and proposes a real-time stop point generation method based on mobile phone multi-motion parameters. This method makes full use of the advantages of GPS positioning and network positioning. At the same time,

combined with the acceleration sensor to make up for the unreliability of the single positioning data, comprehensively use the above conditions to construct the outdoor and indoor stop point recognition model, and carry out the threshold classification of the motion parameter characteristics when moving and stationary, and finally realized on the iOS platform This method verifies its feasibility and better solves the problems existing in the identification method of stop points in current travel surveys.

5.2 Analysis of travel feature extraction results

1. Travel distance

According to the calculation formula of the travel distance in Section 4.2.3, the daily travel distance of the participants is obtained. Table 5-2 is a statistical table of the travel distance distribution of the sample of the population participating in this travel survey, and Figure 5-7 is the travel distance of the survey sample It can be seen that short-distance trips of 500-2000m accounted for 10.8%, trips within 5km accounted for 24.7%, and long-distance trips over 20 kilometers accounted for 53.6%.

Table 5-2 Travel distance distribution

Travel distance	Count	Travel distance	Count
500-1000m	4	6-7km	3
1-2km	17	7-8km	4
2-3km	11	8-9km	4
3-4km	8	9-10km	6
4-5km	8	10-15km	14
5-6km	4	15-20km	7
		Above 20km	104

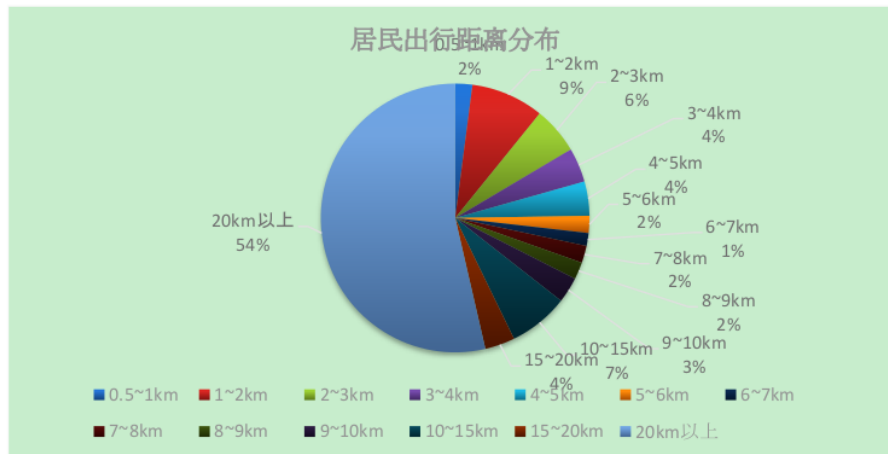


Figure 5-7 Resident travel distance distribution map

2. Travel time and distribution

According to the travel time calculation formula for individual users in Section 4.2.4, the travel time of each travel record on the day is obtained. Table 5-3 is the distribution statistics table of travel time, and Figure 5-8 is the travel time distribution ratio chart of the survey sample. It can be seen that short-term trips with a travel time of 5-30 minutes account for half of the proportion, and as the travel time increases, the proportion gradually decreases.

Table 5-3 Travel time distribution

Travel time	count	Travel time	Count
5-10 min	27	1.5-2h	13
10-20 min	53	2h-2.5h	5
20-30 min	24	2.5h-3h	2
30-40 min	15	3h-4h	6
40-50 min	13	4h-5h	3
50-60min	12	5-6h	0
1h-1.5h	18	6h-	3

After calculating the characteristics of personal travel time, calculate the travel time distribution characteristics according to the time period proposed in 4.2.4, and perform statistics on the time period distribution of all travel records of the movement points. Finally, the total is obtained, as shown in Figure 5-9 Distribution map of travel time.

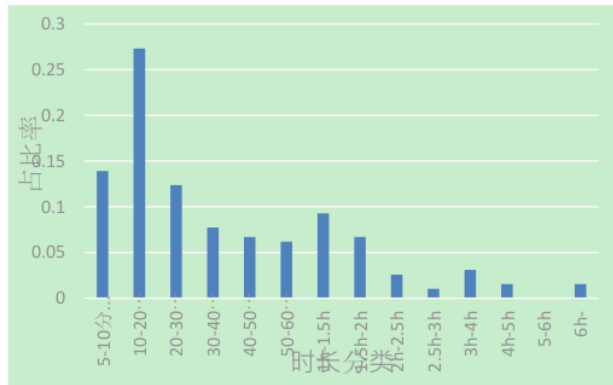


Figure 5-8 Travel time distribution ratio

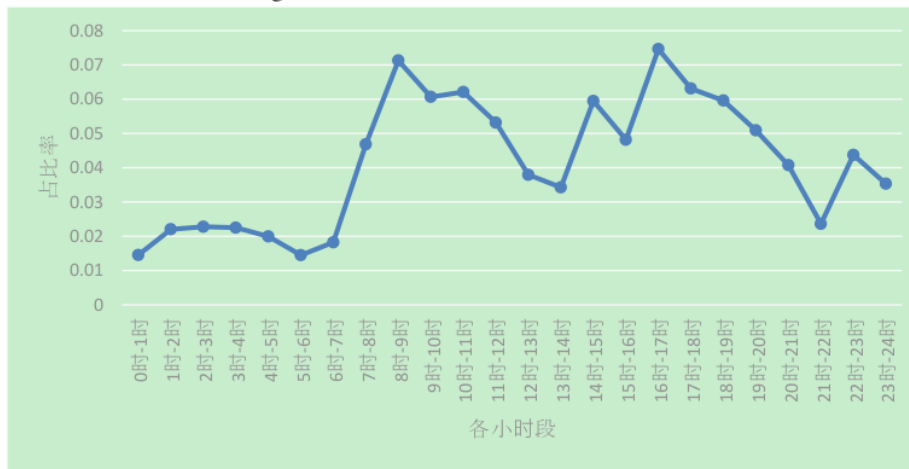


Figure 5-9 Time distribution ratio chart of system travel data

It can be seen from the figure that the hour period with the highest travel rate is between 16:00 and 17:00, the rate is 7.46%, followed by the travel rate between 8:00 and 9:00, which is 7.12%, which obviously corresponds to the evening peak and morning peak in daily travel. Among them, the travel rate from 8 am to 10 am is 13.18%, and the travel rate from 16:00 to 18:00 is 13.77%, which are the two periods with the highest travel rates in the whole day, which are more in line with the travel pattern of morning and evening peaks.

5.3 Hot spot results

According to AGNES hotspot discovery algorithm in 4.3 based on the distance between the center of mass, clustering algorithm is carried out for the stop points uploaded by users, and it is pushed to the iOS client.



Figure 5.10 Hot spot discovery display diagram

Figure 5-10 shows the hotspot area based on web display. The left side is the display of all staying points, and the right side is the map display of the corresponding hotspot distribution area. The hotspot area identified by the web background can also be transmitted to iOS customers through the Https protocol. end.

6 conclusion

This article mainly focuses on the investigation and study of residents' travel behavior in urban traffic. First, the mobile phone GPS sensor, network positioning technology and acceleration sensor are used to collect travel data. On this basis, a stop point recognition algorithm based on mobile phone multi-motion parameters is designed. And through the analysis of sampled data, the reliability and stability of the algorithm are verified, and the accuracy of identifying the location of the user is improved. Based on the sampled data, the collected data can be analyzed in the background of the system

to extract the user's travel characteristics and identify urban hot spots. This also verifies the accuracy of the collected samples from the side.

The main research results and innovations of this paper are as follows:

- (1) Detailed analysis of specific requirements based on identification and feature extraction of mobile phone travel survey stop points, design of system functions and overall framework: analysis of the system's operating process, laying the foundation for system research and implementation.
- (2) In order to efficiently identify the user's staying point, by comparing and analyzing the advantages and disadvantages of the three data and methods of GPS positioning, network positioning and acceleration sensor, a mobile phone multi-sports parameter travel survey data representation structure is constructed. A recognition method of staying points based on mobile phone multi-motion parameters is proposed.
- (3) In order to extract semantic features of user travel, a feature extraction model and calculation method based on personal travel records are proposed for travel data. The user's travel day is used as the analysis and calculation unit to extract various feature parameters, such as travel distance, Travel time length and travel time distribution, etc. Finally, the characteristic semantics of travel area are obtained through global calculation, and the effectiveness of the model and method is verified through examples.
- (4) Based on the identified stay points, an AGNES travel hotspot area discovery method based on the centroid distance is proposed, and the feasibility and effectiveness of the method are verified through examples, and the active discovery of hotspot areas in travel surveys is realized. The heat map is visually displayed on the map, which provides guidance for the construction of urban transportation on a macro level.

The author believes that the next stage of research can be considered from the following three aspects:

- (1) This paper proposes a method for identifying staying points based on mobile phone multi-motion parameters, which can effectively and comprehensively identify the staying points of mobile phone users. Generally speaking, most of the user's travel stops belong to the user's frequently visited places. Therefore, you can use this as a basis to establish a user's stopping point history database and design a method to quickly predict and match the frequently visited places .
- (2) Complement the missing trajectory during the travel survey process based on path matching, and provide a more complete and effective data basis for the subsequent analysis of the results.
- (3) For the discovered travel survey hotspots, further dig out its land use type and POI attributes within its scope for subsequent investigation and analysis. At the same time, it can be matched with the user's basic data, and a recommendation algorithm based on this can be designed. Pushing users' potential POI points provides a mode for discovering points of interest at the user level.

7 Reflection

Based on the existing problems and actual needs of the current urban traffic travel survey, this paper builds a mobile phone-based travel survey stop point identification and feature extraction system. By analyzing the needs of the mobile phone-based travel survey stop point identification and feature extraction system, we study the use of mobile phone multi-motion sensor parameters to construct a travel data representation model to identify the travel stop points of mobile phone users and provide a data basis for subsequent travel feature extraction . Taking personal travel records as the object, a personal travel feature model is constructed, and a comprehensive feature extraction method based on the personal travel feature model is proposed to mine the macro travel characteristics and laws of travel surveys.

In this project development process, the author learned a lot of low-level knowledge, especially when using the ssh framework for web development, because various frameworks in the project need to be manually configured, this is for me to learn the springboot framework and understand the framework. The underlying operating principles have helped a lot. In the development of mobile phone applications, through the learning and use of Object-C language, I have mastered many frameworks in iOS development, such as common AFNetworking, CoreMotion, Runtime mechanism, KVO monitoring mechanism, KVC key-value pair monitoring, NSNotification Broadcast notifications, and have an in-depth understanding of the Runtime mechanism. By understanding the underlying operating principles of these frameworks, I use them more confidently and apply them flexibly to projects.

In addition to learning in technology, through this system, I also have a stronger interest in software development, and even want to engage in this interesting work for a lifetime. And I also learned to analyze the needs of a project, design the overall system architecture, and be able to reasonably arrange the project progress time and strictly plan the project schedule.

8 Reference

- [1]Zhang Wei. (2015). Study on the Application Method of Extracting The Travel Characteristics of Urban Residents Based on Mobile Phone Positioning Data (Nanjing: Southeast University).
- [2]Du, J., & Aultman-Hall, L. (2007). Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. *Transportation Research Part A: Policy and Practice* , 41 (3), 220-232.
- [3]Tsui, S. Y. A., & Shalaby, A. S. (2006). Enhanced system for link and mode identification for personal travel surveys based on global positioning systems. *Transportation Research Record* , 1972 (1), 38-45.
- [4]Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* , 17 (3), 285-297.
- [5]Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies* , 17 (3), 285-297.

- [6]Wan, N., & Lin, G. (2013). Life-space characterization from cellular telephone collected GPS data. *Computers, Environment and Urban Systems* , 39 , 63-70.
- [7]Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E., & Zegras, P. C. (2013). Future mobility survey: Experience in developing a smartphone-based travel survey in Singapore. *Transportation Research Record* , 2354 (1), 59-67.
- [8]Safi, H., Assemi, B., Mesbah, M., Ferreira, L., & Hickman, M. (2015). Design and implementation of a smartphone-based travel survey. *Transportation Research Record* , 2526 (1), 99-107.
- [9]Hariharan, R., & Toyama, K. (2004, October). Project Lachesis: parsing and modeling location histories. In *International Conference on Geographic Information Science* (pp. 106-124). Springer, Berlin, Heidelberg.
- [10]Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies* , 26 , 301-313.
- [11]Angelakis, V., Gundlegård, D., Rydergren, C., Rajna, B., Vrotsou, K., Carlsson, R., ... & Zhao, S. (2013). Mobility modeling for transport efficiency: analysis of travel characteristics based on mobile phone data. In *Netmob 2013-Third International Conference on the Analysis of Mobile Phone Datasets*, May 1-3, 2013, MIT, Cambridge, MA, USA .
- [12]Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C., & Ben-Akiva, M. (2014, August). Activity recognition for a smartphone based travel survey based on cross-user history data. In *2014 22nd International Conference on Pattern Recognition* (pp. 432-437). IEEE.
- [13]Liu Shasha. (2013). Research and Realization of Smartphone-based Behavior Recognition of getting on and off buses (Master's Thesis, Wuhan University of Technology).
- [14]Zhao, F., Ghorpade, A., Pereira, F. C., Zegras, C., & Ben-Akiva, M. (2015). Stop detection in smartphone-based travel surveys. *Transportation research procedia* , 11 , 218-226.
- [15]Gao Jing, & Duan Huichuan. (2011). Research on JSON data transmission efficiency. *Computer Engineering and Design*, 32 (7), 2267-2270.
- [16]Hu Zhongwei, DENG Xiaoyong, GUO Jifu, & Wen Huimin. (2013). Analysis of residents' Travel Demands based on mobile Phone location data. *The 8th China Annual Conference on Intelligent Transportation*, 889-897.
- [17]Wei Wei. (2013). Resident travel survey system based on JavaEE and mobile location services [D] (Doctoral dissertation, East China Normal University).

Paper

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

