www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

OD Matrix Acquisition Based on Mobile Phone Positioning Data

Xiaoqing ZUO, Lei CHEN, Hongchu YU
Faculty of Land and Resource Engineering, Kunming University of Science and Technology, Kunming, China
E-mail: zuoxq@163.com, 954206641@qq.com, 1063716089@qq.com

Received: 30 May 2014 /Accepted: 27 June 2014 /Published: 30 June 2014

Abstract: Dynamic OD matrix is basic data of traffic travel guidance, traffic control, traffic management and traffic planning, and reflects the basic needs of travelers on the traffic network. With the rising popularity of positioning technology and the communication technology and the generation of huge mobile phone users, the mining and use of mobile phone positioning data, can get more traffic intersections and import and export data. These data will be integrated into obtaining the regional OD matrix, which is bound to bring convenience. In this article, mobile phone positioning data used in the data acquisition of intelligent transportation system, research a kind of regional dynamic OD matrix acquisition method based on the mobile phone positioning data. The method based on purpose of transportation, using time series similarity classification algorithm based on piecewise linear representation of the corner point (CP-PLR), mapping each base station cell to traffic zone of different traffic characteristics, and through a series of mapping optimization of base station cell to traffic zone to realize city traffic zone division based on mobile phone traffic data, on the basis, adjacency matrix chosen as the physical data structure of OD matrix storage, the principle of obtaining regional dynamic OD matrix based on the mobile phone positioning data are expounded, and the algorithm of obtaining regional dynamic OD matrix based on mobile phone positioning data are designed and verified. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: Mobile phone traffic data, Mobile phone positioning data, Traffic OD zone, Dynamic OD matrix.

1. Introduction

In the intelligent transportation system (ITS), OD matrix (transportation production matrix) is a very important initial data, which is essential foundation data for traffic management, traffic planning and traffic control. In addition, OD matrix is also the most direct, the most reliable simulation input data in traffic simulation system [1]. Roadside inquiring method, form survey method, family access method, postcards survey method, vehicle license plate method and other traditional obtaining method of OD matrix, need large-scale artificial sampling survey, consume a mass of manpower, financial resources and time, in addition, these methods mostly depend on the subjects of investigation having a recall of "travel time, travel purpose" and other travel information. In practice, there are often unclear information memory, impatient such investigation, and random answer to travel information given, making the OD matrix investigated with low accuracy and unreliable. This survey generally is carried out after a lapse of a few years, which has long update cycle, and can't reflect the characteristics of the dynamic OD information and soon on [2]. It is imperative to adopt new methods and new technology to study the OD matrix acquisition.

According to Ministry of Industry and Information Technology of the People's Republic China (MIIT) statistics, as of 2013 August, Chinese mobile phone users have reached 1.196 billion; Mobile phone ownership in big cities is already close to 100 %. Huge mobile phone users, provides a vast and stable data source for obtaining the OD matrix based on the mobile phone positioning data. The basic principle to obtain the dynamic OD matrix using mobile phone positioning data is: each mobile phone user as a independent of each other sensor, through the mobile communication network to record mobile phone base station cell switching sequence, the motion trajectory of mobile phone users on traffic road can be calculated out, and through the establishment of corresponding relationship between the base station cell and urban traffic zone in GSM network, map the mobile base station cell to urban traffic zone, so as to obtain the corresponding OD data [3]. The technology is not dependent on wireless positioning technology of GPS, without additional equipment installation and maintenance cost, without the need for the existing terminal hardware upgrade renovation, with short construction period, low installation cost, and less susceptible to outside interference, and it obtain traffic information of high coverage and all-weather. OD Matrix acquisition based on mobile phone positioning data has distinct advantages over traditional method, through continuously tracking the position change of a target object, to reflect the dynamic change characteristics of OD matrix at various times, with enough traceable samples, the dynamic OD matrix well representative, high data quality, authentic and reliable.

In this paper, according to the general change pattern of time series of typical traffic zone's mobile phone traffic, use time series similarity classification algorithm based on piecewise linear representation of the corner point (CP-PLR), to realize the mobile base station cell to the city traffic zone mapping. On this basis, through the tracking record of masses of mobile phone users' point trajectories, and the statistics of OD information between each traffic zone, obtain the regional dynamic OD matrix of the corresponding period.

2. The Research Background

2.1. GSM Communication Network

The GSM communication network consists of many components and ports which provide communications services for the mobile station (usually refers to a mobile phone), and sets the information transceiver, control, switching, routing, registration and other functions into one [4]. Fig. 1 shows the architecture of GSM network [5]. Users carrying mobile station MS (Mobile Station), which are connected to BSS (Base Station System) and NSS (Network Subsystem) of the mobile Station and Network, as well as to the GSM communication network, PSTN (Public Switched Telephone Network) and other external network. The BSS can be subdivided into the Base Transceiver Station (BTS) and the Base Station Controller (BSC). BTS will be responsible for the mobile station to access the communication network, which is realized through Um port, the range that a BTS service coverage as a small base station (SITE). So the coverage area of GSM communication network can be classified as a series of base station cell collection, called LAC (Location Areas). It is worth pointing out that, with the aid of signal exchange between mobile phones and communications networks, the location of the mobile phone can be positioned in the range of base station cell SITE_ID after the positioning treatment. This is because the mobile phone service operators record the coordinates of each BTS, which can determine the approximate spatial position of the mobile phone connection with each BTS.

In addition to providing mobile phone location information, GSM communication network also provides mobile phone activity parameters, which reflect the utilization rate of network. The mobile traffic data reflects the degree of the use of communication networks, which is a standard for a lot of communication network operators to measure the network usage rate.

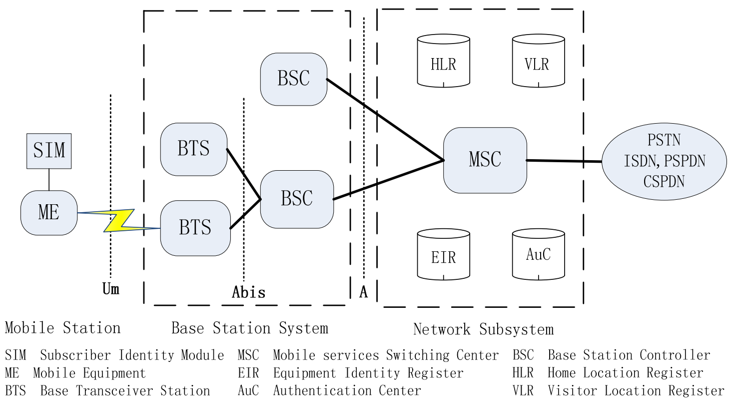

Fig. 1. The architecture of GSM communication network [5].

2.2. Voronoi Diagram to Represent Base Station Service Area (Base Station Cell)

Mobile phone service operators record the basic information of each mobile phone base station, whose purpose is to conduct the daily management and maintenance of the base station. This information includes base station code, latitude, longitude and other field. Some part of the data is illustrated in Table 1. We only need to know the spatial coordinates of the base station and base station cell coverage radius, to determine the approximate spatial scope of the mobile phone users communicating with the base station (COOpositioning method). Therefore, determining the scope of service area of the base station is one of the important problems we face, after thinking and research, we introduce the Voronoi diagram as the service area of the base station, namely the base station cell (SITE) [6, 7].

Table 1. The attributes structure of the base station cell.

Field Name	Field data type	Field description
SITE_ID	Long integer	The base station code
LAC	Long integer	The location area code of base station
SHAPE	Geometric objects	On behalf of planar geometry of the base station cell
LONGITUDE	Double	The longitude of the base station location
LATITUDE	Double	The latitude of the base station location

Voronoi diagram, also called the Tyson polygon, is composed of a group of continuous polygons which are formed by perpendicular bisectors of straight lines connecting two adjacent points. N distinguishing points in a plane, according to the proximity principle divide plane; each point is associated with its nearest neighbor area. Although the signal strength from base station is changing, the location area range will also have a certain degree of change. But in order to facilitate the study, the assumption that all signals sent from the base station are the same and the range of location areas is fixed. So base station as the discrete points, Voronoi distribution graphs of all base stations are established to represent the service area of the base station (base station cell), as shown in Fig. 2.

Fig. 2. The distribution graph of base station cell.

2.3. The Mobile Telephone Traffic Data

In a base station cell, the product (unit for Erlang) of the total number of calls per specific unit time and the average occupancy time each call is mobile phone traffic of the base station. Usually, mobile phone traffic data uses base station as the unit to record, which is a macroeconomic indicator for communication busy degree in the base station cells. Mobile phone traffic data with location attributes, can according to the base station code (SITE_ID) determine its position. It reflects strength of communication activity of different place. Using these data, unearth the time-spatial distribution law of resident activity, reflect the traffic characteristics of the base station cell on the side. As shown in Table 2, SITE_ID for base station code, telephone traffic i(i) for telephone traffic of the base station cell in the period numbered i, the row in Table 2 represents the time series of a base station cell traffic data.

Table 2. The mobile telephone traffic data.

SITE _ID	Telephone traffic 0	Telephone traffic 1	Telephone traffic 2	Telephone traffic 3	Telephone traffic 4	Telephone traffic 5	Telephone traffic 6
1	7.21	5.04	3.46	2.37	0.83	0.14	0.69
2	1.08	0.76	0.27	0.08	0.21	0.14	0.11
3	6.17	2.91	1.88	0.96	0.34	0.03	0.17

2.4. The Mobile Phone Anchor Point Data

Mobile phone anchor point data describes a series of ordered discrete points recorded when the mobile phone users using the COO positioning method. Each point represents the center of gravity of the mobile phone user's base station cell in a moment, these ordered discrete points as the beginning and the end points of mobile phone users, so the set of ordered discrete points that have the same USER_ID represents a mobile phone user travel path (that is the mobile phone user travel chain), with point trajectory of each mobile phone user for statistical analysis through a certain algorithm, to obtain the corresponding regional dynamic OD matrix. The attribute structure of mobile phone anchor point data is shown in Table 3.

Table 3. The attribute structure of mobile anchor point data.

Field Name	Field data type	Field description
USER_ID	Long integer	Mobile phone user code
TIME	Text	Mobile phone user positioning time
SITE_ID	Long integer	The base station code of the mobile phone user's base station cell
SHAPE	Geometric objects	On behalf of geometric shapes of the mobile anchor point data
CORE_X	Double	The abscissa of the center of gravity of the mobile phone user's base station cell
CORE_Y	Double	The ordinate of the center of gravity of the mobile phone user's base station cell

3. The Traffic Zone Division Based on Mobile Phone Traffic Data

For traffic travel having two endpoints (the starting and ending points), without considering traffic tools, the purpose of travel constitutes a necessary condition for traffic. The traffic zone division method based on mobile phone traffic data is to focus on the purpose of the traffic travel, and divide each base station cell into zone of different function and different traffic characteristics, with a certain cell merge optimization algorithm to realize the mapping from base station cell to traffic zone. Mobile phone traffic data is an index that mobile base station records talk activity in its service area within a certain time. It has a close relationship with the total activity of the people in its corresponding areas, and it can be qualitatively describe as a kind of positive correlation between total mobile telephone traffic and human activities. The change of mobile phone traffic of different base station cell along with time is different, but the changes of mobile phone traffic of the base station cells with similar transportation characteristics generally consistent over time. Based on this, researcher Li Xiaopeng conduct self-organizing competitive network clustering analysis for the mobile phone traffic data, through unguided self-learning clustering process with multi-step iterative correction, and using mobile phone traffic time distribution characteristics to analyze urban regional traffic characteristics and distinguish land unit that has uniform traffic characteristics [8]. The clustering analysis process of the method from the strict sense is a kind of unsupervised data mining process, in the case of delimiting classification unknown in advance, according to the similarity principle of information gathering of information. It is unavoidable to have the shortcomings that clustering category is single and clustering results cannot be artificially controlled. This paper proposed a data mining process with the supervision and guidance, in the case of predefined categories, dividing each object of the data set into the known object class, the specific approach is first through statistical methods to analyze the change characteristics of mobile phone traffic in all kinds of typical traffic zone artificially determined in advance along with time, to get the general pattern of the time series variation; And as a base station cell classification standard with this, according to the different characteristics of general pattern, design corresponding classifier; Then use time series similarity classification algorithm based on piecewise linear representation of corner points (CP-PLR) to realize the base station cell to traffic zone mapping; Finally, merge the same type of location adjacent base station cells into the same traffic zone, and conduct the corresponding mapping optimization; Thus divide the final city traffic zone of the whole research area, and realize the transition of mobile base station cell to urban traffic zone.

3.1. Time Series and the Similarity Measurement

Time Series is the effective collection which is made up of the binary pairs of real values and time, expressed as, said the real value at the moment in time series. When =0 and ∆t=-=1, time sequence can be abbreviated to X= [, , …,] [9]. For the two time series K= [[, , …,] and Q = [[, , …,], their Euclidean distance is defined as Distance (K, Q) =; If distance, the time series of K and Q are similar, or dissimilar, where is similarity threshold of the time series.

3.2. The Piecewise Linear Representation Method Based on Corner Points (CP-PLR)

The object of time series similarity analysis is time series data with high dimensions and mass characteristics, which generally has the characteristics of serious noise disturbance and frequent short-term fluctuations, direct similarity analysis on the original time series not only inefficiency, and its accuracy and reliability cannot be guaranteed [10, 11]. Therefore, in order to improve efficiency, accuracy and reliability of similarity analysis, we need to deal with the original time series before the time series similarity judgment, which contains approximate representation of the original time series as far as possible to keep the original time series main morphological characteristics, reduce the column dimension of the original time series, remove noise, and compress data effectively. In this article study, use piecewise linear representation based on the corner point (CP-PLR) as the approximation of time series. Through searching the corner points (① the starting and ending points; ② time series extreme value points but not extreme noise) in the original time series, connect these corner points end to end by straight line segments, resulting in piecewise linear representation method of the original time series [12]. Let's suppose there is a period of time sequence (A, …, F). Where A is the starting point and F is the end point of time series; B, C, D, E are the extreme points of time series; C, D are extreme noise. Then the time series piecewise linear representation process based on the corner points, as shown in Fig. 3.

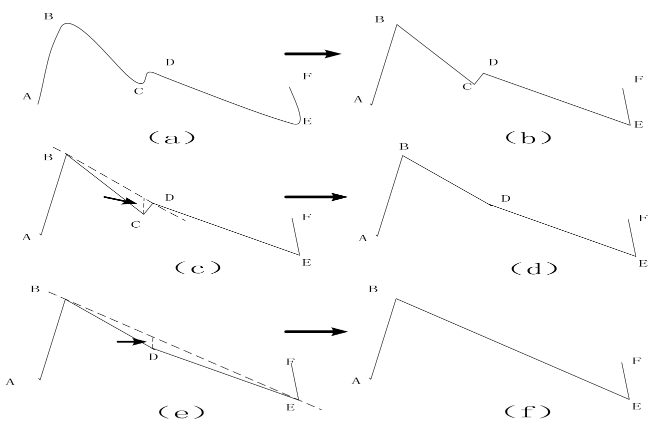

Fig. 3. Piecewise linear representation sample process of CP – PLR algorithm.

3.3. Time Series Similarity Classification Algorithm Based on CP – PLR

The research content of this paper is, according to the general pattern of typical traffic zone's mobile phone traffic time series change, designing the corresponding classifier and realizing the mobile phone communication base station cell to the city traffic zone mapping.

The time series similarity classification algorithm based on CP – PLR is critical. The specific algorithm is as follows:

```
1.  The algorithm begins
2.  Traffic data preprocessing and statistics  to get time series set P of each base
station cell traffic and time series set  S of all kinds of typical traffic zones
corresponding classifier traffic
3.  The time series of set P and set S for piecewise linear representation based on corner
points  CP-PLR) to obtain corresponding time series collections PM and SM
4.  Given a time series similarity threshold ε ;  MinDistance was defined as the minimum
Euclidean distance between the elements PM[i] in the collection PM  and the elements in
the collection SM ; Minindex is when obtaining the minimum Euclidean distance,  the
corresponding element index number in the collection SM;
5.  for i=0 to PM.Count-1           // PM. Count is the total number of time series in
collection  PM
6.        minDistance= Distance(PM[i],SM[0])  //The initialization of minDistance
7.        for j=0 to SM.Count-1
8.              if( Distance(PM[i],SM[j])<minDistance) then
9.                   minDistance= Distance(PM[i],SM[j])
10.                  minindex=j
11.             end if
12.       end for
13.       if(minDistance ≤ ε ) then
14.           do The division of PM[i] corresponding base station cells into SM[minindex]
corresponding  traffic zone types
15.       else
16.        do  The division of PM[i] corresponding base station cells into the transition
area (Comprehensive Area)of each traffic zone
17.       end if
18.  end for
19.  The algorithm is over
```

4. Regional Dynamic OD Matrix Acquisition Based on the Mobile Phone Positioning Data

4.1. OD Storage Structure Based on Adjacency Matrix

In general, the regional OD matrix is a square matrix of equal number of rows and columns, where the number of non-zero elements is far greater than the number of zero elements [13]. The distribution of elements does not follow regularity in the regional OD matrix as a special matrix (such as symmetric matrix, an upper triangular matrix, and a lower triangular matrix) does; And the algorithm to obtain the regional dynamic OD matrix in this paper, need random storage of OD matrix, to find each vertex adjacency point, fast search, and need to modify the side weights between any two vertices and, (namely the travel traffic volume between traffic zone numbered i as the starting point and traffic zone numbered j as the end point); Therefore, in this paper, select the adjacency matrix as the physical data structure which the algorithm to obtain regional dynamic OD matrix based on, with a one-dimensional array Traffic_Areas to store all traffic zones of the research area, and with a two-dimensional array Matrix to store the traffic travel volume(namely the regional dynamic OD matrix) between each traffic zone of the research area. Their class view as shown in Fig. 4.

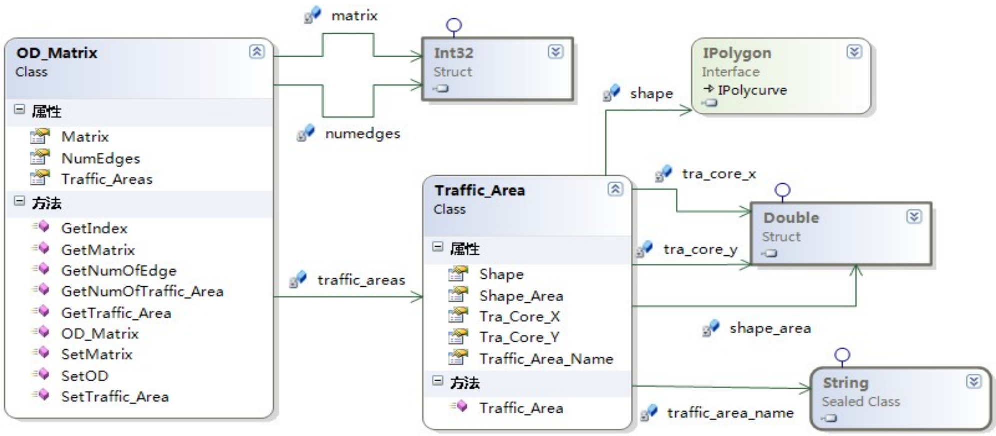

Fig. 4. The class view of traffic zone classes Traffic_Area and OD matrix class OD_Matrix.

4.2. The Algorithm to Obtain the Regional Dynamic OD Matrix

4.2.1. The Acquisition Principle of Regional Dynamic OD Matrix Based on the Mobile Phone Positioning Data

The acquisition principle of regional dynamic OD matrix is on the basis of using mobile telephone traffic data for urban traffic zone division and mapping the mobile base station cell to the city traffic zone; Using the existing mobile phone communication network facilities, with COO (Cell of Origin) positioning method as the basis, use barycentric coordinates of the mobile phone base station cell to determine mobile phone users' position coordinates within the scope of its service; The changing pattern of continuous cell switching of mobile phone users in the network communication layer expressed as a group of continuous ordered discrete points (the center of gravity of mobile phone users' base station cell in a moment) in a plane, these ordered discrete points as the starting point and end point of mobile phone users, so different ordered discrete point set represents point trajectory of different mobile phone users travel(that is the mobile phone user travel chain); Keeping tracking and record of the mass mobile phone users' point trajectory, counting OD information between each traffic zone, obtain regional dynamic OD matrix of the corresponding period [14].

As shown in Fig. 5, in the GSM mobile communication network, the continuous cell switch of a mobile phone user in the network communication layer can be expressed as a group of continuous ordered discrete points in a plane. In Fig. 5, each small hexagon for the scope of one base station cell signal coverage, base station cells of a color grouped together constitute the four urban traffic zones of different functions and different traffic characteristics: residential living zone, transportation hub zone, working zone and entertainment shopping zone. STOP i represents the center of gravity of base station cell where cell switching occurs ($1 \leq i \leq 8$).

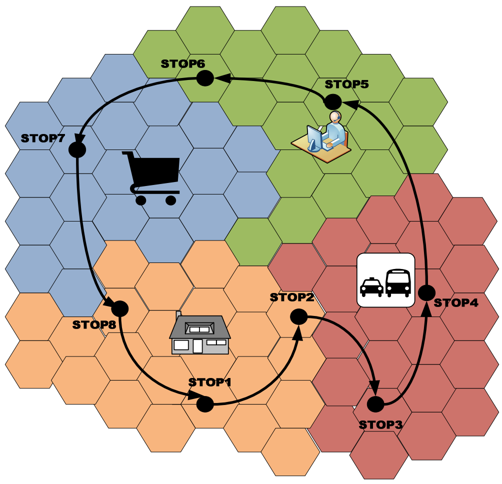

Fig. 5. The acquisition principle diagram of regional dynamic OD matrix.

Due to the mobile phone users with different life entertainment needs at different time in a day, according to the purpose of travel, the travel point trajectory can be expressed as the STOP1 - STOP2 - STOP3 - STOP4 - STOP5 - STOP6 - STOP7 - STOP8 - STOP1. Track and record the travel points' trajectory, to get eight pairs of starting and ending points that include (STOP1, STOP2), (STOP2, STOP3), (STOP3, STOP4), (STOP4, STOP5),

(STOP5, STOP6), (STOP6, STOP7), (STOP7, STOP8), and (STOP8, STOP1). Through the statistical analysis of eight pairs of starting and ending points, eliminating travel starting and ending points within urban traffic zones that include (STOP1, STOP2), (STOP3, STOP4), (STOP5, STOP6), and (STOP8, STOP1), and OD statistics on travel starting and ending points which cross unban traffic zone, you can get OD information of the corresponding period between each urban traffic zone.

4.2.2. Algorithm Design for Acquiring Regional Dynamic OD Matrix

Based on the previous traffic zone class Traffic_Area and OD matrix class OD_Matrix, according to the acquisition principle of regional dynamic OD matrix based on mobile phone positioning data, the design thinking of the regional dynamic OD matrix acquisition algorithm in this paper is as follows:

1) Firstly, using traffic zone division method based on mobile phone traffic data, dividing each base station cell into urban traffic zones of different functions and different traffic characteristics, mapping mobile base station zone to traffic zone through a certain cell merge optimization algorithm, and initialize traffic zone class and OD matrix class by taking each traffic zone data.

2) Using mobile phone user code USER_ID field as a primary keyword, user positioning moment TIME field as the secondary keyword, sort mobile phone anchor points data into ascending order; then develop the set of unique value on base of the mobile anchor points data' USER_ID field, using the unique value in the collection as the query conditions, and extract sequentially the positioning data of each mobile phone user in a particular time period; For these data, each mobile phone user as the basic unit, sequentially extract data of two consecutive anchor points as the starting and ending points of travel, through the SetOD method of OD_Matrix object, to determine the starting and ending points whether across the city traffic zone or not; If across the city traffic zone, then carry on the corresponding record in the two-dimensional array named Matrix of OD_Matrix objects, and if not, directly judge and record the next pair of starting and ending points of travel, until the user' all travel starting and ending points are processed. Finally, by OD judgment and recording of each mobile phone users' all starting and ending point pairs, we can get the total urban regional dynamic OD matrix. The flow chart of the algorithm to obtain the regional dynamic OD matrix as shown in Fig. 6.

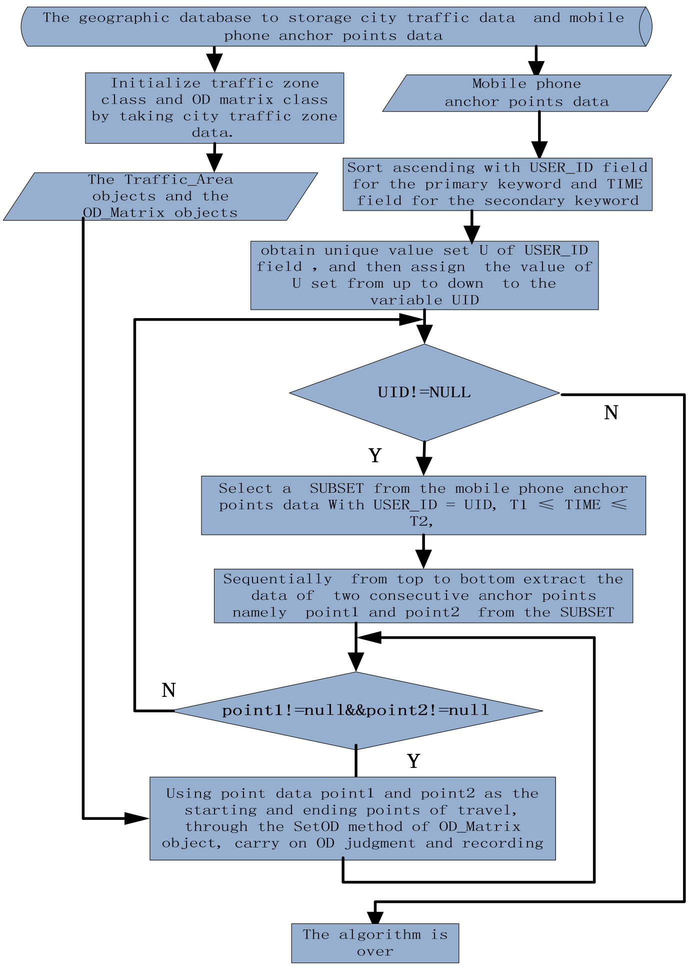

Fig. 6. The regional dynamic OD matrix acquisition algorithm flow chart.

5. Example Analysis

5.1. Traffic Zone Classifier Design

Before traffic zone division based on the mobile phone traffic data, by applying averaging operator to the mobile telephone traffic data, obtain statistically more stable average value of mobile phone traffic of each period of each day of each base station [15]. And the Euclidean distance as the similarity measure of time sequence in the 3.1 section of this article, it is very sensitive to the time sequence offset changes on the time axis, therefore time series data of mobile phone traffic needs further standardization processing. Divide traffic of each time interval of each base station cell by the sum traffic of 24 time intervals of the base station cell, showing traffic by relative size. On this basis, firstly, typical traffic zone in advance can be artificially divided into residential living zone, transportation hub zone, working zone and entertainment shopping zone; Then in the pre-determined training zone, each base station cell as a unit, statistical analysis of geographical names data and the address data in its range, combined with the actual situation of the traffic characteristics of the base station cell, divide it into some kind of typical traffic zone; Based on the base station cell of training area all sorted, all kinds of typical traffic zone traffic data respectively summarizing, then traffic variations characteristics over time as the basis, and get the general pattern of all kinds of typical traffic zone traffic time series change through traffic filtering and fitting. Finally these obtained patterns as classification standards of all base station cells in the research area, design corresponding classifier according to the characteristics of morphological variation of the general patterns.

Relative traffic time sequence change analysis diagram of each designed classifier as shown in Fig. 7.

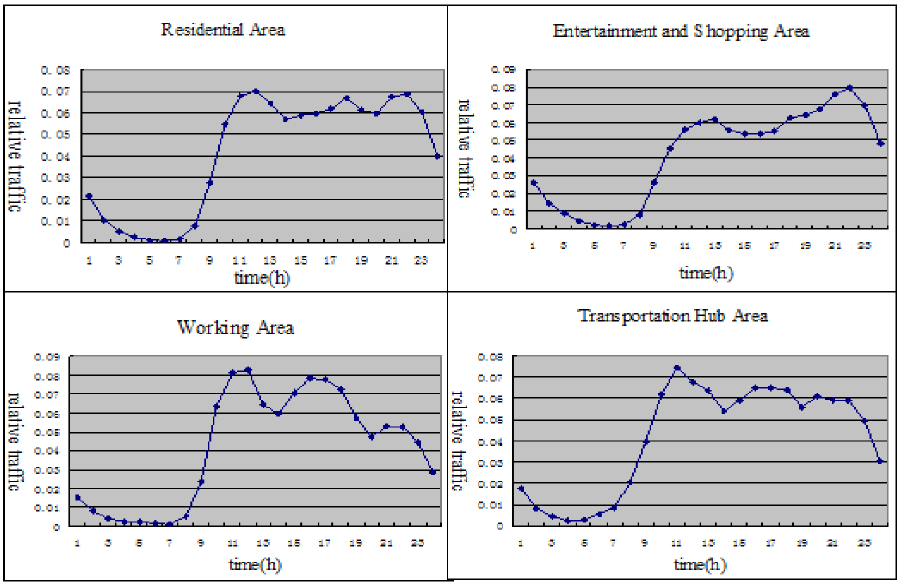

Fig. 7. Time series classifiers of each typical traffic zone.

5.2. The Base Station Cell Classification

Typical traffic zone categories expressed by traffic zone time series classifier as the standard, use time series similarity classification algorithm based on piecewise linear representation of corner points (CP-PLR) to realize the base station cell to traffic zone mapping, and divide each base station cell in Kunming urban area one by one into the corresponding category. In particular, when the minimum Euclidean distance namely MinDistance between base station cell time series and traffic zone classifier time series is greater than the similarity threshold, the base station cell time series and typical traffic zone classifier time series can be considered not similar, and the base station cell should be divided into the transition area of each traffic zone, which is called Comprehensive Area. (See Fig. 8).

5.3. Mapping Optimization Base Station Cell to the Traffic Zone

After each base station cell sorted, divide the same kind of geographic location adjacent base station cells into one traffic zone; In addition, according to the foreign successful experiences of traditional traffic survey, generally, the area of the traffic zone in city center is between one square kilometers and three square kilometers, and the area of the traffic zone at the edge of the city is between five square kilometers and fifteen square kilometers [16]. However, according to Chinese city high population density and complex land property, traffic zone scale determination need to combine with the specific circumstances of different cities.

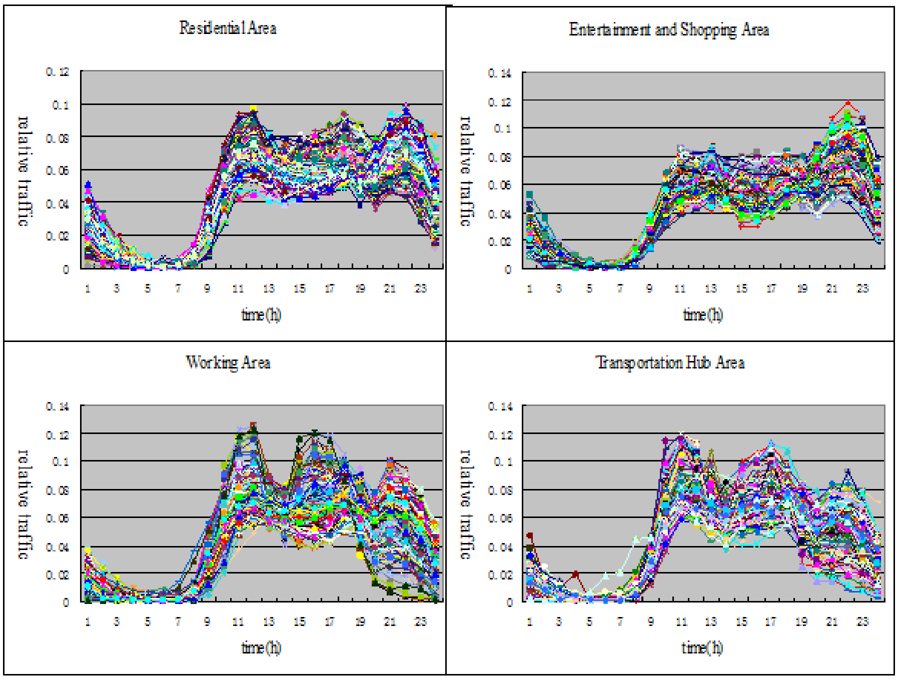

Fig. 8. Dividing base station cell time series into various typical traffic zones.

The minimum area of traffic zone of Kunming city can be set to 0.2 square kilometers in this paper. For mobile base station relatively concentrated area, some traffic zone, the area of which tend to less than the lower limit after the base station cells merged, should be attributed to other kinds of geographic location adjacent traffic area which has the closest connection with these "piecemeal" traffic zones. Here we can put the length of the public side between two cells as the criterion for judging whether they are closely linked, and "piecemeal" traffic zones should be attributed to other kinds of geographic location adjacent traffic area which has the longest public side with them, thus to obtain the final city traffic zone division of the study area. As shown in Fig. 9, the division graph of part of traffic zones in the urban area of Kunming City, where A (i) as the name code of each city traffic zone (1 ≤ i≤ 20).

5.4. The Generation of Regional Dynamic OD Matrix of Traffic Zone

Taking the division graph data of part of traffic zones in the urban area of Kunming City as input data, initialize traffic zone class Traffic_Area and OD matrix class OD_Matrix; Then select part of mobile phone users' positioning data within the 20 traffic zones to verify the regional dynamic OD matrix acquisition algorithm, and the regional dynamic OD matrixes between 20 traffic zones in a certain time are as shown in Table 4. By the regional dynamic OD matrixes in Table 4, we can know travel traffic volume (i, j) between the various traffic zone and other traffic zones in a certain period of time. As shown in Fig. 10, each three-dimensional pie chart said percentage of travel traffic volume between its traffic zone and other various traffic zones.

Fig. 9. The division graph of part of traffic zones in the urban area of Kunming City.

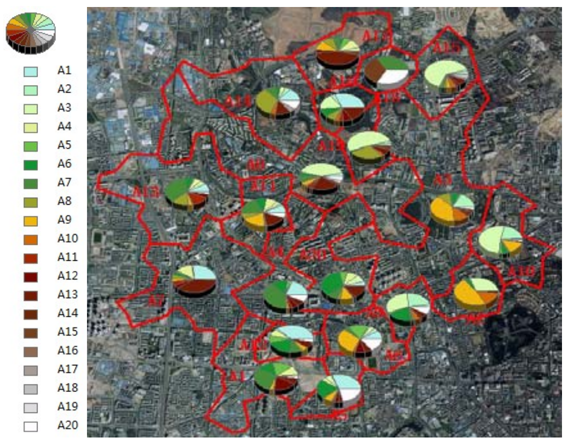

Fig. 10. The three-dimensional pie chart of travel traffic volume between each traffic zone.

Through the three-dimensional pie chart, we can know the distribution of travel traffic volume of each traffic zone very intuitively. For example, through pie chart of the traffic zone A7, we can clearly see that the travel traffic volume of A7 traffic zone is mainly distributed in the A1 traffic zone and A13 traffic zone, travel traffic volume between A7 and A1, A7 and A13 basically accounted for more than two-thirds of the total travel traffic volume of A7 traffic zone.

5.5. Traffic Generation and Traffic Attraction Analysis

By the regional dynamic OD matrixes in Table 4, we can also know traffic generation (i) of each traffic zone in a certain period of time. In order to understand the size and other geographical spatial distribution characteristics of traffic generationof each traffic zone more intuitively, we can do point data interpolation into raster surface processing of traffic generation (i) of each traffic zone. Concrete steps are as follows.

The barycentric coordinates of each traffic zone as the plane coordinate of discrete interpolation points, traffic generation of each traffic zone as interpolation properties, adopting the inverse distance weighted interpolation method, generate three-dimensional surface of the traffic generation as shown in Fig. 11.

In Fig. 11, you can easily see that traffic generation of traffic zones are mainly distributed in traffic zones A9, A3, A6, A1, A7, and A13. Traffic generation of traffic zone A9 is the maximum traffic generation, followed by traffic generation of traffic zone A3 and traffic generation of traffic zone A6, finally to traffic generation of traffic zones A1, A7 and A13.

Traffic attraction similar to traffic generation, the corresponding three-dimensional surface chart of traffic attraction can be obtained with the same step.

Table 4. The two-dimensional table of dynamic OD matrixes of traffic zones.

O\D	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	O
A1	0	8	13	105	203	165	310	11	49	9	12	203	76	1	0	3	1	0	0	24	1193
A2	8	0	235	4	10	33	7	3	374	111	2	13	6	0	2	0	1	0	1	6	816
A3	13	242	0	9	31	106	26	155	570	193	9	20	42	0	32	4	3	0	16	40	1511
A4	101	3	11	0	16	75	203	3	15	2	16	29	29	0	0	0	0	0	0	44	547
A5	207	8	36	26	0	127	9	4	44	14	7	26	28	1	0	1	1	0	0	141	680
A6	116	38	55	19	249	0	62	10	389	28	11	130	76	1	0	2	3	0	1	194	1384
A7	323	2	18	170	18	45	0	42	30	5	50	47	479	2	2	2	5	0	0	13	1253
A8	16	5	208	3	7	16	29	0	22	8	15	8	118	8	2	2	6	0	11	21	505
A9	36	406	515	17	25	343	37	34	0	70	14	35	80	1	1	0	0	0	12	135	1761
A10	5	91	236	2	12	23	8	3	61	0	0	4	8	0	2	0	1	0	2	8	466
A11	21	5	10	28	2	29	52	14	56	3	0	3	47	0	0	0	0	0	0	15	285
A12	215	3	8	42	38	115	60	8	23	3	0	0	41	0	0	5	1	0	0	15	577
A13	92	3	46	57	21	84	383	167	40	12	141	44	0	8	2	6	6	0	2	29	1143
A14	3	1	1	0	0	1	2	7	0	1	0	0	3	0	0	0	1	0	0	2	22
A15	1	1	30	0	1	0	1	1	2	2	1	1	3	0	0	0	1	2	0	0	47
A16	8	1	4	0	0	2	0	2	0	1	1	2	4	0	0	0	0	0	0	0	25
A17	0	0	1	3	2	1	2	3	3	0	0	1	14	0	1	0	0	0	0	0	31
A18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	3
A19	1	1	23	0	0	0	1	13	0	1	0	0	3	0	2	0	0	0	0	0	45
A20	26	4	55	62	46	224	60	23	77	8	5	7	93	0	0	0	0	0	0	0	690
D	1192	822	1505	547	681	1389	1253	503	1755	471	284	573	1150	22	47	25	30	2	45	688	

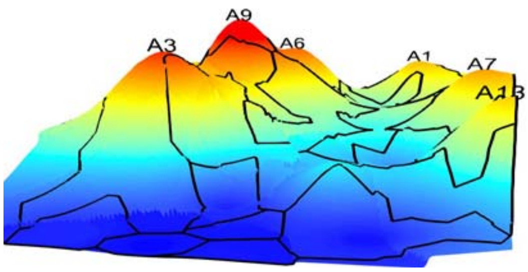

Fig. 11. The three-dimensional surface chart of traffic generation.

6. The Conclusion and Prospect

In this experiment, traffic generation of traffic zone A9 is the maximum, followed by traffic generation traffic zone A3 and traffic zone A6. In Fig. 9, the division graph of part of traffic zones in the urban area of Kunming City obtained by the algorithm overlaid to the GIS platform, you can see that the three zones are just the locations of three universities and several well-known schools in Kunming City, with dense population, traffic generation huge. Therefore, the traffic zones division obtained by this method is practicable, reflects the traffic distribution of traffic zone at various times, captures dynamic changes of the traffic flow well, and has a very high reference value to the division of urban traffic area.

In this paper, the purpose of traffic travel as the direction, connect the mobile phone traffic of base station cell with population activity in its scope. Based on different traffic of the base station cell reflecting the regional traffic characteristics, according to the morphological characteristics of mobile phone traffic time series variation, obtain qualitative description and classification of traffic characteristics of the base station cell, to realize the transformation of mobile base station cell to urban traffic zone. On this basis, through the analysis of the mobile phone positioning data, design and verify the regional dynamic OD matrix acquisition algorithm based on the mobile anchor points data. This research provides a new idea and method for regional dynamic OD matrix acquisition. Future work is as follows. Firstly, study a more accurate method of cells division for the traffic zones in the complex area. Secondly, further improve the efficiency of regional OD matrix acquisition. In the implementation of the base station cell to traffic zone mapping optimization, design the corresponding base station cells mapping table, keep a record of each base station area merge optimization, and store the base station code directly in the mobile phone base station positioning data, without storing plane coordinates of the mobile phone base station, without the geographical spatial relationship judgment for each Anchor point data as well, directly through base station code information of each positioning record to find corresponding mapping information in the base station cells mapping table, to obtain the traffic zone name. This can reduce the data redundancy, and improve acquisition efficiency of the regional OD matrixes.

Acknowledgements

This research work was supported by the National Science Foundation of China (No. 41061043).

Reference

[1]. G. Y. Ma, OD matrix estimation method and its application in the traffic simulation system, PhD Thesis, *Zhejiang University*, 2006.

[2]. J. Wang, et al, Dynamic OD estimation method based on traffic flow parameters, PhD Thesis, *Jilin University*, 2012.

[3]. F. Yang, Traffic OD data acquisition technology based on mobile phone positioning, *Systems Engineering*, Vol. 25, 2007, pp. 42-48.

[4]. GSM for dummies (http://www.gsmfordummies.com/architecture/arch.shtml).

[5]. GSM (http://big.uwaterloo.ca/~jscouria/GSM/gsmreport.htm).

[6]. C. Cai, et al, A non-motion data clustering elimination method of mobile phone positioning data, *Journal of Traffic Information and Security*, Vol. 28, pp. 60-63, 2010.

[7]. R. G. Wang, The information acquisition of starting and ending points of city traffic based on sensory data, PhD Thesis, *Tianjing University*, 2012.

[8]. X. P. Li, et al, Analysis of city activities and land use characteristics based on mobile phone traffic, Shanghai, 2005.

[9]. R. G. Fang, The research of time series data mining algorithm based on the similarity analysis, PhD Thesis, *Zhejiang University*, 2011.

[10]. L. Gao, et al, Analysis approach of relevance trend of time series with multiple uncertain features, *Journal of Application Research of Computers*, Vol. 29, 2013, pp. 3255-3258, 3265.

[11]. H. L. Li, et al., The research of characteristics representation in time series data mining and similarity measurement method, PhD Thesis, *Dalian University of Technology*, 2012.

[12]. Q. Q. Ma, et al, The research of similarity search algorithm in multivariate time series, PhD Thesis, *Lanzhou University of Technology*, 2013.

[13]. Z. Jiang, et al, A novel model for dynamic OD matrix estimation in traffic network, *Journal of Traffic Information and Security*, Vol. 169, 2012, pp. 48-51.

[14]. H. Zhao, et al, The research of dynamic origin-destination matrix on model and algorithm, PhD Thesis, *University of Science and Technology*, Beijing, 2013.

[15]. J. Li, et al, Identifying traffic district based on cell phone traffic data, *Journal of Convergence Information Technology*, Vol. 7, 2012, pp. 454-462.

[16]. C. Q. Ma, et al, City traffic zone radius calculation method based on travel proportion in the traffic zone, *Journal of Traffic and Transportation Engineering*, Vol. 7, 2007. pp. 68-72.