Niched Genetic Algorithm Based Disaggregate Mode Choice Model

Wenzhi LIU[*,1,2], Qiuyan TAO[1] and Huapu LU[3]

[1] *School of Management,* Beijing Union University, Beijing, 100101, P.R. China
[2] *Postdoctoral research station of civil engineering,* Tsinghua University, Beijing, 100084, P.R. China
[3] *Department of civil engineering,* Tsinghua University, Beijing, 100084, P.R. China

***Abstract*—In this paper, it is built the niched genetic algorithm and Monte Carlo Method based disaggregate mode choice model using the 2006 Travel Survey Data of Beijing. Because of the model is both tour based and trip based on household and person level, at the same time, it is also considered the household vehicle allocation. So the model choice structure becomes more complex. The results shows that niched genetic algorithm and Monte Carlo Method based model can fits the survey data very well and the McFadden R-square is equal to 0.286. It means the model describes mode choice behavior fairly well.**

***Keywords* - Disaggregate choice model, niched genetic algorithm, mode.**

I. INTRODUCTION

Mode choice analysis always plays a very important role in transportation planning. Many scholars have conducted a lot of research on this area. The traditional four stage model and random utility maximization theory were used popular in prior research [1-4].

Although all these research have got significant progress, they were lack of research on human's travel behavior. At first, people's mode choice is always rely on their entire trip chain. Second, household car allocation is based on overall household utility maximization. Third, the prediction accuracy of traditional four stage model is not high because of aggregation. In recently, some researchers were used disaggregated method, nested logit and multinomial logit models. They pointed out that mode choice is not an individual behavior and choice sets have tree logit form [5-9]. However, with the complication of model decision structure, all these econometric methods will not be used again. In this paper, using 2006 Beijing travel survey data to build niched genetic algorithm based disaggregate mode choice model and to describe the procedure for estimating a model of tour based mode choice and household car allocation. This research is based on the work of Miller, Roorda and Carrasco in 2005 [10]. A technical method provided by Lerman and Manski in 1981 will be used in this paper[11].

II. BASIC THEORY

In 1985, Ben-Akiva and Lerman introduced the basic principles of the disaggregated model [12]. They pointed out that models are based on random utility theory, which assumes that people make rational decisions in order to maximize their utility. The utility $U(m,t,n)$ can be expressed as follows:

$$U(m,t,n) = V(m,t,n) + \varepsilon(m,t,n) t \in T(c,n);$$
$$m \in f(t,n) \quad (1)$$

where,

$U(m,t,n)$--- the utility of person n for trip t on chain c using mode m ;

$V(m,t,n)$--- the systematic component;

$\varepsilon(m,t,n)$--- the random utility component;

$T(c,n)$--- the set of trips on chain c for person n;

$f(t,n)$--- the set of feasible modes for trip t for person n.

For given parameter set β , the related log likelihood function L can be expressed as follows:

$$L(\beta) = \sum_h \sum_{n\epsilon H(h)} \sum_{c\epsilon T(c,n)} \log(P(m^*, t, n : \beta)) \quad (2)$$

where,

$H(h)$---a set of persons observed in household h ;

$C(n)$---a set of home-based tours for person n;

β---vector of model parameters;

$P(m^*, t, n : \beta)$ ---the simulated probability of person n choosing observed mode m^* for trip t on chain c given model parameters β .

III. MODEL CONSTRUCTION

A. Data

The data used to construct the mode choice model are taken from the Beijing 2006 travel survey. The basic rules of data cleaning are as follows:

- Remove the entire household data if exist some unrealistic data. For example, the member 2 in household 07090400014, trip mode is auto, trip duration is longer to 9 hour 42 minutes, trip distance is only 14.2 km.
- Due to lack of network date, only built the 18 zone travel time matrix through Google Map.

- The concept of trip chain in this paper means first trip is begins from home and the last one is ends at home. Thus the tours which contain only one trip are removed from data set.
- If a trip violated the choice set rules, the entire household was removed from the estimation data set. For example, some people drive from home to work and then walk back to home.
- When there are some unreasonable cases, we need to according to some other related information to make some reasonable modifications. For example, some household have only one car, but two members in this household used auto mode in the same origin and destination, at the same start and arrived time. In such cases, we modify one member's (who have no driver's license) mode from auto to passenger.

Beijing 2006 travel survey data originally contains 3016 households and 17568 trips. According with above four rules, we eventually cleaned 467 households and 3192 trips. After cleaning and modifications, the data used to estimate the model totally contain 2549 households, 14376 trips and 504 passenger mode. Where, totally 28 households contain some people drive without household car were cleaned; There are 34 households, 193 trips were cleaned and 45 auto mode were changed to passenger mode because of they have only one car, but two or more than two members used auto mode. For the household have two cars, totally deleted 1 household, 5 related trips and modified to 13 passenger modes. There are 49 households were cleaned and 446 auto mode were modified to passenger mode because of drive without license. 235 households and 1463 trips were removed because of the trip chain concept. Totally 10 households with zone number is greater than 18 were deleted.

B. Method

The procedures of model construction are as follows:

Step1. Mode choice for an individual trip maker is determined on the basis of a random utility maximization theory. The chosen mode should maximize the utility of entire tour.

Step2. Household car allocation: Car allocation decisions are made at the household level to maximize entire household utility rather than individual utility.

Totally six kinds of modes mentioned in this paper. It is separately Auto, Passenger, Transit, Walk, Bike and Taxi. The passenger mode is considered just inside the household interactions; Transit mode includes the bus, metro and unitcoach.

C. Individual mode choice

The basic rules of individual mode choice behavior are as shown below:

- To a large extent, mode choice is determined simultaneously by each trip purpose in the tour and mode choice decision made by other members in household;
- If a car or bike is to be used on a tour, it must be returned home at the end of the tour. So, if a car or bike is to be used on a tour, it must be used for the entire chain or some trips in the chain can use other mode, but at the end the last trip in the chain must be use the car or bike;
- Two members in a household could not use the same car at the same time at different origin and destination zones for different trip purpose;
- For non-auto or non-bike trips, each trip in the chain can choose mode independently, only relay on the maximum utility theory;
- Mode choice decision structure becomes complex, when there exist sub chains in the entire trip chain. Even if a car is to be used by first trip in a chain, it can use other modes for sub chain, but at the end the last trip in the chain must be use the car. As illustrated in Fig.(1), from home to work one can drive, and then walk to lunch, after that go to conference from lunch by transit or walk, when finish the conference one can back to work by transit or walk again, at the end have to drive to home.

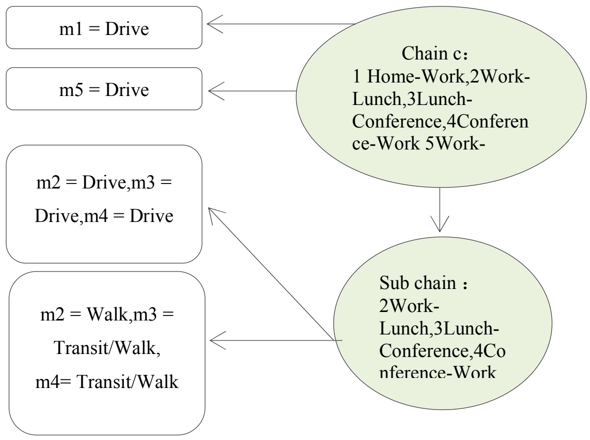


Fig. (1). Decision Structure of Mode Choice

D. Household car allocation

When the number of cars available to the household is less than the number of overlapping with car tours, all possible car allocations are evaluated and allocation decisions must be made at the household level to maximize overall household utility. For example, assume some given household have 3 members, and 2 cars, then the all possible household car allocation results are illustrated in Fig.(2).

Based on above basic principles, the utility of person n for entire tour c using mode set M, can be expressed as follows:

$$U(M,n) = \sum_{t\in T(c,n)} V(m(t),t,n) + \sum_{t\in T(c,n)} \varepsilon(m(t),t,n), \quad M \in F(c,n) \tag{3}$$

where,

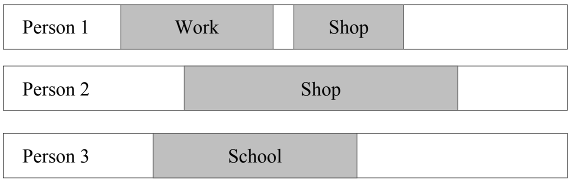
$T(c,n)$--- a set of trips on chain c for person n ;

M--- one set of specific feasible modes for the trips on chain c for person n (the chain mode set);

$F(c,n)$--- a set of chain mode sets for chain c for person n .

According to the utility maximization theory, when the chain mode set chosen is M*, we have,

$$U(M^*,n) \geq U(M,n)\ ,\ \forall\ M, M^* \in F(c,n);\ M^* \neq M \qquad (4)$$



	Person 1	Person 2	Person 3
Allo1	✓	✓	
Allo2		✓	✓
Allo3	✓		✓

Fig. (2). Household Car Allocation

IV. PARAMETER ESTIMATION

A. Log likelihood calculating

Model parameters were estimated by maximizing the log likelihood function which is shown in Equation (2).

Because of the choice decision structure and the constrains of vehicle allocation and non standard trip chain nesting structure shown in Fig. (1) & Fig. (2), no analytical expression could be found for the choice probability P. Thus P is simulated by a Monte Carlo method. Where N sets of random utilities are drawn for each trip for each person for each chain for a given β, Equation (4) is evaluated for each draw, and the frequency with which m^* is predicted to be chosen is accumulated. To account for the possibility that m^* is never chosen within the N draws, P is defined as follows:

$$P(m^*,t,n \vdots \beta) = [F(m^*,t,n \vdots \beta) + 1]/[N + n_t] \qquad (5)$$

where,

$F(m^*,t,n \vdots \beta)$--- the number of times m^* was selected for trip t out of the N draws;

n_t--- the number of feasible modes for trip t .

In order to calculate the log likelihood function, we programming with C Sharp language.

B. Niched genetic algorithm

The search for a parameter set that resulted in the maximization of the log likelihood function was done by using Niched Genetic Algorithm. The exactly NGA design are illustrated in Table **1**.

TABLE 1. NICHED GENETIC ALGORITHM DESIGN

NGA Elements	**Setting**	**Description**
Population Size	200	The number of chromosomes in the population
Initial Population	Random selection	
Niche Distance	2	The distance between parameters to consider a niche
Niche Capacity	7	The max number of population continuing in a niche
Max Mutation	0.15	The maximum amount (in 0 to 1) that a parameter can be mutated
Mutation Probability	3.5	The number of mutations per gene
Mutation Exponent	2	The exponent used for mutation
Cross Exponent	2.2	The exponent used for selecting the parameters to breed
Total Generations	300	How many generations should we process?

V. MODEL ESTIMATION RESULTS

Explanatory variables in utility function are defined as in Table **2**:

TABLE 2. DEFINITION OF EXPLANATORY VARIABLES

Explanatory Variables	**Description**
Income1	Household monthly income is less than 1500 Yuan
Income2	Household monthly income is between 1500 to 3500 Yuan
Income3	Household monthly income is between 3500 to 10000 Yuan
Income4	Household monthly income is greater than 10000 Yuan
LTTW	Linear travel time weight
Driver’s Licence	=1 if the people have driver’s licence; =0 otherwise

Youth	Age is 11 to 15
YoungAdult	Age is 16 to 19
Female	=1 if the people is female; =0 otherwise

TABLE 3. VARIANCE SCALE OF EACH MODE

Mode	Variance Scale
Auto	1
Passenger	6.443931
Walk	7
Bike	6.641128
Transit	5.644899
Taxi	1

TABLE 4. MODEL ESTIMATION RESULTS

Explanatory Variables	Coefficient
Value	-16725.11
McFadden R-square	0.286
Auto.Income1	17.86715
Auto.Income2	12.49719
Auto.Income3	10.78478
Auto.Income4	12.87068
Auto.YoungAdult	3.844018
Auto.Female	1.199191
Auto.LTTW	-0.00005
Transit.Income1	2.904111
Transit.Income2	-0.1114829
Transit.Income3	0.02398401
Transit.Income4	3.725063
Transit.Youth	6.25263
Transit.YoungAdult	7.532133
Transit.Female	4.882453
Transit.DriversLicense	6.412169
Transit. LTTW	-0.003440768
Walk.Income1	7.52139
Walk.Income2	1.148051
Walk.Income3	0.1802166
Walk.Income4	-4.15647
Walk.Youth	4.084092
Walk.YoungAdult	-2.8446
Walk.Female	6.247279
Walk.DriversLicense	2.592262
Walk. LTTW	-0.003665322
Bike.Income1	7.359191
Bike.Income3	-0.3889427
Bike.Income4	2.391879
Bike.Youth	5.736332
Bike.YoungAdult	9.494144
Bike.Female	3.628263
Bike.DriversLicense	4.498754
Bike. LTTW	-0.002138829
Taxi.Income1	-1.372013
Taxi.Income2	-5
Taxi.Income3	-3.16594
Taxi.Income4	-5
Taxi.Youth	-8.473745
Taxi.YoungAdult	-3.394375
Taxi.Female	-8.946586
Taxi.DriversLicense	-7.547933
Taxi. LTTW	-0.004079027
Passenger.Income1	-1.108307
Passenger.Income2	-3.243628
Passenger.Income3	-0.667622
Passenger.Income4	1.17974
Passenger.Youth	2.790644
Passenger.YoungAdult	1.987323
Passenger.Female	-9.038801
Passenger.DriversLicense	-10
Passenger. LTTW	-0.004261622

TABLE 5. COMPARISON OF OBSERVED AND PREDICTED VALUES

Pred\Real	Auto	Walk	Tran	Bike	Taxi	Pass	Row Total
Auto	127388	9832	6559	7429	454	2431	154093
Walk	7183	87289	64983	73891	2671	9405	245422
Tran	11609	61305	56267	57177	2658	8899	197915

Bike	7670	60422	52319	57221	2251	7501	187384
Taxi	0	1151	872	1226	66	114	3429
Pass	0	1	0	6	0	0	7
Colu Total	153850	220000	181000	196950	8100	28350	328231
Pred\Real %	Auto	Walk	Transit	Bike	Taxi	Pass	Row Total
Auto	16.16%	1.25%	0.83%	0.94%	0.06%	0.31%	19.55%
Walk	0.91%	11.07%	8.24%	9.37%	0.34%	1.19%	31.14%
Tran	1.47%	7.78%	7.14%	7.25%	0.34%	1.13%	25.11%
Bike	0.97%	7.67%	6.64%	7.26%	0.29%	0.95%	23.77%
Taxi	0%	0.15%	0.11%	0.16%	0.01%	0.01%	0.44%
Pass	0%	0%	0%	0%	0%	0%	0%
Colu Total	19.52%	27.91%	22.96%	24.99%	1.03%	3.60%	41.64%

Tran: Transit mode, Pass：Passenger mode

The following results are notable from Table 3, 4 and 5:

1) All parameters have the expected signs.

2) The mode choice model fits the data well and the McFadden R-square is equal to 0.286.

3) In Table 5, the predicted and observed percentage of each mode except the taxi and passenger mode is very close. Where, the auto mode is predicted very well, about 83% is correctly predicted. However, there existed some confusion between walk, transit and bike mode.

VI. CONCLUSIONS

Residential mode choice decision is not an independent behavior, but it is closely related with entire trip chain, household car allocation and other member's trip mode. In this paper, the car allocation in household members and trip chains are fully considered. Thus, the choice structure becomes complex, more realistic, and further the model can describe choice behavior well. The model parameters were estimated by Monte Carlo Method and Niched Genetic Algorithm based on maximum utility theory. The model results indicated that our disaggregated mode choice model fits the data well.

ACKNOWLEDGEMENTS

This work was financially supported by Beijing outstanding talent training programme "Forecasting on time and space distribution of travel demand under the affect of household activity decision" (20140000204400001); Thanks very much for the helps of Eric Miller who is professor of University of Toronto. Also thanks James Vaughan for his technology support.

REFERENCES

[1] Ortúzar, J. de D.. Nested logit models for mixed mode travel in urban corridors [J]. Transportation Research, 17A (4): 283-299, 1983

[2] Asensio, J.. Transport mode choice by commuters to Barcelona's CBD [J]. Urban Studies, 39(10): 1881–1895, 2002

[3] Bing'en Liu, Zhicai Jun. Construction of disaggregated residential mode choice model[J]. Journal of highway and transportation research and development, 25(5):116-120, 2008

[4] Hongzhi Guan, Huanhuan Yin. Study of travel mode choice behavior based on disaggregate model [J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 34(5): 1000-1003, 2010

[5] Miller, E. J.. Micro-simulation and activity based forecasting [C]. In: Activity Based Travel Forecasting Conference, June 2-5, 1996 Summary, Recommendations and Compendium of Papers, (Texas Transportation Institute, eds.), pp 151-172, Washington DC, Travel Model Improvement Program, US Department of Transportation and US Environmental Protection Agency, 1996

[6] Algers, S., Daly, A.J. and Widlert, S.. Modelling travel behaviour to support policy making in Stockholm[M]. In: Understanding Travel Behaviour in an Era of Change (P. Stopher and M. Lee-Gosselin, eds.), Pergamon, Oxford, 1997

[7] Vovsha, P., Petersen, E. and Donnelly, R.. Micro-simulation in travel demand modeling: lessons learned from the New York Best Practice Model [J]. Journal of the Transportation Research Board, 1805: 68-77, 2002

[8] Vovsha, P., Petersen, E. and Donnelly, R.. Explicit modeling of joint travel by household members: Statistical evidence and applied approach [J]. Journal of the Transportation Research Board, 1831:1-10,2003

[9] Vovsha, P. and Petersen, E.. Escorting children to school: statistical analysis and applied modelling approach [C]. Paper presented at the 84th Annual Meeting of the Transportation Research Board, Washington, DC, 2005

[10] Miller, E.J., Roorda, M.J. and Carrasco, J.A.. A tour-based model of travel mode choice[J]. Transportation, 32(4): 399-422, 2005

[11] Lerman, S. R. and C. Manski. On the use of simulated frequencies to approximate choice probabilities[M]. In: Structural Analysis of Discrete Data and Econometric Applications (C. F. Manski and D. McFadden, eds.), pp. 305-319, MIT Press, Cambridge, MA, 1981

[12] Ben-Akiva, M., Lerman, S.. Discrete Choice Analysis: Theory and Application to Travel Demand [M]. MIT Press, Cambridge, MA, 1985