Travel Behaviour and Society 11 (2018) 93–100

Contents lists available at ScienceDirect

Travel Behaviour and Society

journal homepage: www.elsevier.com/locate/tbs

Structured random walk parameter for heterogeneity in trip distance on modeling pedestrian route choice behavior at downtown area

Toshiyuki Yamamoto[a,*], Shinichi Takamura[b], Takayuki Morikawa[c]

[a] Institute of Materials and Systems for Sustainability, Nagoya University, C1-3(651) Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
[b] Pacific Consultants Co. Ltd., 2-3-13 Azuchimachi, Chuo-ku, Osaka 541-0052, Japan
[c] Institute of Innovation for Future Society, Nagoya University, C1-3(651) Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

ARTICLE INFO

Keywords:
Expanded path size
Path size logit model
Sampling of alternatives
Expanded logsum
Heteroscedasticity
GPS

ABSTRACT

A nested logit model of mode and walk route choice behavior in a downtown area is developed. The model samples alternatives at the route choice level, which is a lower level of the nesting structure. Alternative routes are drawn using the random walk method. A structured random walk parameter based on trip distance is proposed in order to consider the heterogeneity in trip distance, and a structured scale parameter based on trip distance is also applied for the nested logit model. Dataset is obtained by using mobile phones with GPS functions to track the trajectories in a downtown area at Nagoya, Japan. The results suggest that the proposed structured random walk parameter provides more efficient parameter estimates. Also, heteroscedasticity in the error terms for trips of various distances is found to be significant. Empirical findings include that older pedestrians prefer main shopping streets, and that roadside restaurants affect the route choice differently on weekdays and holidays. The different effects of the roadside environments according to the pedestrian's attributes and the day of the week may indicate a need for more appropriate policies for specific pedestrian categories and days of the week.

1. Introduction

Pedestrian behavior in downtown areas has been investigated for decades in order to forecast the total demand for retail facilities or the effects of changes in the transportation network on the turnover in the shopping streets. Borgers and Timmermans (1986) developed a model system that consists of three sub-models: one for destination choice, one for route choice, and one for impulse stops. In their model, a conventional multinomial logit model was applied using only a single independent variable, distance. More recently, some studies have applied continuous space representations for pedestrian behavior from more microscopic perspectives (e.g., Hoogendoorn and Bovy 2004; Antonini et al. 2006; Asano et al. 2009). This approach has the advantage of representing pedestrian movements at transfer stations or shopping malls, where routes cannot be defined clearly. However, several studies suggest using the continuous space model with a network-based route choice model. Papadimitriou et al. (2009) regards route choice as tactical, and road crossings, obstacle avoidance, and interaction with other pedestrians as operational; the latter is well represented by continuous space models. Antonini et al. (2006) also regard the continuous space model to be used with network-based route choice models, where the destination and route at the network level are treated as given when implementing the continuous space model. Here, a critical problem with network-based route choice models, especially in downtown areas, is the formation of the choice set. Several empirical analyses have been carried out on pedestrian route choice behavior (e.g., Muraleetharan and Hagiwara, 2007; Borst et al., 2009; Guo, 2009; Guo and Loo, 2013), but the formation of choice set is a difficult task for the analysis. The number of physically possible alternative routes becomes extremely large in dense networks such as downtown areas, so the choice set for the network-based route choice models is difficult to define. In this study, the network-based walk-route choice model in a downtown area is developed by explicitly considering the difficulty in formulating the route choice set and combining it with the mode choice as a nested logit model. The nested logit model of mode and route choice enables rigorous examination of the effects of the walking environment on route and mode choice in a downtown area.

Not specific to route choice models, choice sets are important and difficult to formulate, as is widely known. Many studies have been done towards developing operational models with two-stage characterization: choice set generation and choice among them (e.g., Manski, 1977; Horowitz, 1991). At the same time, the integrated choice model, in

* Corresponding author.
E-mail addresses: yamamoto@civil.nagoya-u.ac.jp (T. Yamamoto), shinichi.takamura@os.pacific.co.jp (S. Takamura), morikawa@nagoya-u.jp (T. Morikawa).

https://doi.org/10.1016/j.tbs.2018.02.006
Received 8 June 2017; Received in revised form 24 January 2018; Accepted 10 February 2018
2214-367X/

which choice set generation is embedded, also has been proposed (e.g., Fotheringham, 1983; Horowitz and Louviere, 1995; Sinha, 2000; Swait, 2001). Both approaches have advantages and disadvantages, and the most suitable approach depends on the empirical context. In route choice analyses, several algorithms for generating the choice set have been proposed and evaluated (e.g., Bekhor et al., 2006; Bovy and Fiorenzo-Catalano, 2007; van Nes et al., 2008). In order to represent the limited capability of treating the large choice set by the decision maker, bounded rationality has been recently proposed for route choice behavior (Di and Liu, 2016). Although it is more realistic, the application of the bounded rationality remains as a future research topic since our sample size is small while the parameters on bounded rationality are usually latent and difficult to identify and estimate (Di and Liu, 2016).

Bovy (2009) in his review states that constrained enumeration approaches are promising, and also points to a random walk method by Frejinger et al. (2009) as another promising method. Kaplan and Prato (2012) applied both the simulation approach using randomly distributed errors on link costs and the random walk method in order to develop the choice set of the route choice, and obtained consistent parameter estimates of the route choice model with each other while the sizes of the choice sets are different from each other. Frejinger et al. (2009) assume that all possible paths belong to the choice set, so the method can be regarded as the integrated choice model described above. We apply the random walk method by Frejinger et al. (2009) in the context of walk-route choice in a downtown area having empirical data.

The random walk method is free from misspecification of the choice set, but it is difficult to set a parameter to determine randomness for efficient parameter estimation in the choice model. Frejinger et al. (2009) investigated the effect of the random walk parameter using a hypothetical single origin-destination pair, but it is not clear that the results agree with the empirical data, which contains significant variations in trip distance and network density. This study investigates the effect of heterogeneity in trip distance on sampled alternatives using the random walk method. More specifically, a structured random walk parameter according to the trip distance is proposed in this study, and the effectiveness in terms of parameter estimates of the nested logit model is investigated.

Regarding heterogeneity in the trip distance, the difference in the travel times among the alternatives may also have a varying effect according to the trip distance. For example, a route with a travel time five minutes longer than that of the shortest path is unlikely to be chosen when the shortest path has a travel time of only two minutes, given other conditions are equal. The two routes are comparable, however, when the shortest path takes thirty minutes. In other words, the error terms in the route choice model may have greater variance in longer trips than in shorter trips. Gliebe et al. (1999), Morikawa and Miwa (2006) and Li et al. (2016), among others, applied heterogeneous variances in the error terms for the travel times and trip distances, respectively, and obtained a better goodness-of-fit than conventional models with constant variance in the case of car trips. The heteroscedasticity of the error terms is considered in this study in combination with the sampled alternatives described above.

This study used a nested logit model of the mode and route choice to examine the effects of the walk environment on mode choice and route choice. As the alternatives of modes, only trips on foot and by subway are investigated in this study. Because once the subject came to the downtown area by car, mostly the car had to be used for the next trip within the downtown area. Thus, car trips were excluded from the analysis. Generally, conventional exogenous sample maximum likelihood (ESML) estimation with sampled alternatives provides consistent estimates only for a multinomial logit model. However, Bierlaire et al. (2008) proposed a consistent estimator for generalized extreme-value models with sampled alternatives. Also, Lee and Waddell (2010) proposed a nested logit model with sampled alternatives; this approach is applied in this study using sampled alternatives for the random walk method. Guevara and Ben-Akiva (2013) provided more general framework for the estimation of multivariate extreme value models with sampled alternatives. The approach by Lee and Waddell (2010) is consistent with Guevara and Ben-Akiva (2013).

Before, the route choice analysis had the difficulty in obtaining reliable sufficient data set since reporting the routes used in the traditional travel survey with paper and pencil was laborious. However, it has become practically possible thanks to the development of mobile phones with GPS functions. More and more travel behavior analyses have been carried out using the information extracted from the GPS trajectories (Yue et al. 2014; Chen et al. 2016; Wang et al. 2017). The mobile phones with GPS functions are also used to obtain the data on route choices in the empirical analysis of this study.

2. Methodology

2.1. Nested logit model with sampling of alternatives

According to Frejinger et al. (2009), suppose R_n paths are drawn with replacements from the universal set of paths U, and the chosen path is added to generate a set C'_n for observation n ($|C'_n| = R_n + 1$), where each path $j \in U$ has non-zero sampling probability $q(j)$. Alternative j appears k_{jn} times in C'_n, where $k_{jn} = k'_{jn} + \delta_{jc}$, k'_{jn} is the number of times alternative j is drawn ($\Sigma_{j\in U}\ k'_{jn} = R_n$), c is the index of the chosen alternative and $\delta_{jc} = 1$ if $j = c$ and zero otherwise. Suppose C_n is the choice set containing all alternatives corresponding to the R_n draws ($C_n = \{j \in U\ |k_{jn} > 0\}$). The size of C_n becomes one at the minimum when only duplicates of the chosen alternative are drawn, and becomes $R_n + 1$ at the maximum when the chosen alternative is not drawn and the draws contain no duplicates. The multinomial logit model can be consistently estimated with the conditional probability given as

$$P(i|C_n) = \frac{\exp\{\mu V_{in} + \ln(k_{in}/q(i))\}}{\sum_{j\in C_n}\exp\{\mu V_{jn} + \ln(k_{jn}/q(j))\}} \tag{1}$$

where μ is a scale parameter and V_{in} is the deterministic utility.

In our study, a nested logit model of the mode and route choice is developed with sampled alternatives, where the nesting structure is assumed to include a binomial mode choice between subway and walking at the top level and a multinomial walk-route choice at the bottom level. Eq. (1) is used as the bottom level conditional choice probability. For the top-level marginal choice probability, the logsum should be calculated. The logsum is the log of the denominator of the conditional probability multiplied by $1/\mu$ for conventional nested logit model. Lee and Waddell (2010) proposed the expanded logsum for nested logit models with sampled alternatives at the bottom level, which is given as

$$V'_{wn} = (1/\mu)\ln\left\{\sum_{j\in C'_n}[\{1/q(j)\}\exp(\mu V_{jn})]\right\}. \tag{2}$$

where V'_{wn} is the logsum associated with walk nest. Eq. (2) approximates the true logsum using a scaling adjustment with inverse sampling probability on randomly sampled alternatives. Eq. (2) is consistent with the application to the nested logit of the more general framework proposed by Guevara and Ben-Akiva (2013). Guevara and Ben-Akiva (2013) proved the consistency of the parameter estimates of the expanded logsum in more general framework of multivariate extreme value models including nested logit models. Eq. (2) can be rewritten by considering the duplicates among the draws as

$$V'_{wn} = (1/\mu)\ln\left\{\sum_{j\in C_n}[\exp\{\mu V_{jn} + \ln(k_{jn}/q(j))\}]\right\}. \tag{3}$$

The top level marginal choice probability is given as