

基于 Nested Logit 模型和蒙特卡罗法的通勤者活动链模拟

吴静娴¹ 杨敏¹ 陈学武¹ 吴明进²

1. 东南大学, 交通运输规划与管理江苏省重点实验室, 南京 210096

2. 上海同济城市规划设计研究院, 上海 200092

摘要:为改进传统出行模型难以解析个体活动出行全过程和各阶段前后关联性不足等问题,建立了通勤者个体活动链模拟模型。应用相关性分析等统计学方法研究了 2 434 个通勤者有效样本的活动选择相关影响因素,以此为基础建立以活动性质和发生时段为分层依据的两阶段 Nested Logit 模型,并通过蒙特卡罗法模拟通勤者的活动选择行为,结合具体活动类型选择和发生时段生成个体活动链。结果表明,两阶段 Nested Logit 模型具有较高拟合度和较好的整体性,能够刻画活动性质和发生时段对个体活动出行的双重影响规律,研究成果可为动态模拟个体活动链选择行为提供一种新的途径。

关键词: 交通运输工程活动链; Nested Logit 模型; 蒙特卡罗模拟

中图分类号: U491.1¹¹

文献标识码: A

文章编号: 1672-4747(2016)02-0076-07

DOI: 10.3969/j.issn.1672-4747.2016.02.011

Simulation of Commuter's Activity Chain Based on Nested Logit Model and Monte Carlo Method

WU Jing-xian¹ YANG Min¹ CHEN Xue-wu¹ WU Ming-jin²

1. Jiangsu Provincial Key Laboratory of Transportation Planning and Management, Southeast University, Nanjing 210096, China

2. Shanghai Tongji Urban Planning & Design Institute, Shanghai 200092, China

Abstract: In order to improve the weakness that the conventional travel model hardly analyzed the whole process of individual activity trip and the relation of every trip stage, a simulation model of commuter's activity chain was established. The statistical methods, such as correlation

收稿日期: 2015-06-17.

基金项目: 东南大学研究生创新人才培养基金(项目编号: KYLX_0170)。

作者简介: 吴静娴(1987-),女,汉族,江苏大丰人,东南大学博士研究生,主要研究方向为通勤者出行行为研究。

analysis, were applied to study all the relevant factors, which affected the commuter's activity choices of 2434 valid samples. Then a two-stage nested logit model based on the activity starting and lasting periods. Furthermore, combined with the specific types' selection and time sequences of the activities, the commuter's activity choices were simulated with Monte Carlo method. The results indicate that the two-stage nested logit model has higher fitting degree and better integrity, and can better characterize the dual influences of the time and property of individual's trip activity. The research can provide a new approach for dynamic simulation of individual's activity chain.

Key words : Transportation engineering; activity chain nested logit model; montecarlo simulation

0 引言

交通需求源于人们参与社会经济活动的需要,个体活动链的研究有助于对个体完整出行活动模式和出行中发生的各种活动前后关联性的跟踪,从而克服传统出行模型以单次出行为分析单元的不足,强化对出行生成机理的揭示。以 Ben-Akiva、Bowman、Bhat、Miller 等为代表的西方学者用以非集计模型为核心的计量经济学方法建立了 Boston、Portland、CEMDAP、TASHA 等代表性活动链模型^[1-3], Kuppam 运用结构方程式模型研究了通勤者的个人属性、出行生成与活动链之间的相关关系^[4]。国内学者隗志才、宗芳将活动模式定义为一阶、二阶、中间驻停、工作往返四类,并运用 logistic 回归方法进行模型的标定,杨敏等运用相关性分析和多项 logit 方法建立了基于活动模式的工作者出行生成预测模型,杨扬等基于支持向量机理论对出行链活动类型的识别方法进行了研究^[5-7]。近年来, Arentze 和 Timmermans、Charypar 和 Nagel 等运用人工智能方法研究了活动-出行决策过程中的学习机制,按智能体决策规则对活动链进行了仿真研究^[8,9]。然而,通勤者由于受到工作刚性活动的约束,活动链决策过程中既有稳定因素,又存在动态性。在刚性活动框架下对弹性活动类型和活动发生的时间段进行模拟,可为获取通勤者活动链提供一种可行的方法。因此,本文将构建通勤者活动选择的 2 层

Nested Logit 模型,并采用蒙特卡罗模拟活动选择,模拟通勤者个体活动链。

1 通勤者活动链模拟思路

个体活动链可从参与活动的类型和活动发生的时段来划分,本文将以此着手模拟通勤者活动链,分析刚性活动约束下个体弹性活动选择和发生时序决策行为。具体思路和方法如图 1 所示。

首先,对个体活动进行分类,建立通勤者基础数据库。其次,运用 STATA 软件定量分析通勤者家庭属性、个体属性和出行属性与个体活动选择的相关性,确定模型分层变量。然后,以活动类型和时间发生阶段为分层依据,构建通勤者活动选择的 2 层 Nested Logit 模型,定量分析个体三类属性对其活动选择的影响,并检验模型。最后,根据模型标定结果预测个体活动选择概率,结合蒙特卡罗法生成通勤者个体活动链。

2 通勤者活动选择的两阶段 Nested Logit 模型

2.1 两阶段活动类型细分

本文以 2009 年苏州市居民出行调查为数据基础,提取 2 434 个通勤者有效样本进行分析。通勤者的活动主要以工作为主,工作之外的弹性活动在工作

活动时间和空间的约束下形成多种活动链。因此，以 和重组，得到活动类型如表 1 所示。
 活动类型和发生时间阶段为参考点，将活动进行分类

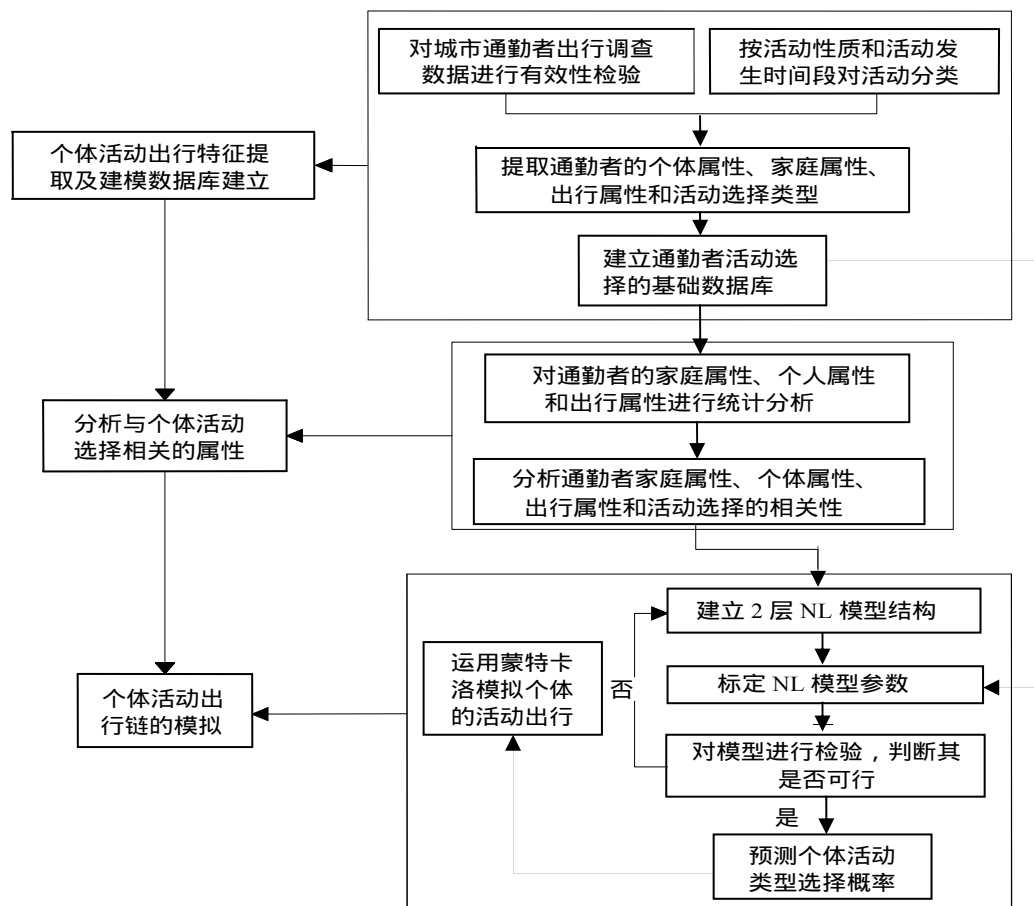


图 1 通勤者活动链模拟流程

Fig.1 Simulation flow chart of commuter's activity chain

表 1 两阶段活动类型细分

Tab.1 Two stage activity classifications

以活动发生时间阶段分类		以活动性质分类	
时间阶段 细分	含义	活动性质 细分	含义
1	$B_{eforework}$ 工作前活动	1	$E_{ntertainment}$ 娱乐活动
2	A_{twork} 工作中活动	2	$S_{hopping}$ 购物活动
3	$A_{fterwork}$ 工作后活动	3	$R_{eturnhome}$ 临时回家活动
注:4 活动性质×3 时间阶段=12 种活动类型			
		4	P_{ickup} 接送他人活动

统计分析发现通勤者工作日参与活动主要为工作，参与其他弹性活动频率很低。剔除频率较小的活动模式，得到以下四种主要的活动组合，见表 2。

表 2 通勤者复杂活动链类型

Tab.2 Types of commuter's complex activity chain

活动链 类型	含义	活动链描述
B_{EP}	工作前接送人	$H_{ome} \rightarrow P_{ickup} \rightarrow W_{ork} \rightarrow H_{ome}$
A_{TR}	工作后临时回家	$H_{ome} \rightarrow W_{ork} \rightarrow R_{eturnhome} \rightarrow W_{ork} \rightarrow H_{ome}$
A_{FP}	工作后接送他人	$H_{ome} \rightarrow W_{ork} \rightarrow P_{ickup} \rightarrow H_{ome}$
A_{FS}	工作后购物	$H_{ome} \rightarrow W_{ork} \rightarrow S_{hopping} \rightarrow H_{ome}$

注： $B_{EP}=B_{eforework}+P_{ickup}$ ； $A_{TR}=A_{twork}+R_{eturnhome}$ ；

$A_{FP}=A_{fterwork}+P_{ickup}$ ； $A_{FS}=A_{fterwork}+S_{hopping}$ 。

2.2 两阶段 Nested Logit 模型建立

Nested Logit 模型结合分层的思想，将个体出行活动选择划分为两层，由上而下分别命名为水平 2 和

水平 1，以此构建个人活动出行的模型决策树。同时，模型根据个人效用最大化理论，将居民个人属性、家庭属性及出行属性的多种变量纳入各层模型中。根据随机效应理论，第 n 个出行者选择方案 (rm) 时的效用 U_{rmn} 可以表示为^[10]：

$$U_{rmn} = V_{(r|m)n} + V_{mn} + \varepsilon_{(r|m)n} + \varepsilon_{mn} \quad (1)$$

式中， V_{mn} ——第 n 个个体选择水平 2 选项 m 的效用；
 $V_{(r|m)n}$ ——在选择 m 的条件下选择水平 1 选项 r 的效用。

当固定项 V_{mn} 和 $V_{(r|m)n}$ 分别与其中的解释变量呈线性关系时，表示为：

$$V_{(r|m)n} = \beta' X_{(r|m)n} = \sum_{k=1}^{K_1} \beta_k X_{(r|m)nk} \quad (2)$$

$$V_{mn} = \theta' X_{mn} = \sum_{k=1}^{K_2} \theta_k X_{mnk} \quad (3)$$

式中， K_1 、 K_2 ——解释变量的个数；

β_k 、 θ_k ——第 k 个变量所对应的待估计参数；

$X_{(r|m)nk}$ 、 X_{mnk} ——对应的第 k 个特性变量。

随机项 ε_{mn} 和 $\varepsilon_{(r|m)n}$ 服从二重指数分布，有：

$$P_n(rm) = P_n(r|m)P_n(m) = \frac{\exp(\lambda_1 V_{(r|m)n}) * \exp[\lambda_2 (V_{mn} + V_{mn}^*)]}{\sum_{r'=1}^{R_{mn}} \exp(\lambda_1 V_{(r'|m)n}) * \sum_{m'=1}^{M_n} \exp[\lambda_2 (V_{m'n} + V_{m'n}^*)]} \quad (4)$$

式中， $P_{n(m)}$ ——第 n 个个体选择水平 2 选项 m 的概率；

$P_{n(r|m)}$ ——在选择 m 的条件下选择水平 1 选项 r 的条件概率。

然后，由极大似然估计法及分阶段估计法，先后估计水平 1 模型中参数 β 和水平 2 模型中参数 λ_2 和 θ 。以通勤者个体参与活动的时段选择作为水平 2 模型，活动类型选择作为水平 1 模型，构建居民活动时间-性质选择相互影响的 Nested Logit 模型，如图 2 所示。

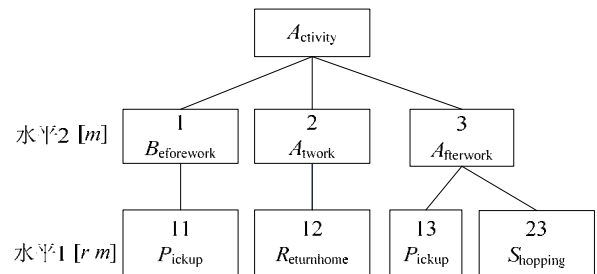


图 2 两阶段 Nested Logit 决策树

Fig.2 Two stages of the nested logit decision tree

2.3 模型的标定与检验

利用相关性分析和交叉列联表法筛选模型的特征变量，得到如表 3 所示的解释变量。

表 3 模型的解释变量及其取值

Tab.3 Explanatory variables and their values

属性类别	编码规则	含义
家庭属性	家庭结构 f_s	$f_s=1$ ：家中无老人； $f_s=2$ ：家中有老人
	家庭人口数 p_{opu}	$p_{opu}=1$ ：家中有 3 人； $p_{opu}=2$ ：家中超过 3 人
	家庭工作人数 w_{orkers}	$w_{orkers}=i$ ：家中共有 i 个人上班
	家庭学龄前儿童 $p_{rechild}$	$p_{rechild}=0$ ：家中无学龄前儿童； $p_{rechild}=1$ ：家中有学龄前儿童
	家庭电动车数 b_{ikes}	家庭拥有电动车的数量值
个人属性	家庭小汽车数 c_{ars}	家庭拥有小汽车的数量值
	家庭年收入 i_{ncome}	$i_{ncome}=1$ ：低于 2 万； $i_{ncome}=2$ ：(2~4) 万； $i_{ncome}=3$ ：(4~6) 万； $i_{ncome}=4$ ：(6~8) 万； $i_{ncome}=5$ ：8 万以上
	性别 s_{ex}	$s_{ex}=0$ ：女； $s_{ex}=1$ ：男
个人属性	职业 c_{areer}	$c_{areer}=1$ ：学生； $c_{areer}=2$ ：商业/服务人员； $c_{areer}=3$ ：高级行政/管理人员； $c_{areer}=4$ ：企业一般职员； $c_{areer}=5$ ：事业单位及公务员； $c_{areer}=6$ ：个体经营者；公司职员； $c_{areer}=5$ ：私营及个体劳动者
	有无驾照 l_{icense}	$l_{icense}=0$ ：无驾照； $l_{icense}=1$ ：有驾照
个人属性	年龄 a_{ge}	$a_{ge}=1$ ：30~39； $a_{ge}=2$ ：40~59

续表 3

个人	有无 IC 卡 i_c	$i_c=0$: 无 IC 卡; $i_c=1$: 有 IC 卡
属性	受教育程度 $e_{educate}$	$e_{educate}=1$: 初中及以下; $e_{educate}=2$: 高中及中专; $e_{educate}=3$: 大专本科及以上学历
属生类别	编码规则	含义
	是否为高峰时间出行 p_{peak}	$p_{peak}=0$: 非高峰时间出行; $p_{peak}=1$: 高峰时间出行
出行属性	出行方式 t_{way}	$t_{way}=1$: 非机动车化出行; $t_{way}=2$ 公共方式出行; $t_{way}=3$: 小汽车出行
	出行时耗 t_{time}	因前去参与某活动而产生的出行时耗
	活动时耗 a_{time}	所参与活动的活动时耗

为保证模型的渐进性、正态性以及有效性, 本文进行标定。然后标定水平 2 的模型, 以 $B_{eforework}$ 作为采用分阶段估计法进行模型参数标定。水平 1 以 A_{FS} 参照类, 对由 $B_{eforework}$ 、 A_{twork} 和 $A_{fterwork}$ 组成的三项作为参照类, 对以 A_{FS} 和 A_{FP} 为选项的两项 logit 模型 logit 模型进行标定, 详见表 4 和表 5。

表 4 水平 1 模型标定结果

Tab.4 Calibrating results of the model at layer 1

$A_{ctivity}$	属性	系数	标准差	z	$P>z$	95%置信区间		
A_{FP}	p_{peak}	-0.742	1.517	-0.490	0.225	-3.714	2.231	
	t_{time}	-0.008	0.030	-0.003	0.184	-0.067	0.050	
	f_s	2.725	1.207	2.260	0.024	0.360	5.091	
	b_{ikes}	-17.819	2 588.473	-0.010	0.195	-0.509	0.506	
	t_{way}	2	4.888	1.391	3.510	0.000	2.161	7.615
		3	0.910	1.227	0.740	0.158	-1.494	3.314
		3	0.161	1.580	0.100	0.219	-2.935	3.257
	Career	4	0.406	1.241	0.330	0.244	-2.027	2.839
		5	-3.260	2.253	-1.450	0.148	-7.677	1.156
	$_{-}c_{ons}$	15.590	2 588.473	0.010	0.095	-5 057.730	5 088.905	
A_{FS}	(参照类)							

注: $_{-}c_{ons}$ 为选择工作后活动的条件下选择 A_{FP} 的效用函数中的常数项。

在该模型中, p_{peak} 以 $p_{peak}=0$ 作为参照类; f_s 以 $f_s=1$ 作为参照类; b_{ikes} 以 $b_{ikes}=0$ 作为参照类;

t_{way} 以 $t_{way}=1$ 作为参照类; c_{career} 以 $c_{career}=2$ 作为参照类。

表 5 水平 2 模型标定结果

Tab.5 Calibrating results of the model at layer 2

$A_{ctivity}$	属性	系数	标准差	z	$P>z$	95% 置信区间		
$B_{eforework}$	(base outcome)							
A_{twork}	c_{ars}	-1.058	1.192	-0.890	0.375	-3.393	1.278	
	s_{ex}	0.689	0.924	0.750	0.456	-1.121	2.499	
	i_c	-0.868	1.011	-0.860	0.390	-2.849	1.113	
	p_{peak}	-5.467	1.690	-3.240	0.001	-8.778	-2.155	
	t_{way}	2	3.314	1.483	2.240	0.025	0.408	6.220
		3	-2.067	1.461	-1.410	0.157	-4.932	0.797
	t_{time}	0.013	0.038	0.340	0.731	-0.062	0.088	
	a_{time}	0.075	0.016	4.780	0.000	0.044	0.106	
	$_{-}c_{ons}$	-1.666	0.985	-1.690	0.091	-3.596	0.264	
	$A_{fterwork}$	c_{ars}	-0.221	0.643	-0.340	0.731	-1.482	1.040
s_{ex}		-1.245	0.514	-2.420	0.015	-2.252	-0.238	
i_c		0.858	0.610	1.410	0.159	-0.337	2.053	
p_{peak}		0.687	0.737	0.930	0.351	-0.757	2.132	
t_{way}		2	2.973	0.860	3.460	0.001	1.287	4.658
		3	-0.392	0.676	-0.580	0.562	-1.718	0.934
t_{time}		0.095	0.027	3.500	0.000	0.042	0.148	
a_{time}		0.063	0.014	4.420	0.000	0.035	0.091	
l_{ogsum}		-0.190	0.086	-2.210	0.027	-0.359	-0.021	
$_{-}c_{ons}$		-3.622	0.946	-3.830	0.000	-5.476	-1.768	

注: $_{-}c_{ons}$ 为常数项。各分类变量均以其最小值作为参照类。 l_{ogsum} 为水平 1 的效用对水平 2 的影响值。

由上可知，各水平模型的参数检验结果基本显著。为确保模型的整体性，先后对水平 1 及水平 2 的多项 logit 模型进行整体卡方检验，结果分别为 LR $\chi^2(9)=48.51$ 和 LR $\chi^2(18) = 242.68$ ，其各自对应的概率为 $\text{Prob} > \chi^2 = 0.0000$ 。根据统计学理论可知，该模型模拟的结果较好。而模型预测的准确性检验采用的是类 R^2 指标，类 R^2 指标值域为 0 至 1，其值越接近于 1 越好，表示自变量可以很好解释因变量。水平 1 和水平 2 模型 R^2 分别为 0.5482 和 0.5849，说明模型具有较好准确性。综上可知，两个水平的模型拟合效果较好，适于个体活动选择的预测。

3 基于蒙特卡罗法的通勤者个体活动链模拟

通勤者活动选择两阶段 Nest Logit 模型的建立，仅是将个体活动选择行为转化为概率数学模型。而个体对活动具体选择结果，则需借助计算机模拟方法实现。蒙特卡罗(Monte Carlo)模拟方法是基于概率模型的计算方法，它可以模拟各选择枝概率已知情况下的选择结果。个体活动的蒙特卡罗模拟过程具体如下：

(1) 计算个体的累积概率：

$$P_{in}^{(0)} = 0, P_{in}^{(l)} = \sum_{i=1}^l P_{in} \quad (l=1,2,3,4) \quad (5)$$

(2) 生成区间[0,1)的随机数 R_n 。

(3) 根据 R_n 值，确定个体的活动选择。记蒙特卡罗模拟的个体活动选择为 ActM。若 R_n 满足下式，

$$P_{in}^{(l-1)} < R_n \leq P_{in}^{(l)} \quad (l=1,2,3,4) \quad (6)$$

则令 ActM=l，即认为个体选择的活动为 l。

从苏州市通勤数据提取 4 个样本，信息如表 6 所示。

由 NL 模型估算个体选择活动的概率，见表 7。

根据随机生成一组随机数，结合表 7 得到各自个体活动选择结果，详见表 8。

表 6 4 个个体、家庭及出行基本信息

Tab.6 Individual information of the 4 samples

个体编号	f_s	b_{ikes}	c_{ars}	s_{ex}	c_{areer}	i_c	a_{time}	t_{time}	t_{way}	P_{eak}
1	2	1	1	0	5	0	120	30	3	1
2	1	1	0	1	4	1	10	50	2	1
3	1	1	0	0	4	1	105	15	1	0
4	2	1	1	1	4	0	5	25	3	1

表 7 4 个个体活动选择累积概率

Tab.7 Cumulative probability of the 4 samples activity choices

个体编号	$P_1=B_{EP}$	$P_2=P_1+A_{TR}$	$P_3=P_2+A_{FP}$	$P_4=P_3+A_{FS}$
1	0.000 624	0.000 806	0.055 886	1
2	0.009 532	0.010 474	0.872 711	1
3	0.002 195	0.577 383	0.630 197	1
4	0.892 247	0.892 374	0.968 119	1

表 8 4 个个体活动类型选择结果

Tab.8 Activity choices results of the 4 samples

个体编号	1	2	3	4
随机数	0.165 021	0.760 604	0.371 38	0.379 541
活动类型	A_{FS}	A_{FP}	A_{TR}	B_{EP}

最后，由表 2 对各活动链类型的描述，得到蒙特卡罗模拟的 4 个个体活动链。

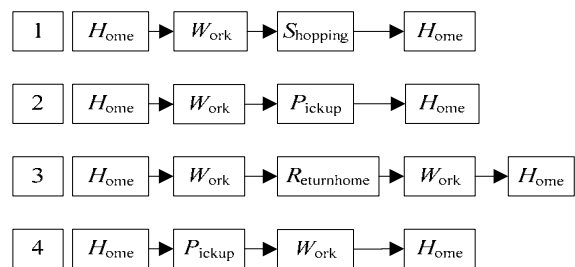


图 3 个体活动出行链模拟结果

Fig.3 Simulation results of individual's activity chain

4 结 论

本文首先提出了根据活动发生时间阶段和活动性质对个体活动进行分类的具体方法，以此为基础建立了通勤者活动选择的两阶段 Nested Logit 模型，并

借助蒙特卡罗模拟方法研究了通勤者在刚性活动约束下对弹性活动类型选择及时序安排的决策行为,综合考虑了活动选择的全过程和动态性。本文的研究成果进一步揭示了居民活动出行的内在规律和决策行为。本研究旨在通过两阶段模型构建描述个体活动决策行为过程,探讨动态模拟个体活动链的新途径。由

于建模所用数据所限,模型考虑的影响因素尚不够全面,预测精度仍有待提高。今后可以通过精细的调查设计和样本挑选,深入活动链模拟的实际大范围应用研究。此外,在以后的研究中可以针对 Nested Logit 模型进行参数的敏感性分析,进而评估交通需求管理政策的实施效果。

参考文献

[1] Bowman J.L. and Ben-Akiva M. Activity-based disaggregate travel demand model system with activity schedules [J]. Transportation Research Part A, 2000, 35(1): 1-28.

[2] Bhat C.R. and Singh S.K. A comprehensive daily activity-travel generation model system for workers [J]. Transportation Research Part A, 2000, 34(1): 1-22.

[3] Roorda M., Miller E.J. and Habib K.M.N. Validation of TASHA: A 24-h activity scheduling microsimulation model [J]. Transportation Research Part A, 2008, 42: 360-375.

[4] Kuppam A.R. and Pendyala R.M. A Structural equations analysis of commuters' activity and travel patterns [J]. Transportation, 2001, 28: 33-54.

[5] 宗芳, 隗志才, 高研等. 拥挤收费政策交通影响评价 [J]. 吉林大学学报(工学版), 2009, 39(2): 343-348.

[6] 杨敏, 陈学武, 王伟等. 基于活动模式的工作者出行生成预测模型[J]. 东南大学学报(自然科学版), 2008, 38(3): 525-530.

[7] 杨扬, 姚恩建, 岳昊等. 基于支持向量机的出行链活动类型识别研究[J]. 交通运输系统工程与信息, 2010, 10(6): 70-75.

[8] Arentze T., Pelizaro C. and Timmermans H. An agent-based micro-simulation framework for modeling of dynamic activity-travel rescheduling decisions [J]. International Journal of Geographical Information Science, 2010, 24(8): 1149-1170.

[9] Charypar D. and Nagel K. Q-learning for flexible learning of daily activity plans [J]. Transportation Research Record, 2005, 1935: 163-169.

[10] 关宏志. 非集计模型——交通行为分析的工具[M]. 北京: 人民交通出版社, 2004.

[11] L'Ecuyer P. Quasi-Monte Carlo methods with applications in finance[J]. Finance and Stochastics, 2009, 13(3): 307-349.

(中文编辑: 刘娉婷)