Method of traffic zone division based on spectral graph theory

Wu Shimei*, Pei Yulong, Cheng Guozhu

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Received 1 March 2014, www.tsi.lv

Abstract

Aiming at the macro planning of traffic district division, the spectral graph theory, graph, spectral and matrix (i.e. Laplace matrix) were introduced, and the method and procedures of the traffic zone division based on spectral graph theory were put forward. Euclidean distance between communities was calculated according to four indexes, i.e., economic indicators, land development intensity, residential population quantity and spatial distance. The balance degree of indexes between communities was regarded as criterion to evaluate community division precision. Taking Dongguan, a city in the south China, as an example, above mentioned method was applied, conducting the traffic zone division, with rail transit network planning flow prediction.

Keywords: Traffic zone division, Spectral graph theory, Euclidean distance, Dongguan

1 Introduction

Traffic zone division is the basis and a key step of predicting the passenger flow. Traffic zones can be divided into macro planning, middle management and micro control [1]. Generally, the macro planning means the planners divide the areas in accordance with the administrative boundaries or roads, railways, mountains, rivers and other natural lines[2], with no comprehensive consideration of the usage of land, population, traffic and other factors within the community, and the result is comparatively subjective. This directly influences the traffic forecast precision. In order to improve the arbitrariness and limitations of traffic zone division at macro planning level, this paper introduces the spectral graph theory and proposes a new method of traffic zone division.

A comprehensive understanding of the traffic flow among the traffic resources is needed when processing the transportation planning management. However, because of the large amount of resources, it is impossible to research the resources individually. Therefore, integrating several traffic zones into one is required. Weather the divisions are proper is directly related to the research, analysis, prediction, labour force and precision. The smaller the traffic zone is divided, the more accurate it is. However, it takes much more labour forces with higher costs when the traffic zones are too small. On the basis of ensuring the accuracy of the research, taking advantage of the existing materials (like income, occupation, etc.), or the previous data given by the cities, which have been investigated on residents travel, and reducing the numbers of the traffic zones being researched, so as to decrease the workload, are difficult. In current China, the size of the traffic zone is determined by the land usage, population distribution, administrative division, geography, and the layout of roads, etc. As for the radius of the traffic zones, no uniform and related standard or theory is at hand, but with some existing data.

2 Literature Review

The method of traffic zone divisions in America is in accordance with the area and population of the zones determined by the size of the scale of the urban area. Details are as follows: when the population is less than 715,000 within the zone, the average area is 1,138 km2 with the average population of 872; when the population is less than 1,510,000, the average area is 2177km2 with the average of population of 954; then the population is less than 10,000,000, the average area is 5155 km2 with the average population of 2828; when the population is more than 10,000,000, the average area is 7183 km2 with the average population of 7339. The method of traffic zone divisions in European countries is on the basis of population, for example, the regulation in Russia is as follows: with population of 100,000 ~ 250,000, the number of the zones should be 5~10; with the population of 250,000~500,000, the number should be 8~20; with the population of 500,000~1,000,000, the number should be 15~20; with the population of 1,000,000~2,000,000, the number should be more than 50. Clustering analysis is applied in Chinese related documents when research on the traffic zone divisions, which include: automatic classification method based on the spatial analysis was proposed by Li Xiaodan and Yang Xiaoguang [3]. The model of the clustering analysis algorithm was given in the document, planning the procession and arithmetic of the divisions. The method of divisions based on the fuzzy analysis was proposed by Zhao Jinhuan, Li Wenquan, etc. [4]. In accordance with the fuzzy clustering of the equivalence matrix in clustering analysis, by making use

* Corresponding author - E-mail: 4770260@qq.com

of the data of ages, occupations, incomes and ratio of car ownership, programming the MATLAB, and using statistical magnitude F to determine the best fuzzy clustering number, the divisions of the traffic zones is realized. Dynamic division method based on the GPS data of the taxis was proposed by Lv Yuqiang [5]. A revised method based on the clustering analysis was proposed by Yang Bo and Liu Haizhou [6]. The method based on the clustering analysis was proposed by Liu Yifei [7].

The method of radius calculation of the traffic zones based on the proportion of travel inside the zones was proposed by Ma Chaoqun [8]. On the basis of the three different combinations of the radius and trip distance, the calculation model of the proportion of travel was elicited in the document, and the quantitative relation of the three was processed. The reasonable range of the radius of the traffic zones were produced by controlling the proportion of travel inside the zones. Double divided method was proposed by Dai Binggui and Hu Ji, which was based on the weighted modification of the fuzzy clustering analysis by adopting the level analysis, and established the method of weighted fuzzy clustering analysis to research the divisions of the traffic zones [9].

3 Spectral graph theory

Spectral graph theory is to study the collection of eigenvalue of a Laplace matrix or adjacency matrix in a graph. It is an effective approach to describe the diagram's structural characteristic with its eigen value and characteristic vector as well as its algebra representation. By analysing the spatial features of graph, spectra graph theory establishes the topology map (especially the graph's various invariants) and connection between the eigen values; By making use of algebraic theory, geometric theory and probability method, .the nature of topology of a graph is revealed. Concepts related with Spectral graph theory that are discussed in this paper include graph, spectrum and Laplace matrix. [10].

3.1 GRAPH AND SPECTRUM

3.1.1 Graph

Graph G is composed of two sets, V and E, and $G=(V,E)$. V is a set of n nonempty vertexes, mathematically, it is expressed as $V=V(G)=\{v_1,v_2,\cdots,v_n\}$; E is a collection of graph edges, mathematically expressed as $E=E(G)=\{(v_0,v_1),(v_0,v_2),\ldots,(v_i,v_j),\ldots,(v_n,v_{n-1})\}$.

3.1.2 Spectrum

λ is known as the n values of n-order matrix A. If there is any non - zero n-order vector which realizes $AX=\lambda X$, then X is called an eigen vector of A corresponding with eigen value.

Known that, λ is the roots of polynomials $\chi_A(\lambda)=\det(\lambda I_n-A)$, in which I_n is an n-order unit matrix, and $\chi_A(\lambda)$ is known as characteristic polynomials.

According to Gauss's theorem, the characteristic equation $\chi_A(\lambda)=0$ has n roots: $\lambda_1,\lambda_2,\cdots,\lambda_n$. Since $\lambda_1,\lambda_2,\cdots,\lambda_n$ are not necessarily distinct, it is called the spectrum of matrix A, renamed as $spec\ A$:

$$spec\ A=\begin{pmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_s\\ m_1 & m_2 & \cdots & m_s\end{pmatrix}.$$

$\lambda_1,\lambda_2,\cdots,\lambda_n$ are distinct, and m_i are the algebraic multiplicity of λ_i.

3.2 LAPLACE MATRIX

Laplace matrix is matrix L, which is a $n\times n$ dimensional symmetric matrix whose element L_{ii} on the diagonal line is the degree of node i, while the elements L_{ij} on other off-diagonal means the connection of i and j. Laplace matrix usually has three kinds of representations, respectively, $D-A$, $A-D$ and $D+A$, and the calculation formula are as follows:

$$L_{D-A}=[l_{i,j}]=\begin{cases}d_{ij} & i\neq j\\ -\sum_{k=1}^{n}d_{i,k} & i=j\end{cases},$$

$$L_{A-D}=[l_{i,j}]=\begin{cases}-d_{ij} & i\neq j\\ \sum_{k=1}^{n}d_{i,k} & i=j\end{cases},$$

$$L_{D+A}=[l_{i,j}]=\begin{cases}d_{ij} & i\neq j\\ \sum_{k=1}^{n}d_{i,k} & i=j\end{cases}$$

d_{ij} is the Euclidean distance between two feature points of G.

If G is connected, then the second smallest eigen value λ_2 of Laplace matrix is positive, and its corresponding eigen vector (called Fiedler vector) contains important information of the graph, i.e. the value of elements or (including positive and negative ones) reflects the relationship between the corresponding vertices.

4 The method of traffic zone division

4.1 THE PROCESS OF DIVISION

Making use of the spectral graph theory, the process of division is shown as follows:

1) Primarily divided the study area into subareas according to the administrative division and number them. The size of planed area should be considered and comparatively small administrative division should be applied (such as small village, town or county administrative division);
2) Calculate and normalize the Euclidean distance in appliance with the index value and construct the Laplace matrix.
3) Calculate the second eigen value λ_2 of Laplace matrix and the corresponding eigen vector (Fiedler vector)
4) Integrate the original traffic zones into two, satisfying the value of each element of eigen vector.
5) New residential parks containing most original traffic zones are selected to determine whether the number of original traffic zones is greater than the fixed M;
   1. If it is larger than M, establish the Laplace matrix for the new community, and return to step 3) for further division of the community;
   2. If it is smaller than M, stop the division.

The value of M is determined by the preliminary division of the traffic zone's administrative and planning level, as well as the accuracy. For example, urban traffic planning if applied, and the planning is required to be as detailed as traffic facilities of inner-countries and intra-countries., then the preliminary divided traffic zone should be limited within the county, in accordance with the administrative divisions below the county level (such as town). If there are 20 towns in average, then the minimum M can be set as 20 or below. Then, in the (4th) step, according to the value of elements of eigen vector, the original traffic zone is divided into two parts in the following way: choose the critical value S of eigen vector $F=(f_1, f_2, \cdots, f_n)$ according to certain rules, and the original traffic zone corresponding to $f_i \geq S$ is regarded as one traffic zone, and the rest is set as another traffic zone. The ways to determine the critical value of S are as follows:

1. Bisection method. To order the elements in eigen vector (Fiedler vector) from small to large (or from large to small), and the value of the element in the middle is set as the critical value, i.e., the new traffic zones contain the same number of the original zones.
2. The average method. It can be proved that the sum of all elements of eigen vectors is zero, so is their mean, i.e. $S=0$.
3. The method of maximum distance Order the elements in eigen vector (Fiedler vector) from small to large (or from large to small), and calculate the distance between two adjacent elements, then set the maximum distance as S.

4.2 CALCULATION OF EUCLIDEAN DISTANCE

Euclidean distance is an arithmetic that calculates the distance between two points in space. In this paper, the Laplace matrix is constructed by making use of the Euclidean distance calculated by various index. Firstly, make sure each traffic zone shares the same index. Knowing from above analysis, the factors affecting the division of the community are economic indicators, land development intensity, population of the community, area of community and the distance between the communities. Taking the four as examples, suppose that the economic index of the community i is A_i, the land development intensity is B_i, the population is set as C_i, area is set as D_i, and the spatial distance is set as L_{ij}, then the Euclidean distance between the two different communities can be expressed as [11]:

$$d_{ij}=\sqrt{(A_i-A_j)^2+(B_i-B_j)^2+(C_i-C_j)^2+(D_i-D_j)^2+L_{ij}^2}$$

d_{ij} is the Euclidean distance between i and j, and the Euclidean distance between each two communities can be obtained, then Laplace matrix based on Euclidean distance can be constituted as shown above.

Economic index A_i, population C_i and area D_i can be obtained by the mean of the number of communities divided by all communities; the index of land development intensity B_i is obtained by floor area ratio (FAR) of this zone divided by that of the whole studied area; spatial distance index L_{ij} is the value of the distance between geometric centres of two residential parks divided by the radius of a circle with the same area in the researched area [12].

4.3 CALCULATION PROCEDURE

Generally, the iterative calculation of matrix adopts MATLAB for programming for now. The MATLAB software provides a comprehensive solution for many scientific fields that need numerical calculations by integrating the numerical analysis, matrix calculation, data visualization, modelling and simulation of nonlinear dynamic systems and many other powerful functions into an easy-to-use Windows environment, [13] This paper, using MATLAB to program, calculate the Euclidean distance, eigenvalue, eigen vector, etc., between every two communities for each step.

5 Evaluation of division accuracy

The feasibility and validity of these traffic zones divided by using above methods should be evaluated according to

certain indexes, because the division has an important influence over the subsequent traffic forecast and traffic distribution, the appropriate division of traffic zone is the key point of the research. The following criteria can be adopted for evaluation:

(1) The population of each traffic zone being finally divided should keep a balance. In order to ensure the precise value of traffic flow between traffic zones, the population of each traffic zone should not differ too much. Assuming the suitable population is P, the population of each community is $P_i(i=1,2,3..n)$, the allowed error limit is ΔP, and then the population should satisfy the following relations: $\begin{cases}|P-P_i|\le\Delta P(i=1,2,3..n)\\|P_i-P_j|\le\Delta P(i\ne j)\end{cases}$.

(2) The uniformity of area of traffic zones: As a key factor for traffic zone division, the size of the areas not only influences the internal travel probability, but also affects the flows among traffic sources. Assume that suitable area is D, the subarea is $D_i(i=1,2,3..n)$, the allowed error limit is ΔD, then the area of residential park should satisfy the following relations (if a residential park contains hills, waters, then that area shall be deducted): $\begin{cases}|D-D_i|\le\Delta D(i=1,2,3..n)\\|D_i-D_j|\le\Delta D(i\ne j)\end{cases}$

(3) The economic level of all traffic zones should reach a certain standard equally. The economic level is an important factor to affect traffic generation and attraction. Developing area usually embraces little transportation, while developed area embraces large traffic flow, with huge travel and attraction. Therefore, the economic level of all traffic zones should be kept at an appropriate level.

6 Case studies

Taking Dongguan rail transit network planning and passenger flow forecast as an example, analysing the divisions of the traffic zone by making use of the above-mentioned methods. Dongguan, borders among Guangzhou, Huizhou and Shenzhen, covers a total area of 2465 km2, with 35 town-level administrative units and 591 village-level administrative units, which are in jurisdiction. Dongguan, due to historical reasons, has no county-level administrative structure as the usual administrative regions would have, but has 35 town-level administrative units directly under jurisdiction.

6.1 PRELIMINARY DIVISION OF THE TRAFFIC ZONE

Dongguan holds 35 town-level administrative units and 591 preliminary divided traffic zones with 16.8 town-level administrative unit in average. In order to appropriately improve the prediction accuracy, M is set as 10.

6.2 PROCESS AND RESULTS OF TRAFFIC ZONE DIVISION

In this paper, the most economically developed south-western area is taken as an example for traffic zone division according to the above method. The south-western area embraces 4 town-level administrative units, including Houjie, Shatian, Humen and Changan, and 84 village-level administrative units. According to the 2011 Dongguan statistical yearbook and regulatory detailed planning of the four towns, the data of population, GDP and land development intensity of the 84 village-level administrative units can be easily obtained. See Table 1.

TABLE 1 Data summary of some village-level administrative

Village name	The town it is located in	GDP (billion CNY)	FAR	Population	Area (km2)
Chenwu	Houjie	0.96	1.7	11503	2.0
Mintian	Shatian	0.48	0.8	14028	5.9
Zexu	Humen	1.05	1.4	32764	2.5
Wusha	Changan	2.58	1.9	112356	10

In 2011, the average population in the 84 village-level units is 21,022, with the average GDP of 780,000,000 CNY, the average FAR of 1.1, and the average area of 6.03 km2, The radius of the circle with the same area is 12.7km.

The steps of division of traffic zones are as follows:

1) Normalizing the data, and calculate the index of the population, economic development, land development -intensity and the area, (calculate the spatial distance respectively) of the village-level administrative units seeing table 2.

TABLE 2 Village administrative unit index summary

Village name	The town it is located in	A_i	B_i	C_i	D_i
Chenwu	Houjie	1.23	1.55	0.55	0.33
Mintian	Shatian	0.62	0.73	0.67	0.98
Zexu	Humen	1.35	1.42	1.56	0.41
Wusha	Changan	3.30	1.73	5.34	1.66

2) Calculate the Euclidean distance among villages, and create the Laplace matrix. See Table 3.

TABLE 3 Euclidean distances for some village-level administrative

Village name	Chenwu	Mintian	Zexu	Wusha
Chenwu	0	1.3846	1.2788	5.5594
Mintian	1.3846	0	1.9467	5.7084
Zexu	1.2788	1.9467	0	4.5508
Wusha	5.5594	5.7084	4.5508	0

3) According to the calculation, λ_2 = 1.6, the corresponding eigen vector (containing 84 elements) can be calculated, and its critical value = 0 derived from the average method;

4) At first, combine the divided zone into two traffic zones, with 38 and 46 village-level administrative units respectively.

5) Then divide the latter one into two subzones, with 25 and 21 village-level administrative units individually.
6) Finally, divide the area into 10 subzones. (See Figure1).

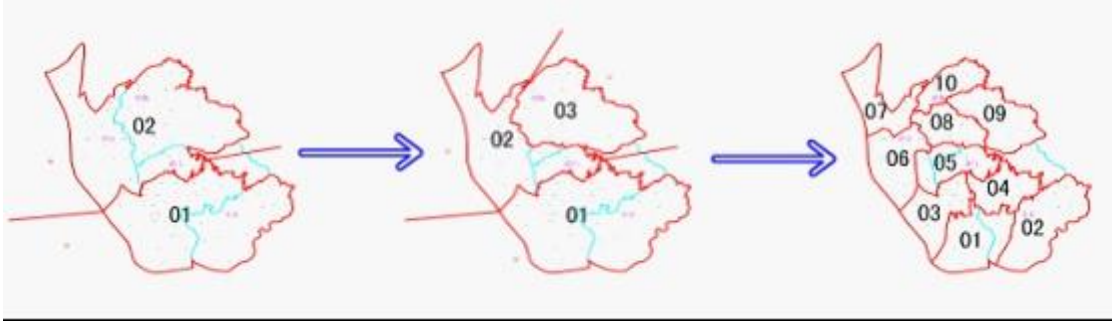

FIGURE 1 Traffic zone division process of Dongguan south-western area

With the similar steps, Dongguan is divided into 66 traffic zones as a whole. As a results, some traffic zones break the town-level administrative region, (for example, the No. 05 traffic zone is composed by some village-level administrative units of Houjie Town and Humen town), and they are connected with Dongguan town-level administrative units for continuous development, forming a dense urban cluster pattern, with closely related economy and traffic with adjacent towns.

6.3 EVALUATION OF DIVISION ACCURACY

In this paper, the average population and area of the 66 traffic zones are set as standard, (not including the large-scale mountain and water), and the allowing error is controlled within 15%. At the same time, Dongguan is a prefecture-level city, with medium-sized area, but its quantity of population and economic development level are above the medium city scale, so it is reasonable that the traffic zones are more than 50.

According to the requirements of planning administrative departments, the largest difference between traffic zones in population, GDP and area should be controlled within 30%. As is calculated, the largest difference of population is 13%, between NO.05 and No .07; largest GDP difference is 22%, between No.17 and No, 29; largest area difference is 28%, between No.33 and No.64. , Thus, the precision of division is totally in accordance with the planning requirements.

7 Conclusions

Traffic zone division is the priority and the most basic step for all kinds of transport planning and passenger flow prediction, and the results are directly related to the precision and difficulty of the passenger flow prediction. Through the introduction of spectral graph theory, a new method for traffic zone division is proposed. The divisions of the traffic zones are realized by adopting the key factors that affect the divisions, normalizing the index, calculating the Euclidean distance and creating the Laplace matrix. Dongguan, as an example, is divided into several traffic zones according to this method, and the division results conform to the planning administrative departments' requirements in precision.

References

[1] Li Xiaodan 2009 *Theory and method of city road network traffic zone division* Doctoral dissertation of Tongji University
[2] Lu Huapu 2006 *Theory and method of traffic planning* Beijing: Tsinghua University press
[3] Li Xiao-dan, Yang Xiao-guang, Chen Hua-jie 2009 Study on traffic zone division based on spatial clustering analysis *Computer Engineering and Applications* **45**(5) 19-22
[4] Zhao Jin-huan, Li Wen-quan 2009 Improvement of the Traffic District Partition in Resident Trip Investigation *Journal of Transportation Engineering and Information* **7**(2) 110-115
[5] Lu Yu-qiang, Qin Yong, Jia Li-min etc. 2010 Dynamic Traffic Zone Partition Based on Cluster Analysis of Taxi GPS Data *Logistics technology* **2010**(216) 86-89
[6] Yang Bo, Liu Hai-zhou 2007 Improvement of the Method about the Partition of Traffic Zone *Traffic and Transportation* **7**
[7] Liu Yi-fei 2011 Traffic District Division Method and Application Based on Fuzzy Cluster Analysis *Logistics Sci-Tech* **2011**(9) 25-28
[8] Ma Chao-qun, Wang Rui, Wang Yu-ping etc. Calculating Method of Traffic Zone Radius in City Based on Inner Trip Proportion *Journal of Traffic and Transportation Engineering* **7**(1) 68-72
[9] Dai Bing-kui, Hu Ji 2010 Double-level Division Method of Traffic Zone *Railway Transport and Economy* **32**(11) 86-89
[10] Zhuang Zhenhua 2010 *Data classification of gene expression spectrum based on Laplace spectrum* Anhui University master's paper
[11] Chung F R K 1997 Spectral Graph Theory *American Mathematical Society Providence* Rhode Island
[12] Luo Fengtao 2012 Application of Euclidean distance matrix analysis (EDMA) in in medical morphology research *Anatomy research* **34**(1) 50-54
[13] Chen Huaichen, Wu Dazheng, Gao Xiquan 2003 *MATLAP and its application in electronic information courses* Beijing: Publishing house of electronics industry

Authors

Wu Shimei, born in July 13, 1978, China

Current position, grades: School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, P. R. China.
University studies: Ph.D. candidate of transportation science and engineering
Scientific interest: urban planning, transportation planning, traffic management, traffic demand forecasting, etc.
Publications: 10 publications
Experience: He has often worked as a business consultant in more than 10 enterprises.

Pei Yulong, born in October 20, 1961, China

Current position, grades: Traffic College, Northeast Forestry University, professor
University studies: Ph.D. in Transportation Planning and Management(2002, Southeast University)
Scientific interest: transportation planning, transportation safety, traffic control etc.
Publications: 300 scientific papers and 18 books

Cheng Guozhu, born in August 16, 1977, China

Current position, grades: Ph.D. in Road and Railway Engineering(2007, Harbin Institute of Technology)
University studies: School of Transportation Science and Engineering, Harbin Institute of Technology, associate professor
Scientific interest: transportation safety
Publications: 70 scientific papers and 6 books