Journal of Southeast University (English Edition) Vol. 26, No. 4, pp. 633 – 637 Dec. 2010 ISSN 1003—7985

Combined model of trip mode and destination

Jiang Yu[1] Lu Jian[2]

([1] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)
([2] School of Transportation, Southeast University, Nanjing 210096, China)

Abstract: This paper analyzes the characteristics of the destination distribution of trips and proposes a stratified sampling strategy for travel mode choice. The stratified sampling strategy can reduce the size of the alternative set; thus, the computation burden of simulation is decreased. Using the stratified sampling strategy, a combined choice model of the trip mode and destination is developed based on the Bayesian theory. Simulations are carried out to verify the proposed model. The results show that the combined choice model of the trip mode and destination can efficiently simulate travelers' choice behaviors. Furthermore, the forecasting accuracy of the combined choice model is higher than the one of the gravity model. Therefore, the proposed model is a powerful tool with which to analyze travelers' behaviors in selecting the trip mode.
Key words: combined choice model; discrete choice; trip mode and destination; sampling

Since 1970s, the four-step model has been the most popular method in the transportation planning field in China, and it has successfully been used to evaluate urban infrastructure construction projects. In the four-step model, the trip distribution model and the mode choice model are very important. A typical trip distribution model is the gravity model, which can consider impedance information (e.g., trip time) between all zones. However, trip time cannot accurately reflect the accessibility and convenience between two zones in general. For example, there are multiple trip modes between two zones. And, it is difficult to estimate the accessibility and convenience of other modes using the trip time of a mode. The sharing rate curve in mode split prediction is a common method that considers the curves of time differences and time ratios. However, the curve method lacks a definite theory basis and cannot reflect the changes of trip structures in the future. Therefore, traditional methods in traffic planning are usually difficult to be applied in practice.

Recently, the discrete choice theory, which provides the foundation to simulate travelers' trip behaviors by using the micro-choice model, has been introduced into the transportation planning field. In 1974, McFadden proposed logit discrete choice which is based on the premise of a hypothesis of utilities and randomness[1-2]. In the studies, it is assumed that everyone in the system faces J alternatives; X represents the characteristics of each alternative branch, and Z represents the personal characteristics which can be observed. For example, when choosing the trip mode, the alternative branches may be car, bus and metro. X represents travel time, cost, and information which is related to each alternative branch. Z includes the personal information of travelers, such as age, income, and education. Thus, the variables of the model include traffic policy and travelers' information. Other variables which cannot be observed are estimated by the error term ε. Travelers choose the most satisfactory trip mode based on the characteristics of alternative branches and travelers' satisfaction levels for various alternative branches. In 1974, McFadden discussed the logit model and its characteristics, and then proposed a theoretical system of the discrete choice model[1].

This paper studies the combined choice behaviors of travelers' trip and destination based on discrete choice. Then the performance of the combined choice model is illustrated by simulations. Finally, the model is compared with the gravity model and some conclusions are obtained.

1 Combined Choice Model of Trip Mode and Destination

Assuming that a traveler can select a trip mode and a destination in set C_n, the utility of the choice branch is $U_{\mathrm{md}} = \bar{V}_{\mathrm{m}} + \bar{V}_{\mathrm{d}} + \bar{V}_{\mathrm{md}} + \varepsilon_{\mathrm{md}}$, which is composed of the trip mode and the destination. Here, $\bar{V}_{\mathrm{m}}$ is the deterministic part of the utility which changes with the trip mode; $\bar{V}_{\mathrm{d}}$ is the deterministic part which changes with the destination; $\bar{V}_{\mathrm{md}}$ is the deterministic part which changes with the trip mode and destination simultaneously; $\varepsilon_{\mathrm{md}}$ is a random item of the utility, which obeys the duplicate-exponent distribution, and it is independent.

According to the discrete choice theory, the combined model, which is relevant to the trip mode and the destination of travelers, is described as[3-4]

$$P_n(i) = P_{\mathrm{md}} = \frac{\mathrm{e}^{\bar{V}_{\mathrm{m}} + \bar{V}_{\mathrm{d}} + \bar{V}_{\mathrm{md}}}}{\sum_{(\mathrm{m}', \mathrm{d}' \subset C_n)} \mathrm{e}^{\bar{V}_{\mathrm{m}'} + \bar{V}_{\mathrm{d}'} + \bar{V}_{\mathrm{m}'\mathrm{d}'}}} \tag{1}$$

The trip mode includes walk, bike, bus, taxi, motorcycle, private car, coach and official car. In the investigation, there are 204 traffic zones. Therefore, the number of choices, which is relevant to the trip mode and destination, is $8 \times 204 = 1\,632$. It is difficult to define the utility of the alternative branches, and the solution algorithm of the model is also complex. According to the IIA characteristics of the logit model, we can simplify the model.

Assume that D_n is a subset sampling in C_n. $\pi_n(D_n \mid i)$ is the conditional probability of D_n, but the precondition is that the i-th alternative branch is given. Then the combined probability density, which is composed of the selection of alternative branch i and D_n, is defined as

Received 2010-04-22.
Biography: Jiang Yu (1975—), female, doctor, lecturer, jiangyu07@nuaa.edu.cn.
Citation: Jiang Yu, Lu Jian. Combined model of trip mode and destination [J]. Journal of Southeast University (English Edition), 2010, 26(4): 633 – 637.

$$\pi_n(i, D_n) = \pi_n(D_n \mid i) P_n(i) \tag{2}$$

According to the Bayesian theory[5], if subset D_n is given, the probability that traveler n selects alternative branch i is defined as

$$\pi_n(i \mid D_n) = \frac{\pi_n(D_n \mid i) P_n(i)}{\sum_{j \subset D_n} \pi_n(D_n \mid j) P_n(j)} \tag{3}$$

Then

$$\pi_n(i \mid D_n) = \frac{\exp[\mu^* V_{ni} + \ln \pi_n(D_n \mid i)]}{\sum_{j \subset D_n} \exp[\mu^* V_{nj} + \ln \pi_n(D_n \mid j)]} \tag{4}$$

As a general rule, travelers always select the nearest destination if the purpose of travel is satisfied. Therefore, weight should be considered during selecting the subset D_n.

Then, C_n is divided into R kinds of sets according to the distance from C_n to the original zone. The number of the alternative branches in the r-th type is J_{rn}, i. e., $\sum_{r=1}^{R} J_{rn} = J$. Assuming that D_n is composed of $\bar{J}_{rn}(r = 1, 2, \ldots, R)$ in the r-th type, then the total number of D_n is $\sum_{r=1}^{R} \bar{J}_{rn} = \bar{J}$. Assuming that the selected probabilities are equal, then the probabilities are defined as [6]

$$\pi_n(D_n \mid i) = \binom{J_{r(i)n} - 1}{\bar{J}_{r(j)n} - 1}^{-1} \prod_{r \neq r(i)}^{R} \binom{J_{rn}}{\bar{J}_{rn-1}}^{-1} \quad i \in D \tag{5}$$

Let $Q_n(D) = \prod_{r=1}^{R} \binom{J_{rn}}{\bar{J}_{rn-1}}^{-1}$, then

$$\pi_n(D_n \mid i) = \frac{J_{r(i)n}}{\bar{J}_{r(i)n}} Q_n(D) \tag{6}$$

Let $q_{in} = \bar{J}_{r(i)n} / J_{r(i)n}$, then

$$\pi_n(i \mid D_n) = \frac{\exp(\mu^* V_{ni} - \ln q_{in})}{\sum_{j \subset D_n} \exp(\mu^* V_{nj} - \ln q_{jn})} \tag{7}$$

The difference between Eqs. (5) and (7) is that there is an adjunctive item ($-\ln q_{in}$) in Eq. (7). The coefficient of the adjunctive item is 1.

2 Strategy of Destination Alternative Sampling

This paper intends to sample destination zones to reduce the number of combined alternative branches. All the zones are sorted according to their distances from the original zone, and the cumulative percentage is counted up according to the condition whether the destination zone is in an appropriate position or not. Fig. 1 shows that the probability of the destination in the ten sampled zones, which are the nearest ones to the original zone, is about 57.6%. Therefore, all the zones are divided into two groups according to the order. The sampling strategy is that it randomly takes 6 zones from zone No. 1 to zone No. 10 and randomly takes 4 zones from the remaining zones, i. e., zone No. 11 to zone No. 204.

According to the combined choice formula (1), the irregular sample needs to amend the utility, and the corrected item is $-\ln q_{in}$. According to the sampling method, we take 6 zones from zone No. 1 to zone No. 10. The ratio of the sampling is 60% and the corrected value is 0.510 8. We take 4 zones from the remaining zones, i. e., zone No. 11 to zone No. 204. The sampling ratio is 2.06% and the corrected value is −3.881 6. After sampling, the number of alternative branches of the trip mode and destination is reduced to 80.

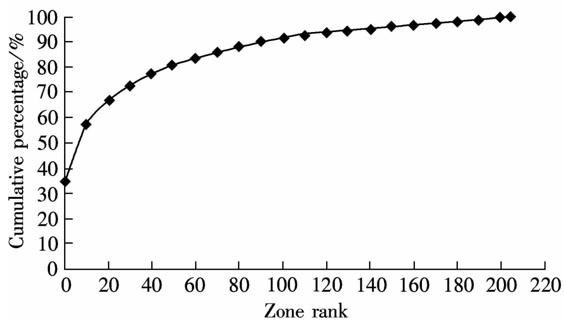

Fig. 1 Cumulative percentage of destination zone rank

3 Results of the Combined Choice Model

There are a number of factors that influence the choice of the trip mode and destination. We divide the influence factors into two groups: the socio-economic variables and the destination variables. The definition of the variables is described in Ref. [7].

We define the utility of various alternative branches according to the variables in Ref. [7]. The definition of the utility of the combined model can be seen in Ref. [8]. We discard the variables whose T-test value is less than 0.8, and repeatedly calibrate the model. The results are shown in Tabs. 1 and 2.

Tab. 1 Descriptive indicators of the combined model

Indicators	Value
Number of estimated parameters	85
Null log-likelihood	−46 823
Final log-likelihood	−31 759
Likelihood ratio test	30 128
Rho-square	0. 323
Adjusted rho-square	0. 312

From the estimated results we can see that

1) The effectiveness of the parameters' estimation is improved greatly. The number of the estimated parameters is 85; the number of the parameters, whose confidence is equal to or greater than 90%, is 76. And other parameters' T-test values are above 0.8.

2) Eight impedance variables are negative, which shows that the variation of the variables' impedance is contrary to the attraction of the variables.

3) The travelers whose ages are from 6 to 14 prefer to select the walk mode, while the travelers whose ages are from 20 to 39 are contrary. The travelers whose monthly expenditure in traffic is less than 100 yuan prefer to walk. In addition, females prefer to select to walk than males. Under the same conditions, the probability of selecting walk is improved when the original zone and the destination zone are in the same zone.

Tab. 2 Estimated results of the combined model

Variables	Estimated value	Standard deviation	T-test	Variables	Estimated value	Standard deviation	T-test
ASC2	1. 607 5	0. 375 1	4. 285 0	BETA46	−0. 901 4	0. 094 6	5. 994 3
ASC3	−1. 784 8	0. 279 9	−6. 376 9	BETA47	0. 603 2	0. 111 8	3. 651 3
ASC4	−1. 053 6	0. 537 7	−1. 959 3	BETA49	−1. 559 1	0. 339 4	−4. 593 1
ASC5	−1. 632 6	0. 239 9	−6. 805 5	BETA50	−1. 449 7	0. 364 2	−3. 980 4
ASC6	−4. 255 4	0. 445 7	−9. 547 9	BETA51	1. 389 7	0. 432 7	3. 211 5
ASC7	−3. 038 9	0. 551 7	−5. 508 1	BETA561	−0. 113 4	0. 001 6	−69. 707 8
ASC8	−3. 234 4	0. 455 5	−7. 101 1	BETA562	−0. 064 8	0. 0038	−17. 230 5
BETA1	0. 885 9	0. 365 6	2. 423 2	BETA563	−0. 143 0	0. 013 5	−10. 584 3
BETA10	−0. 858 6	0. 597 6	−1. 436 9	BETA564	−0. 116 2	0. 0044	−26. 414 3
BETA101	1. 785 2	0. 774 0	2. 175 8	BETA565	−0. 103 1	0. 0139	−7. 428 7
BETA105	0. 534 8	0. 267 1	2. 306 6	BETA566	−0. 217 9	0. 0442	−4. 925 6
BETA108	1. 169 9	0. 264 5	−2. 223 5	BETA567	−0. 064 2	0. 0012	−52. 107 6
BETA109	0. 795 8	0. 148 3	2. 002 1	BETA568	−0. 034 3	0. 004 0	−8. 516 3
BETA11	−1. 264 5	0. 753 9	4. 423 6	BETA57	−0. 581 2	0. 607 9	−0. 956 0
BETA110	−0. 445 0	0. 390 2	5. 367 9	BETA58	2. 088 6	0. 328 8	6. 351 7
BETA112	−0. 483 6	0. 138 6	−1. 677 5	BETA59	−1. 805 3	0. 328 3	−5. 498 6
BETA115	1. 355 4	0. 215 9	−1. 140 3	BETA6	−0. 394 2	0. 206 3	−1. 911 0
BETA116	1. 880 4	0. 636 1	−3. 489 7	BETA60	2. 502 7	0. 341 2	7. 335 1
BETA117	−1. 739 2	0. 609 1	6. 277 9	BETA61	−0. 614 4	0. 391 1	−1. 570 7
BETA118	0. 943 6	0. 105 9	2. 956 1	BETA62	0. 948 4	0. 334 6	2. 834 1
BETA119	−1. 724 8	1. 062 8	−2. 855 2	BETA65	−1. 757 5	0. 340 1	−5. 167 5
BETA120	2. 617 7	0. 268 9	8. 911 9	BETA66	0. 874 0	0. 134 6	6. 492 7
BETA122	1. 798 7	0. 409 8	−1. 622 9	BETA67	0. 290 5	0. 113 0	2. 570 7
BETA124	1. 930 7	0. 218 2	9. 734 3	BETA68	−2. 627 4	0. 451 1	−5. 824 1
BETA125	1. 858 7	0. 208 6	4. 388 8	BETA69	0. 930 2	0. 130 9	7. 107 7
BETA13	0. 850 6	0. 651 5	8. 848 3	BETA7	0. 897 0	0. 330 6	2. 713 4
BETA14	−0. 558 8	0. 467 3	8. 909 2	BETA70	0. 390 1	0. 109 7	3. 556 0
BETA15	0. 668 6	0. 091 7	1. 305 7	BETA701	1. 379 6	0. 069 9	19. 724 4
BETA153	0. 010 8	0. 004 8	−1. 195 8	BETA702	2. 000 2	0. 168 5	11. 868 7
BETA154	9.34×10^{-6}	7.40×10^{-7}	7. 292 5	BETA71	−2. 441 7	0. 430 9	−5. 666 8
BETA17	0. 866 9	0. 300 5	2. 265 1	BETA72	−2. 558 2	0. 557 9	−4. 585 5
BETA21	0. 874 2	1. 058 3	12. 617 6	BETA75	−1. 184 2	0. 513 2	−2. 307 7
BETA271	0. 320 0	0. 074 7	2. 885 2	BETA77	−0. 213 5	0. 134 9	−1. 582 5
BETA3	−0. 628 3	0. 050 2	0. 826 0	BETA78	0. 357 3	0. 130 2	2. 744 0
BETA35	0. 185 6	0. 049 6	4. 284 3	BETA79	−1. 320 4	0. 404 9	−3. 260 8
BETA36	0. 135 8	0. 088 0	−12. 516 0	BETA8	1. 308 9	0. 262 9	4. 978 7
BETA38	−0. 382 0	0. 155 2	3. 743 4	BETA80	−0. 374 1	0. 386 8	−0. 967 0
BETA4	−1. 040 7	0. 365 5	1. 543 5	BETA82	2. 298 1	1. 066 3	2. 155 3
BETA40	−0. 992 8	0. 130 3	−2. 462 3	BETA84	−0. 367 7	0. 388 3	−0. 947 0
BETA42	0. 173 7	0. 141 9	−2. 847 7	BETA85	3. 494 7	1. 142 4	3. 059 1
BETA43	−2. 332 2	0. 323 7	−7. 621 0	BETA87	−1. 746 2	0. 561 0	−3. 112 8
BETA44	1. 525 0	0. 254 4	1. 223 7	BETA901	3. 383 3	0. 039 4	85. 834 4
BETA45	1. 364 4	0. 373 7	−7. 204 1				

4) The travelers whose ages are from 6 to 14 or over 59 do not prefer to select the bike mode. But the low-income group tends to select bike than the high-income group. The possession of bikes in a family and the ratios of bike number to the family members have an important impact on choosing the bike mode.

5) The high-income group has a positive effect on the bus mode. The travelers who work in public institution have a positive effect on the bicycle mode.

6) The travelers who choose a taxi as the trip mode have these characteristics: monthly expenditure in traffic more than 200 yuan or the self-employed.

7) The travelers in the 20 to 39 age group prefer to select the motorcycle mode, and the travelers in the 6 to 14 age group do not prefer to select the motorcycle mode. The travelers whose monthly expenditure in traffic is between 100 and 200 yuan do not prefer to select the motorcycle mode. The travelers whose yearly income is between 30 000 and 50 000 yuan prefer to select the motorcycle mode. motorcycle ownership and the ratios of motorcycle number to the family members have an important impact on the motorcycle mode.

8) The travelers, whose age is in the 20 to 30 age group, or the monthly expenditure in traffic is above 400 yuan, or the monthly income is above 1 000 yuan, or the self-employed, prefer to select the car mode.

9) The travelers, who are more than 59 years old or work in public institutions, usually select official cars as the trip mode.

10) The travelers whose monthly expenditure in traffic is over 200 yuan do not prefer to select the coach mode. The travelers who work in public institutions prefer to select the coach mode.

4 Contrast between Combined Model and Gravity Model

This paper uses the forecast of the OD matrix to verify the effectiveness of the combined model and the gravity model. 4/5 of the original data is used for modeling, and 1/5 of the original data is used to test data.

The matrix $\boldsymbol{U} = (u_{ij})$ is composed of trip modes and the variables of the model. Trip modes are arranged as the row of the matrix, and the variables of the model are arranged as the array of the matrix. The elements of the matrix are the estimated values of the corresponding variables. If some variable is not included in the utility definition of the trip mode, the corresponding value is zero. The utility vector of the trip mode is $\boldsymbol{U}_i$, and it is divided into three parts: $\boldsymbol{U}_i = (\boldsymbol{S}_i \mid \boldsymbol{D}_i \mid \boldsymbol{C}_i)$, where $\boldsymbol{S}_i$, $\boldsymbol{D}_i$ and $\boldsymbol{C}_i$ represent the socio-economic characteristics, the trip distance, and the characteristics of the zones, respectively.

The vectors formed by the variables of the j-th trip records are divided into three parts: $\boldsymbol{V}_{sj}$, $\boldsymbol{V}_{dj}$, and $\boldsymbol{V}_{cj}$. The numbers of the rows of the three vectors are 1, 204 and 204, respectively. Therefore, the utility of trip mode i and combined alternative branches of the destination zone is

$$(\boldsymbol{M}_i \otimes \boldsymbol{D}) = \boldsymbol{V}_{sj} \cdot \boldsymbol{S}_i \otimes \boldsymbol{I} + V_{dj} \cdot \boldsymbol{D}_i + \boldsymbol{V}_{cj} \cdot \boldsymbol{C}_i \tag{8}$$

where $\boldsymbol{I}_{204 \times 1}$ is a matrix whose elements are all 1. From Eq. (8), we can obtain the utility matrix $(\boldsymbol{M} \otimes \boldsymbol{D})$.

The combined choice probability of the j-th trip record is obtained as

$$P_j = \left(\frac{\exp(m_i d_k)}{\sum_{i,k} \exp(m_i d_k)} \right)_{8 \times 204} \tag{9}$$

From Eq. (9), we can obtain the OD matrix.

We compare the combined model and the gravity model. The statistical results show that the total number of the round-trips is 13 892 and the number of the round-trips in the zones is 4 305, which is 31.0% of the total trips. From the results of the gravity model, the number of the round-trips in the zones is 1 450, which is 10.4% of the total trips. From the results of the combined model, the number of the round-trips in the zones is 4 457, which is 32.08% of the total trips, and it is slightly higher than the actual situation.

Further comparisons between the gravity model and the combined model are shown in Fig. 2 and Fig. 3. From Fig. 2 and Fig. 3, we can see that the differences between the forecast values of the combined model and the actual values fluctuate around zero. But the differences between the forecast values of the gravity model and the actual values are less than zero.

Taking the 7th zone as an example, the differences of the trip distribution between the forecast values of the two models and the statistical values are shown in Fig. 4.

It can be seen that the differences between the forecast values of the combined model and the actual values and the differences between the forecast values of the gravity model and the actual values are consistent. This validates the rationality of the two forecast models. From Fig. 4, it can be seen that the forecast results of the combined choice mode is closer to the actual data than those of the gravity model.

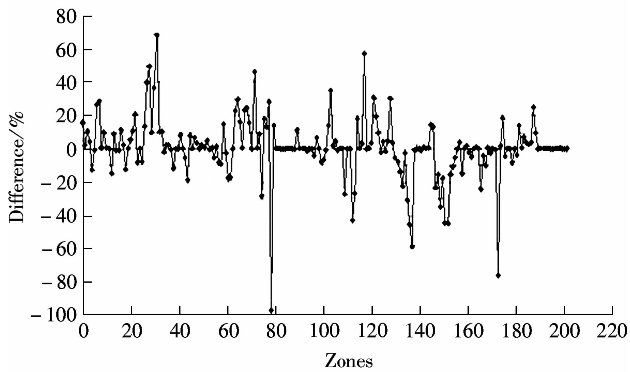

Fig. 2 Difference of zone trips between evaluation results of the combined model and actual results

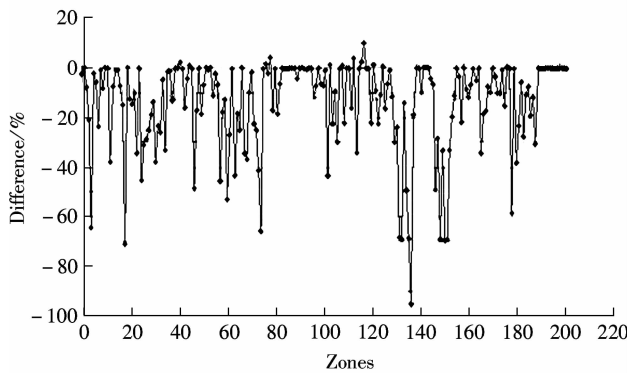

Fig. 3 Difference of zone trips between evaluation results of the gravity model and statistical results

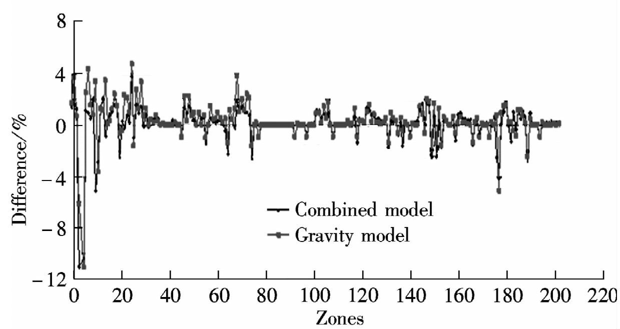

Fig. 4 Difference between the forecast values of the two models and the statistical values in the 7th zone

The combined model is more reasonable than the gravity model from the viewpoint of integrity and distribution. But, the forecast values of the two models have a similar imperfection; i. e., the forecast values of some zones are greater than the actual value. The main reason is that the restriction in selecting the destination is not enough. However, from a theoretical point of view, the combined choice model is a micro-model which simulates the personal choice of trip. The limitation of the alternative branches of destinations can flexibly add these types of constraints[9-10]. But the difficulty in data investigation and processing is increased. With the development of GIS, GPS and the computer, this problem can be solved.

5 Conclusion

A combined choice model is established based on the discrete choice theory. The results show that, compared with the gravity model, the combined choice model is better. The combined choice model is a micro-model which simulates the personal choice of travel. It has many advantages in explaining the trip activities. With the breakthroughs in computer technology and trip survey techniques, the combined choice model can deal with more alternative branches and consider more variables. Therefore, the combined choice model is a powerful tool for transportation planning in the future.

References

[1] Train K E. *Discrete choice methods with simulation* [M]. Berkeley: University of California and National Economic Research Associates, Inc, 2002: 138 − 139; 151 − 154.

[2] Rich J, Holmblad P M, Hansen C O. A weighted logit freight mode-choice model [J]. *Transportation Research Part E*, 2009, **45**(2): 1006 − 1019.

[3] Cramer J S. The early origins of the logit model [J]. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2004, **35**(5): 613 − 626.

[4] Kalouptsidis N, Psaraki V. Approximations of choice probabilities in mixed logit models[J]. *European Journal of Operational Research*, 2010, **200**(1): 529 − 535.

[5] Ka Edi. Foundations of Bayesian theory[J]. *Journal of Economic Theory*, 2007, **132**(4): 167 − 188.

[6] Hess S, Bierlaire M, Polak J W. Estimation of value of travel-time saving using mixed logit models [J]. *Transportation Research Part A*, 2004, **9**(3): 15 − 21.

[7] Jiang Y, Lu J, Wang S S. Mode and destination combined based on discrete choice theory [J]. *Journal of Highway and Transportation Research and Development*, 2007, **24**(7): 137 − 142. (in Chinese)

[8] Wang S S. On activity-based discrete choice travel demand analysis model [D]. Nanjing: School of Transportation of Southeast University, 2007. (in Chinese)

[9] Walker Joan. Mixed logit (or logit kernel) model: dispelling misconcepts of identification [J]. *Transportation Research Record*, 2002(1805): 86 − 98.

[10] Fosgerau M. Investigating the distribution of the value of travel time saving [J]. *Transportation Research Part B*, 2006, **40**(3): 688 − 707.

出行方式和目的地联合模型

姜 雨[1] 陆 键[2]

([1] 南京航空航天大学民航/飞行学院,南京 210016)
([2] 东南大学交通学院,南京 210096)

摘要:在对出行目的地空间分布特点进行分析的基础上,提出了减少选择枝个数的目的地小区分段抽样策略,该策略的应用减少了仿真的计算量.采用该抽样策略建立了基于贝叶斯理论的出行者出行方式和目的地联合选择模型.通过对该模型进行仿真计算,结果表明出行方式和目的地的联合选择模型能够有效地模拟出行者的行为选择.而且,联合选择模型的预测精度明显高于重力模型.因此,在分析出行者行为选择方面采用出行方式和目的地的联合选择模型是一种有效的方法.

关键词:联合选择模型;离散选择;出行方式和目的地;抽样
中图分类号:U491.2