# Travel-Activity Choice Set Generation within the Discrete-Continuous Extreme Value Models using Probe Person data

2012.10.26 Travel behavior and network model courses

**Sachiyo Fukuyama,**
PhD Student, Dept.of Urban Engineering,
The University of Tokyo
fukuyama@bin.t.u-tokyo.ac.jp

# Outline

1. Research background
2. Purpose of the study
3. Frame of the study
4. Data
   1) Probe Person data
   2) Basic analysis
   3) Virtual network
5. Choice set generation
6. Choice model
7. Estimation result
8. Conclusion

## Research background

Effective approaches for reactivating the city central areas have been required and attempts to renovate the places to be attractive are getting much attention.

But the problem is whether people will go there or not.

For the places being used frequently and continuously, considering the relation between the places and daily life activity patterns is necessary.

# Research background

## Activity-based models

The demand for travel is derived from the demand for activity

**Discrete choice models**

Choice behavior is generated based on utility-maximization

Bowman and Ben-Akiva(2000)
PETRA(Fosgerau,2001)
METRO(Bradley et al.,1998) etc.

The size of choice set become enormous

**Rule-based models**

Choice behavior is generated based on huristic rules

ALBATROSS
(Arentze and Timmermans, 2005)
TASHA(Roorda et al., 2008) etc.

Not easy to understand the characteristics of parameters

# Purpose of the study

1) To propose efficient and realistic way for the choice set generation in location choice situations.

2) To find common patterns of the spatial distribution of the location choice as domains of daily activities, which are assumed to be formed by acceccibility and strengthened by familiarity.

3) To find the characteristics of the places where people stay longer by estimating the value of places by duration of stay as well as frequency.

# Frame of the study

**Trajectory based approach**

Applying **MDCEV**(Multiple discrete-continuous extreme value) model (Bhat 2005),
to handle the choice of multiple alternatives simultanelusly
and allocation of time-budget

**Generating the choice set**
in trajectory-based and data-oriented

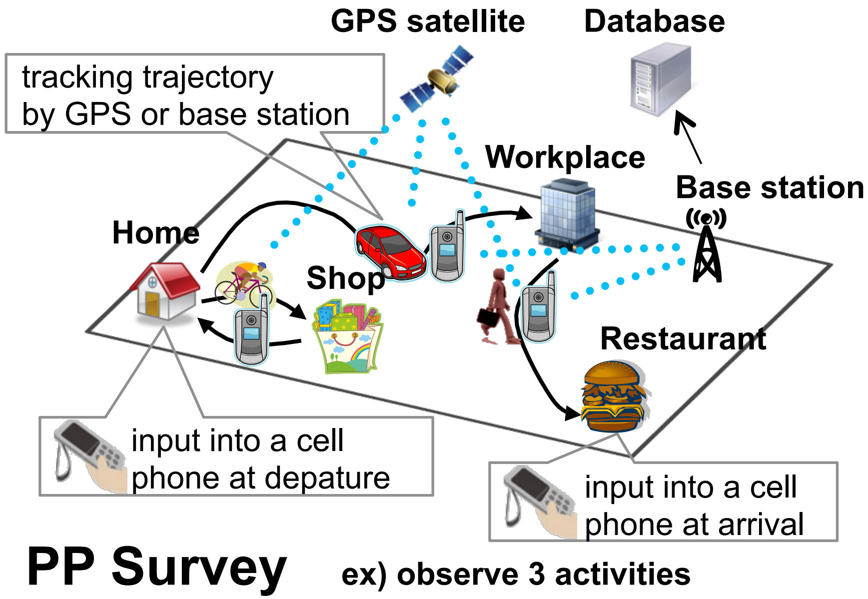
Using **Probe Person data,**
individuals' precise position and time data for long term

Using the **Virtual Network**
for analyzing GPS data without road network data

## Probe Person survey

is the method for tracking individual travel behavior in urban space by using an automatic position and time recording system based on GPS and internet communications

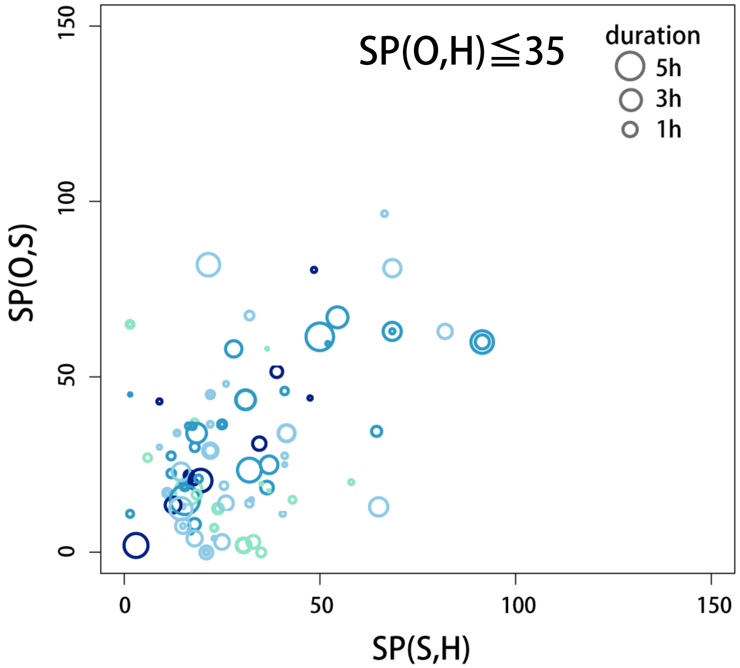
→ Provides us Individuals' precise position and time data
→ Provides us long term data observed for same respondents

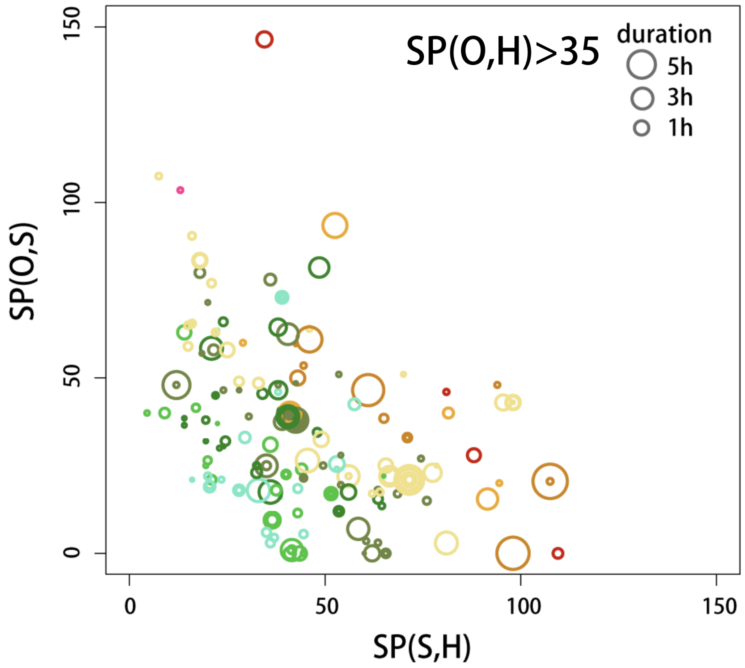



| | |
|---|---|
| Surveillance period | from 2 to 4 weeks, 10 terms in 2007/11/12-2008/1/27 |
| Area | Matsuyama urban area |
| The number of respondents* | 109 |
| The number of tours* | 276 |

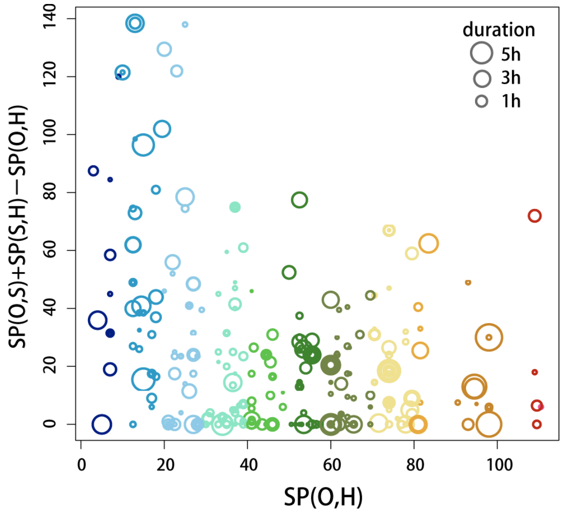
* After data cleaning and extracting tours by car

# Basic analysis




SP: Total link cost of the shortest route
O: Office
H: Home
S: Stopping place



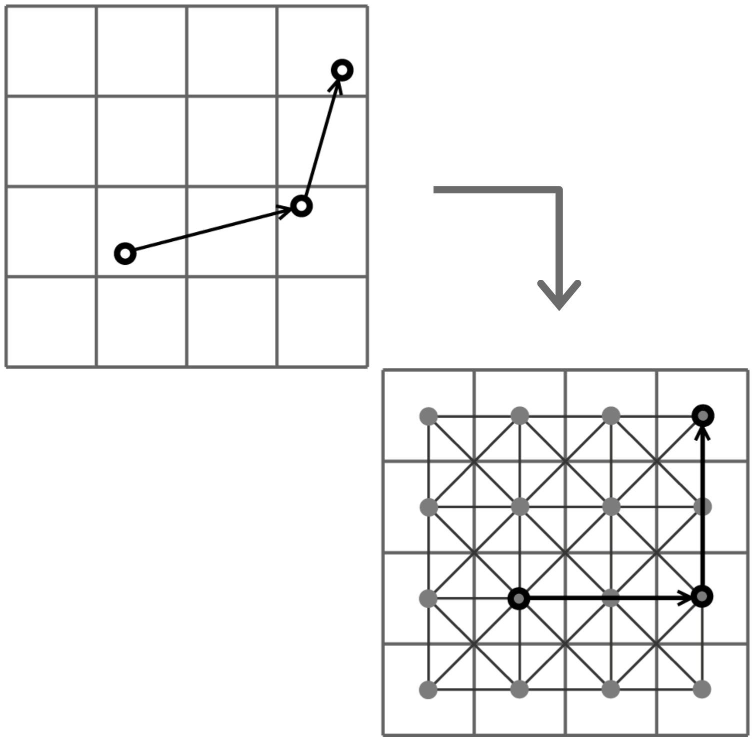

Stopping places distribute around the shortest routes between the office and home in the case of SP(O,H)>35.

The differences between SP(O,S)+SP(S,H) and SP(O,H) get shorter as SP(O,H) longer.

# Virtual Network

## Using the **Virtual Network** for analyzing GPS data without road network data

→ Avoiding the bias caused by map matching
→ Avoiding the computational burden of map matching
→ Saving the cost of development of the road network data



① Divide the area into mesh cells as nodes and set virtual links between adjacent nodes

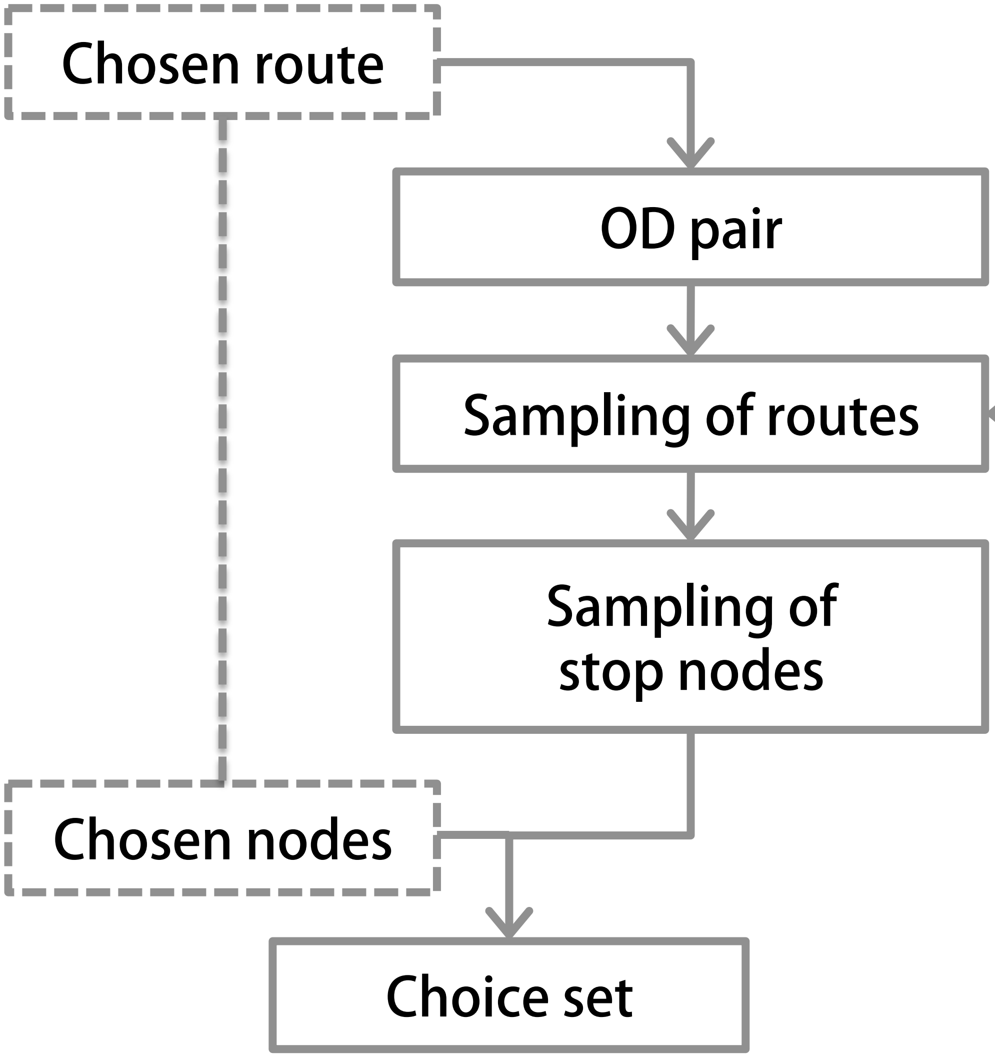
② Relate each GPS point with nodes

③ Compute each link cost by counting the points passing the link

# Choice set generation

## Sampling alternatives of stopping nodes in trajectory-based and data-oriented

$s_d$ : destination
link $l = (v,w) \in E_v$
$E_v$ : set of outgoing links from $v$
$C(l)$ : generalized cost of link $l$
$SP(v_1,v_2)$ : generalized cost of the shortest path between nodes $v_1$ and $v_2$
$b_1,b_2$ : shape parameters




**a biased random walk algorithm (Frejinger et al. 2009)**

**weight**

$$\omega(l|b_1,b_2) = 1-(1-x_l^{b_1})^{b_2}$$

$$x_l = \frac{SP(v,s_d)}{C(l)+SP(w,s_d)}$$

**probability**

$$q(l|E_v,b_1,b_2) = \frac{\omega(l|b_1,b_2)}{\sum_{m\in E_v}\omega(m|b_1,b_2)}$$

## 4 patterns of sampling:

1. Sampling of routes based on the probability distribution by the distance to the shortest path and,

   -a: Sampling of stop nodes based on the distribution of the stop nodes in each individual's real data to consider the difference between individuals

   -b: Samling of stop nodes based on the distribution of the stop nodes in the real data of all members to consider the tendency to ovwrall distribution

   -c: Sampling of stop nodes randomly to see the effects of a and b

2. Sampling of stop nodes randomly

# Choice model

**MDCEV** ( Multiple discrete-continuous extreme value model )

Bhat ( 2005, 2008 )

## Random utility function:

$$U(t) = \sum_k \frac{\gamma_k}{\alpha_k}\left[\exp(\beta' z_k + \varepsilon_k)\right] \bullet \left\{\left(\frac{t_k}{\gamma_k} + 1\right)^{\alpha_k} - 1\right\}$$

$\alpha_k$ : satiation parameter
$\gamma_k$ : satiation and translation parameter
$z_k$ : explanatory variables
$\beta$ : parameter
$\varepsilon_k$ : error term
$t_k$ : time spent in activity purpose $k$

In this research, variables are
1) distance from home to the stop node
2) distance from office to the stop node
3) average of duration and frequency of visit at each node
4) difference between total link cost of the route and the shortest path

# Estimation results

| | Segment1 (SP(O,H)≦35) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Choice set: 1-a | | Choice set: 1-b | | Choice set: 1-c | | Choice set: 2 | |
| | Estimate | *t*-stat | Estimate | *t*-stat | Estimate | *t*-stat | Estimate | *t*-stat |
| Parameters | | | | | | | | |
| SP(S,H)≦50 | -0.987 | -104.701 | -0.899 | -93.087 | -0.795 | -82.723 | -0.550 | -50.815 |
| SP(O,S)≦50 | 0.648 | 51.130 | 0.516 | 42.661 | 0.380 | 31.893 | 0.525 | 40.513 |
| SP(O,S)+SP(S,H)-SP(O,H) | 0.0006 | 0.054 | 0.003 | 0.307 | 0.006 | 0.560 | -0.029 | -2.430 |
| Cumulative sojourn time on each node | -0.013 | -0.501 | -0.060 | -1.855 | -0.055 | -1.482 | 0.048 | 1.255 |
| Constant | -0.203 ～-0.191 | -0.453 ～-0.039 | -0.204 ～-0.194 | -0.428 ～-0.054 | -0.203 ～-0.195 | -0.457 ～-0.059 | -0.203 ～-0.189 | -0.330 ～-0.070 |
| γ | 0.310 | 1.436 | 0.310 | 1.437 | 0.308 | 1.419 | 0.310 | 1.424 |
| Final log-likelihood value | -413.149 | | -405.352 | | -399.326 | | -351.89 | |
| Adj. rho bar sq. | 0.265 | | 0.276 | | 0.275 | | 0.381 | |

# Estimation results

| | Segment2 (SP(O,H)>35) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Choice set: 1-a | | Choice set: 1-b | | Choice set: 1-c | | Choice set: 2 | |
| | Estimate | *t*-stat | Estimate | *t*-stat | Estimate | *t*-stat | Estimate | *t*-stat |
| Parameters | | | | | | | | |
| SP(S,H)≦50 | -0.489 | -129.559 | -0.517 | -148.723 | -0.459 | -123.822 | 0.652 | 152.573 |
| SP(O,S)≦50 | 0.335 | 83.534 | 0.103 | 27.689 | 0.143 | 38.267 | -0.243 | -65.639 |
| SP(O,S)+SP(S,H)-SP(O,H) | -0.035 | -5.252 | -0.027 | -4.408 | -0.026 | -4.189 | -0.050 | -7.596 |
| Cumulative sojourn time on each node | 0.010 | 0.225 | -0.043 | -0.956 | -0.026 | -0.539 | 0.034 | 0.680 |
| Constant | -0.202 ~ -0.195 | -0.522 ~ -0.018 | -0.201 ~ -0.197 | -0.514 ~ -0.080 | -0.201 ~ -0.196 | -0.509 ~ -0.104 | -0.206 ~ -0.175 | -0.475 ~ -0.083 |
| γ | 0.303 | 1.795 | 0.302 | 1.794 | 0.302 | 1.793 | 0.302 | 1.745 |
| Final log-likelihood value | -714.01 | | -715.263 | | -711.045 | | -555.285 | |
| Adj. rho bar sq. | 0.279 | | 0.278 | | 0.282 | | 0.429 | |

3 cases considering route lengths (case1-a,b,c) result worse than the case generated choice set randomly(case2).
This may be the effect of moderate dispersion by random sampling.
But in case 2 the parameters' signs are not stable when computing with the number of alternatives varied.

# Couclusion

1) Proposing a trajectory-based and data-oriented approach for choice set generation in location choice situations.

2) Proposing the location choice model with considering the tendency of the spatial distribution of daily activities and bringing duration of stay in the value estimation of of the places by applying MDCEV model.

3) As a result, though the accuracy of the model was still low, we got the stable and logically appropriate model for future study.