Journal of Southeast University (English Edition) Vol. 27, No. 1, pp. 83 – 87 Mar. 2011 ISSN 1003—7985

Application of discrete choice model in trip mode structure forecast: a case study of Bengbu

Ren Gang[1] Zhou Zhuping[1] Zhang Haoran[2]

([1] School of Transportation, Southeast University, Nanjing 210096, China)
([2] Zhuhai Transport Bureau, Zhuhai 519001, China)

Abstract: In order to find the main factors that influence the urban traffic structure, a relational model between the travelers' characteristics and the trip mode choice is built. The data of urban residents' characteristics are obtained from statistical data, while the trip mode split data is collected through a trip survey in Bengbu. In addition, the discrete choice model is adopted to build the functional relationship between the mode choice and the travelers' personal characteristics, as well as family characteristics and trip characteristics. The model shows that the relationship between the mode split and the personal, as well as family and trip characteristics is stable and changes little as the time changes. Deduced by the discrete model, the mode split result is relatively accurate and can be feasibly used for trip mode structure forecasts. Furthermore, the proposed model can also contribute to find the key influencing factors on trip mode choice, and restructure or optimize the urban trip mode structure.

Key words: trip mode split; trip mode structure; discrete choice model; forecasting

doi: 10.3969/j.issn.1003 – 7985.2011.01.017

Discrete choice analysis is an important method in transportation research. In the early 1970s, McFadden[1] at the Massachusetts Institute of Technology had made great progress on the theoretical study, which led Ben-Akiva et al.[2] to put this model into practice. Traditionally, multinomial logit (MNL) and nested logit (NL) models have been used to model the choice among alternative modes or classes in travel demand modeling[3].

Based on the hypothesis of the utility maximization rule in the trip mode choice, the discrete choice model has two basic hypothesis conditions: 1) Travelers are the basic units in the traffic behavior determining process; 2) Travelers will choose the maximal utility scheme, and the scheme utility is determined by the scheme and travelers' characteristics[4].

Nowadays, a massive amount of literature on modal split forecasting using the discrete choice model has been published in China. The method of the discrete choice model used in the modal split has two main steps: ① Collecting data of urban residents characteristics and trip mode split, and ② Building the functional relationship by the discrete choice model[5-7]. The logit model is widely used in a variety of transport-related choice contexts, and it has also been promoted into practices in the transportation planning field[8]. However, in China, few discrete choice models have been proposed by analyzing the survey data, whether stated data or revealed preference data. Concerning this aspect, researchers in other countries are more interested in the survey data, and recently have focused on the design of stated choice experiments. For example, Bliemer et al.[9-10] focused on designing several choice situations in stated choice models, which can lead to improvements in the reliability of parameter estimation derived from discrete choice models. In addition, constructing a logit model of both revealed and stated preference data is also a great development in this field[11-12].

1 Review of Discrete Choice Models

Limited by the data collecting method, this paper cannot obtain the revealed data. As shown in Fig. 1, the stated trip survey data is used to validate the following results: There are some stable relationships among travelers' personal characteristics, family characteristics, trip characteristics and the mode choice, and the urban traffic structure changes constantly with the passage of time due to the changes in mode choice's influencing factors such as personal characteristics, family characteristics, and trip characteristics.

It has been difficult to acquire urban traffic structure statistical data in the past years, as the data has originated from trip surveys, which is an academic survey consuming massive manpower and material resources. From the urban statistical data, personal characteristics, family characteristics, and trip characteristics can be extracted easily. Based on these analyses, the relation model for traveler characteristics factors and mode choice has a high practical value.

The relationship model between traveler characteristics factors and the individual trip mode choice can be built by the discrete choice model. After modeling, we can identify the model applicability by demonstrating its effectiveness. An example of Bengbu city is discussed in this paper, using the trip survey data in 2007 for modeling and the trip survey data in 2002 for validation. The data used in this paper is the stated data.

In discrete choice models, the selection process is based on the random utility function $U(k)$, which is expressed as[13]

$$U(k) = V(k) + e(k) \tag{1}$$

where $V(k)$ is the fixed utility of scheme k, and $e(k)$, that obeys a certain probablity distribution, is the random term. Random utility shows the sensation evaluation value in the selection process, so scheme k is chosen when its $U(k)$ is greater than that of any other schemes, and the selection probability of scheme k is

$$p(k) = [U(k) > U(j), \forall j(\neq k) \in K] \tag{2}$$

Received 2010-09-19.

Biography: Ren Gang (1976—), male, doctor, researcher, rengang@seu.edu.cn.

Foundation item: The National Natural Science Foundation of China (No. 50738001, 51078086).

Citation: Ren Gang, Zhou Zhuping, Zhang Haoran. Application of discrete choice model in trip mode structure forecast: a case study of Bengbu [J]. Journal of Southeast University (English Edition), 2011, 27(1): 83 – 87. [doi: 10.3969/j.issn.1003 – 7985.2011.01.017]

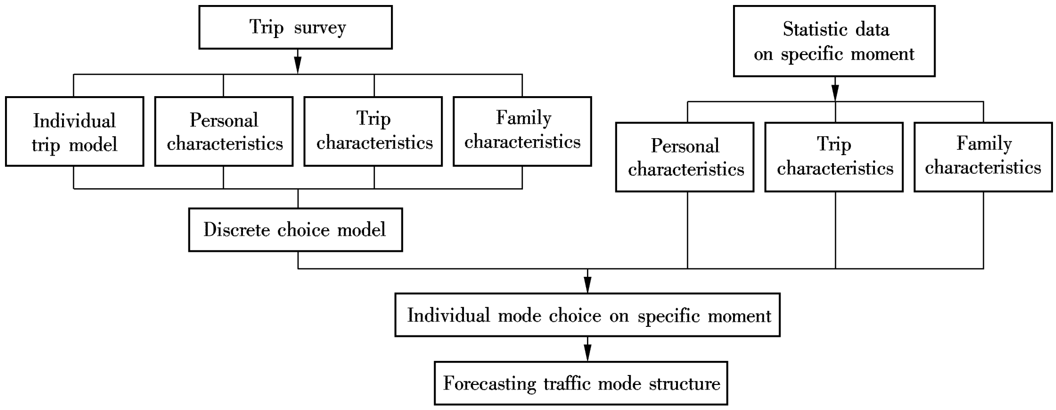

Fig. 1 Discrete model forecasting trip mode structure

where K is the scheme set. Substituting Eq. (2) into Eq. (1), we can obtain

$$p(k)=p[e(j)<V(k)-V[j]+e(k),\ \forall j(\neq k)\in K]=\int_{e(k)}F[V(k)-V(j)+e(k),\ \forall j(\neq k)\in K]f_k(x)\,\mathrm{d}x \tag{3}$$

where $F(x)$ is the probability distribution function, and $f_k(x)$ is the probability density function of probability variable $x(x=e(k))$. First, assume that $e(k)$ is fixed and calculate the value of $e(j)$, which is the simultaneous probability distribution function value of scheme j; then multiply with the probability changes of $e(k)$ when it changes; finally, integrate the multiplication result.

It is assumed that the random term $e(k)$ of the utility function in Eq. (1) is independent and it obeys the same Gambel distribution. Using probability variable x to represent $e(k)$ and θ as the parameter, the distribution function of the random term can be deduced as

$$F_e(x)=\exp\{-\theta\mathrm{e}^{-x}\}\qquad \theta>0;\ -\infty<x<\infty \tag{4}$$

Substituting Eq. (4) into Eq. (3), we can obtain

$$\begin{aligned}p(k)&=\int_{-\infty}^{\infty}\prod_{j\neq k}\exp\{-\theta\exp[-V(k)+V(j)-x]\}\cdot\\&\quad\theta\mathrm{e}^{-x}\exp(-\theta\mathrm{e}^{-x})\,\mathrm{d}x=\\&\quad\int_{-\infty}^{\infty}\prod_{j}\exp\{-\theta\exp[-V(k)+V(j)-x]\}\theta\mathrm{e}^{-x}\mathrm{d}x=\\&\quad\int_{-\infty}^{\infty}\exp\Big\{-\theta\mathrm{e}^{-x}\sum_j\exp[V(j)-V(k)]\Big\}\theta\mathrm{e}^{-x}\mathrm{d}x=\\&\quad\frac{\mathrm{e}^{V(k)}}{\sum_j\mathrm{e}^{V(j)}}\end{aligned}$$

This equation is the basic formula of the multinomial logit (MNL) model and it belongs to the discrete models. In the application of the MNL model, the key point is the relationship between the chosen probability and the characteristic variables. Hence, according to the characteristic variables x_i, the probability of selector i choosing scheme k can be expressed as

$$P(y_i=k)=\frac{\exp(\beta_k' x_i)}{\sum_{l=0}^{K}\exp(\beta_l' x_i)}\qquad k=0,1,\dots,K \tag{5}$$

The relationship between each selector and his/her choice probability is described by the above model, but the model also has uncertain solutions in the solving process. So coefficient standardization is necessary. Let $\beta_0=0$, then the MNL model is

$$P(y_i=k)=\frac{\exp(\beta_k' x_i)}{1+\sum_{l=1}^{K}\exp(\beta_l' x_i)}\qquad k=1,2,\dots,K$$

and

$$P(y_i=0)=\frac{1}{1+\sum_{l=1}^{K}\exp(\beta_l' x_i)} \tag{6}$$

Based on Eq. (6), we denote the probability for selector i to choose scheme k as P_{ik}, and each logarithmic probability can be given by

$$\log\left(\frac{P_{ik}}{P_{ij}}\right)=x_i'(\beta_k-\beta_j)\qquad k=j;\ j=0,1,\dots,K \tag{7}$$

If $j=0$ and considering the former assumption that $\beta_0=0$, Eq. (6) can be written as

$$\log\left(\frac{P_{ik}}{P_{ij}}\right)=x_i'\beta_k\qquad k=j;\ j=0,1,\dots,K \tag{8}$$

2 Modeling and Calculation

17 954 groups data from the trip survey in 2007 of Bengbu city are obtained to model and verify the discrete choice model. Furthermore, the survey data in 2002 is used for validation.

The main trip modes in Bengbu include walking, bicycle (electric bicycle), motorcycle, car, bus and taxi. Personal characteristics, family characteristics, and trip characteristics are the factors that affect mode choice. Sex, age, education level, occupation, bus monthly ticket ownership and driving license ownership are in the category of personal characteristics, while income, bicycle (electric bicycle), motorcycle and car ownership belong to family characteristics; trip purpose, trip distance, travel time, and trip costs belong to trip characteristics. Tab. 1 shows the variables and the corresponding meanings in the discrete choice model.

Tab. 1 Discrete choice model variables and corresponding meanings

Influencing factors		Variables	Instruction
Personal characteristics	Sex	SEX	Female 0, male 1
	Age	AGE0-AGE7	Divided into 8 age segments
	Education level	EDU0-EDU3	Divided into 4 levels
	Occupation	OC1-OC9	Divided into 9 occupations
	Bus monthly ticket ownership	MT	No 0, yes 1
	Driving license ownership	DL	No 0, yes 1
Family characteristics	Family income	HINC	divided into 6 grades
	Family bicycle(electric bicycle)	BIC	0, 1, 2, …
	Motorcycle ownership	MOT	0, 1, 2, …
	Car ownership	CAR	0, 1, 2, …
Trip characteristics	Trip purpose	TP1-TP9	Divided into 9 purposes
	Trip distance	TD	Trip kilometers
	Travel time	TT	Time difference of start and arrival
	Trip cost		Reference variables

The MNL model concerning personal, family, trip characteristics and mode choice is built. Then, model parameters are solved according to the following equation:

$$\log \frac{P_k}{P_1} = \mathrm{CONS}_k + \mathrm{C\_SEX}_k\mathrm{SEX} + \sum_{m_1=0}^{7}\mathrm{C\_AGE}_{m_1k}\mathrm{AGE}_{m_1} + \sum_{m_2=0}^{3}\mathrm{C\_EDU}_{m_2k}\mathrm{EDU}_{m_2} + \sum_{m_3=1}^{9}\mathrm{C\_OC}_{m_3k}\mathrm{OC}_{m_3k} + \mathrm{C\_MT}_k\mathrm{MT} + \mathrm{C\_DL}_k\mathrm{DL} + \mathrm{C\_HINC}_k\mathrm{HINC} + \mathrm{C\_BIC}_k\mathrm{BIC} + \mathrm{C\_MOT}_k\mathrm{MOT} + \mathrm{C\_CAR}_k\mathrm{CAR} + \sum_{m_4=1}^{9}\mathrm{C\_TP}_{m_4k}\mathrm{TP}_{m_4k} + \mathrm{C\_TD}_k\mathrm{TD} + \mathrm{C\_TT}_k\mathrm{TT} \tag{9}$$

where subscripts $k = 1, 2, \ldots, 6$ represent modes of walking, bicycle (electric bicycle), motorcycle, car, bus and taxi, respectively. Bus is taken as the basic variation for standardization in parameters solving[14]. The modeling and calculation process is shown in Fig. 2 and the results are listed in Tab. 2.

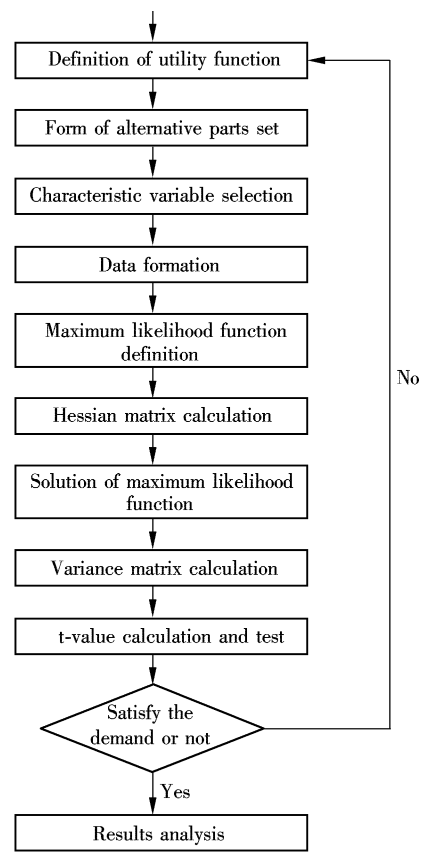

Fig. 2 Modeling and solving discrete model

3 Model Results Analysis

3.1 Model precision evaluation

t-value, McFadden determination coefficient ρ^2 and modified McFadden determination coefficient $\bar{\rho}^2$ are common test indices. t-value can judge whether a single variable significantly affects the selection result or not after the t-test. The whole and superior degrees are judged by ρ^2 and $\bar{\rho}^2$, respectively. Just as the square of the correlation coefficient, ρ^2 covering 0-1 and the value approaching to 1 show the better goodness of fit[15]. In the model of this paper, the high McFadden determination coefficient of $\rho^2 = 0.645$ and $\bar{\rho}^2 = 0.639$ proves that this method has the advantage of high goodness of fit.

The discrete choice model is constructed by the modeling data of 2007, and the model precision evaluation results are shown in Tab. 3. In addition, deducing results for traffic structure in 2002 are given in Tab. 4. Tab. 3 and Tab. 4 show that this model has high precision in both fitting current traffic structure and deducing the traffic structure at other times.

The model precision evaluation shows that the model is accurate and thus appropriate. However, the calculation-precision of the taxi mode is not as high as expected since we ignore some important influencing factors. The main reason for choosing the taxi mode is to save time in an emergency, occupying 74% of all the taxi travelers.

3.2 Model-based analysis

By analyzing the discrete choice model on the basis of the trip survey data of Bengbu, we obtain some useful results as follows.

1) Bus monthly ticket ownership, trip distance and travel time are the main factors influencing other mode choices.

2) Travelers with the purposes of leisure, entertainment and sports prefer to choose walking.

3) Bicycle (electric bicycle) ownership significantly influences bicycle or electric bicycle mode choice. Commuters are

Tab. 2 Discrete choice model coefficient and t-test

Corresponding independent variables	$\log(P_1/P_5)$		$\log(P_2/P_5)$		$\log(P_3/P_5)$		$\log(P_4/P_5)$		$\log(P_5/P_5)$	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
SEX	0.04	0.59	0.49	8.22	1.26	9.29	0.80	3.34	0.38	2.22
AGE1	−0.48	−3.09	−0.12	−0.76	−1.60	−3.24	−3.51	−3.94	−1.46	−2.28
AGE2	−0.70	−2.52	−0.41	−1.64	−2.48	−3.96	−2.97	−2.41	−1.49	−1.70
AGE 3	−0.61	−2.46	0.04	0.16	−2.08	−3.62	−3.21	−2.70	−1.02	−1.24
AGE 4	−0.51	−2.21	0.17	0.75	−1.79	−3.18	−2.66	−2.27	−1.41	−1.74
AGE 5	−0.34	−1.44	0.16	0.70	−2.30	−4.06	−2.77	−2.37	−1.12	−1.38
AGE 6	−0.30	−1.22	0.32	1.34	−2.49	−4.17	−3.76	−3.07	−1.71	−2.03
AGE 7	−0.09	−0.32	−0.54	−1.96	−2.61	−3.50	−3.32	−2.28	−2.45	−2.63
EDU1	−0.05	−0.64	0.16	2.23	0.27	1.66	0.63	1.75	−0.03	−0.12
EDU2	−0.10	−0.88	−0.03	−0.28	0.38	1.92	1.28	3.26	0.56	2.12
EDU3	3.78	3.73	1.73	1.99	3.36	3.45	6.10	5.94	3.93	3.46
OC1	−0.20	−0.90	−0.55	−2.75	−2.42	−4.81	−1.77	−1.58	−1.06	−1.39
OC2	−0.03	−0.21	0.21	1.91	−0.06	−0.30	0.11	0.24	−0.48	−1.22
OC 3	−0.03	−0.18	0.08	0.57	−1.53	−4.39	−0.24	−0.38	0.21	0.51
OC 4	−0.26	−2.02	−0.16	−1.44	−0.13	−0.64	−0.33	−0.81	0.18	0.55
OC5	0.11	0.63	0.14	0.92	0.48	2.02	1.76	4.22	1.92	5.75
OC6	0.03	0.14	−0.73	−4.00	−0.90	−1.73	−0.39	−0.38	0.31	0.59
OC7	0.04	0.21	−0.97	−4.99	−0.93	−1.64	0.42	0.49	−0.34	−0.61
OC8	−0.60	−1.67	0.18	0.69	−0.88	−1.79	−5.78	−0.53	0.67	1.10
MT	−3.39	−39.04	−3.33	−43.63	−3.36	−17.89	−2.90	−9.28	−1.91	−9.26
DL	0.04	0.26	0.51	4.59	2.48	17.99	2.27	9.67	1.04	4.65
HINC	0.07	2.36	−0.01	−0.32	0.17	3.65	0.24	4.68	0.12	2.10
BIC	−0.09	−2.05	0.92	24.07	−0.32	−4.15	−0.54	−3.44	0.00	−0.03
MOT	0.35	3.79	0.07	0.86	2.16	19.58	−0.67	−2.55	−0.42	−1.75
CAR	−0.21	−0.99	−0.33	−1.57	−1.04	−3.33	2.81	12.25	0.80	2.72
TP1	−0.08	−0.40	0.97	5.22	0.53	1.62	−0.30	−0.59	−1.80	−5.46
TP2	−0.25	−1.15	0.38	1.75	−0.07	−0.17	−0.77	−1.08	−2.30	−3.68
TP3	−1.32	−3.31	−0.30	−0.98	0.05	0.10	1.39	2.17	−0.14	−0.29
TP4	0.64	3.09	0.04	0.19	−0.63	−1.42	−0.29	−0.44	−1.31	−2.98
TP5	1.86	7.22	0.93	3.33	−0.05	−0.09	0.44	0.44	1.20	2.80
TP6	−0.18	−0.66	0.61	2.31	0.79	1.58	0.55	0.80	0.16	0.38
TP7	0.04	0.24	0.71	3.91	0.34	1.07	−0.23	−0.46	−1.02	−3.53
TP8	−0.49	−1.32	0.68	2.11	1.29	2.47	2.02	2.99	−0.45	−0.73
TD	−2.19	−51.10	−0.12	−10.28	0.18	12.47	0.20	8.62	0.16	7.73
TT	0.05	13.65	−0.04	−15.36	−0.13	−19.08	−0.11	−9.79	−0.13	−11.97
CONS	4.62	14.94	0.52	1.72	0.39	0.57	−0.60	−0.48	1.06	1.19

Tab. 3 Modeling data of 2007 to evaluate the accuracy of the model

Trip mode	Survey in 2007		Model calculation		Error of split rate/%
	Quantity	Proportion/%	Quantity	Proportion/%	
Walking	6 488	36.1	7 165	39.9	3.8
Bicycle, electric bicycle	4 450	24.8	4 268	23.8	−1.0
Motorcycle	998	5.6	931	5.2	−0.4
Car	262	1.5	233	1.3	−0.2
Bus	5 558	31.0	5 336	29.7	−1.2
Taxi	198	1.1	21	0.1	−1.0

Tab. 4 Speculation result of trip mode structure of 2002

Trip mode	Survey in 2002		Model calculation		Error of split rate/%
	Quantity	Proportion/%	Quantity	Proportion/%	
Walking	22861	39.0	20720	35.4	−3.7
Bicycle, electric bicycle	19 119	32.7	15 796	27.0	−5.7
Motorcycle	1 328	2.3	1 288	2.2	−0.1
Car	203	0.3	216	0.4	0.0
Bus	14 472	24.7	20 477	35.0	10.3
Taxi	562	1.0	48	0.1	−0.9

inclined to use bicycle or electric bicycle. Compared to female travelers who prefer to choose bus, male travelers are more likely to choose bicycle or electric bicycle.

4) Factors for motorcycle mode choice include motorcycle ownership and sex differences in travelers.

5) The ownership of car and driving license obviously influences car mode choice. In addition, people with a high education level are more inclined to use the car for trips.

6) Private and individual operators probably adopt the taxi mode. Moreover, work travelers also have a greater probability to choose taxi.

7) It is interesting that travelers with driving licenses are likely to use car, motorcycle, bicycle (or electric bicycle) and taxi instead of bus.

4 Conclusion

This paper builds a discrete choice model for trip mode forecasting using the survey data of Bengbu city. The following conclusions can be drawn:

1) The trip mode split deduced by the discrete model is relatively accurate, which shows that this method for traffic structure forecasting is feasible.

2) There are stable relationships between the trip mode split and the personal characteristics as well as family characteristics and trip characteristics.

3) It is difficult to acquire urban traffic structure statistical data of past years, but personal characteristics, family characteristics and trip characteristics data can be easily extracted from the urban statistical data.

4) The model in this paper can also be applied to restructuring and optimizing the trip mode structure. Through the guidance and adjustment of controllable factors, we can achieve the goal of optimizing the urban trip mode structure. Hopefully, some feasible implementing schemes for urban traffic structural congestion can be put forward.

References

[1] McFadden D. Conditional logit analysis of qualitative choice behavior [J]. *Frontiers in Econometrics*, 1974, **1**(2): 105 - 142.

[2] Ben-Akiva M, Lerman S R. *Discrete choice analysis: theory and application to travel demand* [M]. Cambridge, MA, USA: The MIT Press, 1985.

[3] Koppelman F S, Sethi V. Incorporating variance and covariance heterogeneity in the generalized nested logit model: an application to modeling long distance travel choice behavior [J]. *Transportation Research Part B*, 2005, **39**(9): 825 - 853.

[4] Guan Hongzhi. *Disaggregate model: a tool of traffic behavior analysis* [M]. Beijing: China Communications Press, 2004. (in Chinese)

[5] Liu Zhen, Zhou Xizhao. Application on nested logit mode of trip mode choice [J]. *Journal of Shanghai Maritime University*, 2006, **27**(3): 66 - 70. (in Chinese)

[6] Jin An. On methodology of parameter estimation in logit model [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2004, **4**(1): 71 - 75.

[7] Zhou Jian, Wang Shusheng, Sui Shuixian. Study on nested logit model for transport mode split and its solution [J]. *Journal of Shandong Jiaotong Universiy*, 2005, **13**(4): 28 - 31. (in Chinese)

[8] Luo Qingyu, Sun Baofeng, Wu Wenjing, et al. Predicting traffic mode split under congestion pricing based on mixed logit model [J]. *Journal of Jilin University: Engineering and Technology Edition*, 2010, **40**(5): 1230 - 1234. (in Chinese)

[9] Bliemer M C J, Rose J M, Hensher D A. Efficient stated choice experiments for estimating nested logit models [J]. *Transportation Research Part B*, 2009, **43**(1): 19 - 35.

[10] Rouwendal J, de Blaeij A, Rietveld P, et al. The information content of a stated choice experiment: a new method and its application to the value of a statistical life [J]. *Transportation Research Part B*, 2010, **44**(1): 136 - 151.

[11] Hensher D A, Rose J M. Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: a case study [J]. *Transportation Research Part A*, 2007, **41**(5): 428 - 443.

[12] Hensher D A. Empirical approaches to combining revealed and stated preference data: some recent developments with reference to urban mode choice [J]. *Research in Transportation Economics*, 2009, **23**(1): 23 - 29.

[13] Shao Chunfu. *Traffic planning* [M]. Beijing: China Railway Publishing House, 2006. (in Chinese)

[14] Jiao Pengpeng, Lu Huapu, Yang Lang. Disaggregate traffic mode choice model based on combination of revealed and stated preference data [J]. *Tsinghua Science and Technology*, 2006, **11**(3): 351 - 356. (in Chinese)

[15] Train K E. *Discrete choice methods with simulation* [M]. Cambridge: Cambridge University Press, 2003.

交通方式划分的非集计模型——以蚌埠为例

任　刚[1]　周竹萍[1]　张浩然[2]

([1] 东南大学交通学院,南京 210096)

([2] 珠海市交通运输局, 珠海 519001)

摘要:为研究影响城市交通结构的主要因素,建立出行者属性与个体出行方式选择的合理关系模型. 所有的出行者相关属性数据均由统计资料获得,方式选择数据由蚌埠市居民出行调查的数据获得. 应用非集计模型来建立个人属性、家庭属性和出行属性与方式选择间的函数关系. 建模结果表明:居民出行交通方式选择与个人属性、家庭属性和出行属性之间有较稳定的关系,其随着时间的推移变化甚微. 非集计模型所推算的交通方式结构较为精确,可用于交通方式结构的预测. 此外,所建模型亦能识别方式分担的主要影响因素,从而实现交通方式结构优化.

关键词:交通方式划分;交通方式结构;非集计模型;预测

中图分类号:U491.1