Destination choice set enumeration using a behavior-similarity human network

Y. Fujiwara[a,*], J. Urata[b], M. Chikaraishi[c], A. Fujiwara[d]

[a,c,d] Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, Japan
a.m213720@hiroshima-u.ac.jp, b.urata@bin.t.u-tokyo.ac.jp, c.chikaraishim@hiroshima-u.ac.jp, d.afujiw@hiroshima-u.ac.jp
[b] Department of Civil Engineering, The University of Tokyo, Tokyo, Japan
* Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

March 31, 2022

Keywords: Destination choice, Smart card data, Choice set, Non-daily trip, Behavior similarity

1 INTRODUCTION

Accurate prediction of destination choice behavior for rare-happened traffic demand (RHTD) is difficult, and particularly because the number of choice sets is finite but extremely large. In the context of the route choice problem, many researchers have worked on choice set enumeration, and various methods have been systematically compiled (Frejinger et al., 2009, Yao et al., 2020). In contrast, less study has been conducted on choice set reduction or enumeration methods for the destination choice problem, although its importance has been pointed out (Crompton and Ankomah, 1993, Decrop, 2010). This study proposes a novel method of choice set enumeration to address the destination choice problem. The proposed approach is inspired by e-commerce recommendation systems, where a destination choice set for non-daily public transport usage is enumerated using the personal behavioral history of people who behave in a similar way. There are two specific objectives. The first is to develop an association network-based rare-happened destination (RHD) prediction method based on behavioral similarity of public transport usage. The second is to demonstrate the prediction accuracy of the proposed method by empirically verifying the model using smart card data and comparing the results with those of other conventional destination choice models and a deep neural network model.

2 METHODOLOGY

2.1 Association network and enumeration

The main idea of our approach is as follows. First, an association network is formed based on the degree of similarities identified by means of the spatiotemporal patterns of non-daily trips using long-term behavior history data. A similarity is defined as the ratio of the number of stations commonly used in the past to the total number of stations used by two users. We then form a choice set for non-daily destination choice from the destinations visited by others who are close to each other in the association network. In this method, destinations that the user has never visited can be included in the choice set. The idea was inspired by collaborative filtering, which is often used in the context of e-commerce recommendation systems. A number of time-series models have been proposed, where an individual's own behavioral history information is used to predict their next travel behavior. To the authors' knowledge, however, there has been no study to date that uses the behavioral history of others for choice set enumeration.

When configuring the association network, the optimal network size is searched for each individual with two indicators. (1) Cover ratio (CR): the percentage of user *i*'s visited destinations that is covered by the enumerated alternatives using the proposed method. (2) Enumerated destinations (EDs): the number of enumerated alternatives. The second indicator is introduced to evaluate whether the choice set size is kept as small as possible. We prefer to keep the choice set as small as possible in principle, largely because the low predictability of the destination choice model essentially comes from the large number of alternatives that cannot really be distinguished by their alternative-specific attributes. Note that a trade-off relationship exists between the CR and the EDs: if the number of EDs is too small, the choice set contains only a few or no stations that an individual has visited before, resulting in the lower value of the CR. Therefore, the optimal network size for each individual should be determined by taking the trade-off into account.

Given the above discussions, we propose an enumeration algorithm shown in Figure 1. In the algorithm, a behavior-similarity network is constructed using top *n* similarity users, where *n* is optimized to maximize the ratio of the CR and the EDs for each individual.

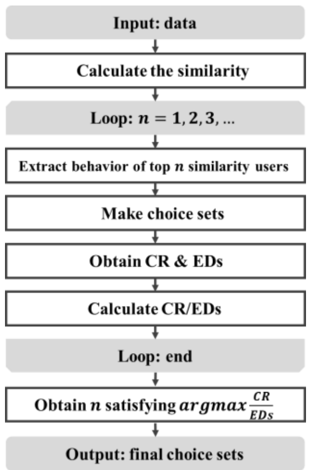

Figure 1 – *Framework of our enumeration algorithm*

2.2 Evaluation

To confirm the performance of the proposed method, we compared the following five destination models:

(a) Multinomial logit (MNL) model with the proposed choice set,
(b) MNL model with choice set from user's own history,
(c) MNL model with choice set from major destinations,
(d) Deep learning model with full choice set, and
(e) MNL model with full choice set.

In (a), the choice set is generated using the approach described above, applying the conventional MNL model. In (b), the choice set is generated using the past non-daily trips of the users. For the model (c), 71 major destinations are used as a choice set. The models (d) and (e) used all the 3654 destinations. Models from (a) to (c) used the same parameters obtained by estimating the MNL model (e). The parameters were estimated by the maximum likelihood method using approximately 10,000 randomized trips for RHDs. Additionally, we compared a model with alternative-specific constants only, for the comparison.

We conducted an out-of-sample validation by comparing the average of predictability. The average predictability is defined as:

$$Predictability = \frac{\sum_i \delta_i P_i}{|I|}, \qquad \delta_i = \begin{cases} 1 \; if \; i = s \\ 0 \; if \; i \neq s \end{cases}, \tag{1}$$

where, s is the actual station chosen, and the predictability means the average choice probabilities for chosen stations. If the choice set does not contain the alternative, P_i is set as zero. In our proposed model, the enumerated choice set does not necessarily include the actual destinations. Therefore, there is a possibility that log-likelihood cannot be calculated. To eliminate this problem, we use predictability as an evaluation index.

3 DATA

This study empirically verifies the performance of the proposed method using one-year smart card data (SCD) in the Hiroshima metropolitan area in Japan. The SCD used in this study were collected over a 365-day period. The study analyzed 236,179 card holders who had traveled more than 100 times during this period. The number of observed stations within the public transportation system was 3655, which was used as a full choice set. Note that this study regarded the station as the traveler's destination. We defined RHDs as trips to particular destinations that had not been visited during the past 30 days, and/or destinations that had not been visited in the past 10 trips in an individual's travel record. From the personal behavior histories, 6% of all trips were RHTDs and did not involve major destinations.

4 RESULTS

4.1 Association network and enumeration

To validate our enumeration algorithm, we divided the data set into two: data for the first 10 months are used as a training set, and those the last two months are used as a validation set due to data constraints. Figure 2 shows the distributions of CR and EDs for the selected 9,872 users who had RHDs in their travel histories. The average number of the enumerated choice sets is 15.4. This number is less than 1/200 of the full choice set, which means that our algorithm was able to largely reduce the choice set.

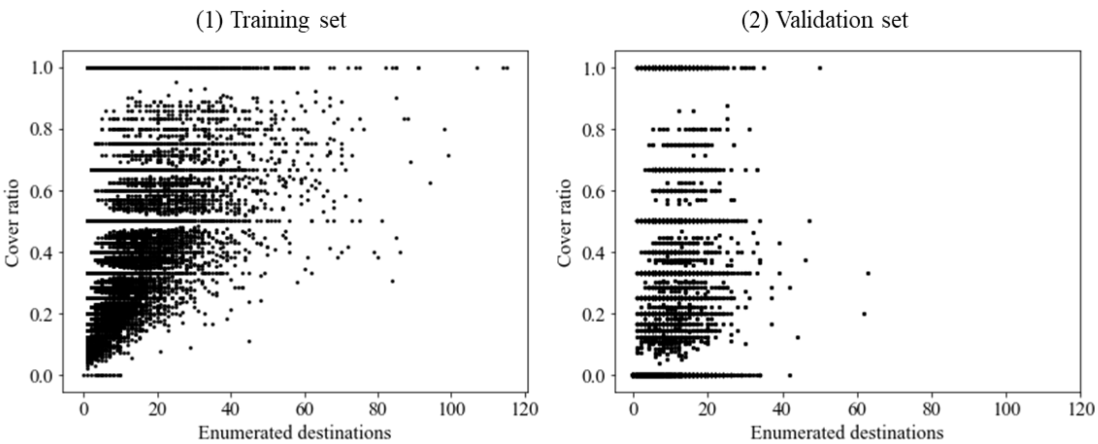

Figure 2 – *Cover ratio variations as a function of the number of enumerated choice sets for training (1) and validation (2) sets*

4.2 Model predictability

The results of model predictability are shown in Table 1. The results show that our proposed model outperforms all other models. This indicates that focusing on the method of enumerating the choice sets, rather than increasing the complexity of the model structure as in the case of methods such as a deep learning model, contributes to improving the model accuracy. In addition, focusing on the method of enumerating choice sets, it is clear that enumerating choice sets based on the behavioral history of those with similar behavior is more predictive than an approach based on individuals' own past behavioral history.

Table 1 – *Predictability of each model*

Models	*Predictability*
(a) MNL model with the **proposed choice sets**	0.1020
(b) MNL model with choice sets from **user's own history**	0.0588
(c) MNL model with choice sets from **major destinations**	0.0230
(d) **Deep learning model** with full choice sets*	0.0105
(e) MNL model with **full choice sets**	0.0036
(e') MNL**-ASC** with full choice sets	0.0028
(a+c) MNL model with **proposed choice sets** and **user's own history**	0.0700

*Deep learning model architecture: 4 parameters (*3654 stations), 4 layers, multilayer perceptron

5 CONCLUSION

In this study, we proposed a new choice set enumeration method for the destination choice prediction of non-daily public transport usage. Specifically, we developed an association network-based RHD prediction method based on behavioral similarity of public transport usage and showed that the predictability of the proposed model is superior to that of conventional methods that do not enumerate choice sets or other possible simple choice set enumeration methods. It is also very interesting to note that for non-patterned behaviors such as RHTD, the proposed method performed better than data-driven methods such as deep learning. We have also shown that the choice of enumeration methods could have a larger impact on predictability than the choice of model structure.

References

E. Frejinger, M. Bierlaire, M. Ben-Akiva, 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B,* **43** pp.984-994.

R. Yao, S. Bechor, 2021. Data-driven choice set generation and estimation of route choice models. *Transportation Research Part C,* **121**, 102832

J.L. Crompton and P.K. Ankomah, 1993. Choice set propositions in destination decisions. *Annals of Tourism Research,* **20**, pp.461-476.

A. Decrop, 2010. Destination choice sets an inductive longitudinal approach. *Annals of Tourism Research,* **37** pp.93-115