Estimation of city tourism tours with survey data from Kyoto

Kai SHEN[1], Jan-Dirk SCHMÖCKER[2], Yuhan GAO[3] and Ali Gul QURESHI[4]

[1]Dept. of Urban Man., Kyoto University
(Katsura C-Cluster C1-2-434, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail: kshen@trans.kuciv.kyoto-u.ac.jp
[2]Member of JSCE, Associate Professor, Dept. of Urban Man., Kyoto University
(Katsura C-Cluster C1-2-436, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail: schmoecker@trans.kuciv.kyoto-u.ac.jp
[3]Dept. of Urban Man., Kyoto University
(Katsura C-Cluster C1-2-434, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail: gao@trans.kuciv.kyoto-u.ac.jp
[4]Member of JSCE, Associate Professor, Dept. of Urban Man., Kyoto University
(Katsura C-Cluster C1-183, Nishikyo-ku, Kyoto 615-8540, Japan)
E-mail: qureshi@trans.kuciv.kyoto-u.ac.jp

In order to manage the growing tourist numbers, new concepts of tourist behavior modeling are required. So far, many studies used statistical approaches to predict tourists' movement, which only gives descriptive results but is of limited help for tourism "management". In order to understand tourists' behaviors and predict the routes they take when traveling among urban attractions, we take an analytical approach and model the tourist movements as a problem for tourism experience (utility) maximization.

We adopted a similar objective function given in most literature, where tourists evaluate and rank the points of interests (POIs) according to their tastes, and try to minimize their cost on traveling by planning an optimal route, taking into account time and other constraints.We also included following features: 1. the choice of destinations will be "history dependent" in that there is diminishing marginal utility gained by visiting additional POIs; 2. since attractions might be classified into several categories their intrinsic utilities are evaluated over multiple dimensions. The utility-based tourist behavior model allows us to estimate the effects if changes to the transport system or entrance fees are made. Our goal is to simulate tourism management strategies under various scenarios.

Key Words : *1. behavior modelling, 2. utility maximization, 3. tourist preference, 4. Tourist Trip Desgin Problem, 5. tourism demand management*

1. Background and Objectives

City tourism is a major business across the world and has become an essential part of the economy. At the same time its steep growth is creating congestion inside cities. In Kyoto, Japan, for example, the number of tourists exceeded 55million in 2016[16)]. The resulting crowding especially at point of interests (POIs) is leading to frustration among tourists. Furthermore, many Kyoto residents perceive the large number of tourists often as negative and avoid visiting the city center or Kyoto tourist sites.

So far, many studies used statistical approaches to predict tourists' movement. Popular studies include using mobile phone statistics to predict tourist flows and OD matries. Other studies are utilizing Wi-Fi packet sensor data to explore popular travel patterns of tourists. For example, Zheng et al used on-site GPS trajectories for movement pattern minning and predict the tourist's next location within a given attraction[11)]. Zheng et al analyzed tourist movement patterns and topological characteristics of travelers' routes based on movement trajectories of photographers generated from geotagged photos on social media[12)].

However, the limitations of above studies are that they only gives descriptive results which are of limited help for tourism "management" as they can not evaluate the effects of travel demand strategies.

Therefore, to manage the growing tourist numbers new travel demand management concepts are required. The basis for such concepts is understanding and predicting the routes tourists travel inside cities. These will allow estimating the effects if changes to the transport system (or entrance fees or other type of regulatory measures) are made. There appears to be, however, very limited literature describing the travel behaviour of tourists inside cities, possibly partly because choices are difficult to estimate and partly because traditional travel surveys tend to focus on residents instead of visitors. This research aims to reduce this gap by aiming to define a choice model that describes the tours of tourists.

2. Literature on (tourist) tour modelling

A "tour" defines a sequential visit of different destinations. Applied to city tourism, route choice and decisions such as time spent at an attraction will be closely linked. Sometimes it is not even a particular, singular destination the tourist aims to reach but generally the goal is to explore a place within a given time (and with a given budget). Features making tourists' decision-making process unique are that they evaluate sightseeing sites based on limited information resources and travel according to a personalized itinerary which can maximize their satisfaction[8). Factors that influence tourist movements identified in the literature generally include a set of destination characteristics, trip characteristics and a set of tourist characteristics[1,2)].

Operations research literature describing individual tourist behaviour through utility functions sees the decision of choosing a specific tour route as a process of solving an optimization problem. In this kind of tour route planning problem (TRPP) or tourist tour design problem (TTDP), tourists evaluate and rank the POIs according to their tastes, and plan an optimal route between these, taking into account time budgets and other constraints. For this kind of tour planning which is a variant of the well-known orienteering problem (OP), heuristics are used for deriving an optimal tour, searching for results that may have higher utilities until the constraints are reached or travel costs outweigh the utility that might be gained from visiting any further POI[3~8)].

However, applying the TRPP or TTDP requires first to understand the choice parameters. Furthermore, although above optimization approach for modelling tourists' trip chain making process is reasonable for many experienced travelers who spend time planning trips and have a strict budget on time and cost, there is also a large group of leisure travelers whose behaviour will be influenced by additional factors such as "fatigue". We suggest that the choice of destinations will be "history dependent" in that there is diminishing marginal utility gained by visiting additional POIs over the course of the tour. In other words, once a few attractions have been visited, the likelihood of skipping attractions even if there would still be sufficient time will increase.

Furthermore, the TTDP literature usually optimizes a single objective function, whereas attractions might be classified into several dimensions. For instance, a hiking area with spectacular scenery could have a high score in terms of natural beauty and outdoor exploration but will never be labeled as a place for leisure activities such as entertainment parks or shopping malls. Similarly, museums and galleries are given high scores in terms of cultural and art activities but will have relatively low values for natural sceneries. Not only does the destination have a range of intrinsic utilities in multiple dimensions, but also the preference of tourists varies with respect to these dimensions and a tourist might want to satisfy several of these dimensions at least to some degree over the course of his tour.

Sasaki et al[9)] summarized sightseeing facilities in Kyoto, Japan into the three categories "Downtown", "Shrine and Temple" and "Natural beauty". Yuichiro et al[10)] collected eleven significant adjectives paired with opposite meanings for expressing the characteristics of tourist attractions and reduced them subsequently into three categories using factor analysis; Becken et al[1)] used factor analysis to reduce attraction categories into 5 dimensions (factors).

In our research, we follow the three-category approach and label them as "Natural and Scenery", "Cultural and Art" and "Leisure and Entertainment". We evaluate both intrinsic utilities of POIs and tourists' preference with respect to these three dimensions.

3. Tourist choice problem formulation

Denote a complete and undirected graph network as G = (V, E), where the vertex set N is a combination of the POI (attraction) set $Q = \{v_1, v_2 \ldots v_n\}$, and the origin and destination set $S = \{v_{n+1}, v_{n+2}, \ldots, v_{n+s}\}$. The edge set $E = \{(v_i, v_j): v_i, v_j \in V, i<j\}$ represents the paths connecting the vertices V. To do so, we therefore preprocess the transport network to find the mode specific shortest paths between different POIs, origins and destinations.

Each vertex in Q corresponds to a POI or an attraction area that has an intrinsic utility denoted as $\boldsymbol{U_i} = (u_{i,1}, u_{i,2}, u_{i,3})$, where each entry has a value regarding the attractiveness of the POI across the above named three dimensions. We suggest that these utilities can be roughly estimated according to guidebooks, user ratings and popularity. Each edge is associated with a non-negative travel time t_{ij} and cost c_{ij}.

We then model the tourist movements as a problem for tourism experience (utility) maximization, in which tourists choose the destinations that are best tailored to their preference and an optimal order to visit them within time and monetary budgets. It is assumed that the preference of a tourist n is characterized by a vector $\boldsymbol{P}_n = (p_{n,1}, p_{n,2}, p_{n,3})$, with an entry for each dimension regarding a traveler's taste. Further, in line with above discussion, instead of making the total utility achieved by tourists additive to each individual destination visited, we consider interactions between destination visits by introducing a diminishing marginal utility along with more utilities being achieved.

The objective function of the traveler is thus to decide an ordered combination of POIs which satisfy his/her interests most including consideration of the route costs. This objective function is formulated in (1):

$$\max U_n | o, d = u_{oi_1} + \sum_{k=1}^{m_p-1} \left(u_{ni_k}^P + u_{i_k i_{k+1}}^T\right) + u_{ni_m}^P + u_{i_m d}^T \quad (1)$$

where we presume that origin o and destination d (e.g. a common entry point to the city, or a hotel) of the person are given; i_k denotes the k'th POI visited and the tourist aims to maximize his utility by visiting m attractions before reaching his/her destination. In (1) u_{ni}^P denotes the positive attraction of person n to visit POI i which we presume to be a function of the previously visited POIs. Further, u_{ij}^T defines the negative utility of travelling from i to j. These utilities can be further specified as follows:

a) Utility of traveling on each edge

The utility of traveling on each edge is assumed to have a linear time and cost function as in (2) where $\varphi = 1/\mathrm{VOT}$ transfers the monetary cost into time:

$$u_{ij}^T = \alpha_1 t_{ij} + \alpha_2 \varphi c_{ij} \quad (2)$$

b) Utility of visiting nodes

The utility of visiting the k'th POI in the journey for person n, $u_{n,k}^P$, is decided by the interest of a tourist in a specific POI (personalized score of the location), which is determined by both the tourist's preference and the intrinsic utility of that POI:

$$u_{n,k}^P = \beta_1 \boldsymbol{P}_n^T * (\boldsymbol{U}_{i_k} \circ \exp(\beta_2 \boldsymbol{A}_{n,k})) + \beta_3 T_{i_k} \quad (3)$$

The '$\circ$' mark in (3) stands for the entrywise product of two vectors. $\boldsymbol{A}_{n,k}$ is a vector with 3 entries similar to the intrinsic utility of POIs, that represents the accumulated utility gained from the POI visits before arriving at current POI i_k. Parameter β_2 is assumed to be negative since more cumulated satisfaction is reducing the benefit to visit another place. If β_2=0 then the travel history does not have any influence on a person's subsequent decision.

4. Data description

Table 1 Illustration of observed tourist trip chain data

p-ID	origin	dest.	dep. H	dep. M	arr. H	arr. M	food cost	souvenir
60222	56	1	7	30	9	0	0	1500
60222	1	20	9	15	10	0	700	2000
60222	20	25	12	0	12	15	2940	0
60222	25	25	13	30	13	40	0	1500
60222	25	99	15	0		0	0	0
240102	38	27	9	0	10	0	0	0
240102	27	24	10	20	10	40	0	0
240102	24	25	11	0	11	20	0	0
240102	25	38	11	30	12	0	0	0

The data used for model estimation are taken from a survey of tourist movement in Kyoto city, conducted in November 2006. The questionnaires were distributed in attraction areas and train terminals and include:

a) socio-demographics such as age, occupation, ownership of vehicles, home city, etc.;
b) tour related attributes such as travel purpose, schedule, travel group, impressions about Kyoto city;
c) a trip diary of detailed trip chains which consists of destinations, travel time and mode choice.

There were about 3,400 valid questionnaires received. We note that there are also two complementary PT surveys conducted in Kansai area of Japan that surveyed socio-demographics and travel patterns of each participant as well as their detailed trip chains. Participants were not specified to tourists exclusively but mainly included commuters and residents. Table 1 provides a glance at how observed trip chains are presented in the data. Note that expenses (food and souvenir) in JPY at attractions were also collected and missing values also exist in the dataset.

5. Methodology

(1) Overview

Figure 1 illustrates the calibration process of the utility parameters that describes tourist behavior when making tours.

As we mainly focus on the behavioral model, the intrinsic utilities of attractions and tourists' preference are evaluated before the calibration process and are taken as input. Tour related attributes such as origin, destination and time budget are utilized for setting the constraints in the optimization problem.

A complete and undirected graph network is then constructed where each node stands for an attraction area or transit entry point to Kyoto. Mode-specific travel time, distance and transit fare matrices between any two nodes are measured by quering Google Maps API and are averaged from different periods throughout a day.

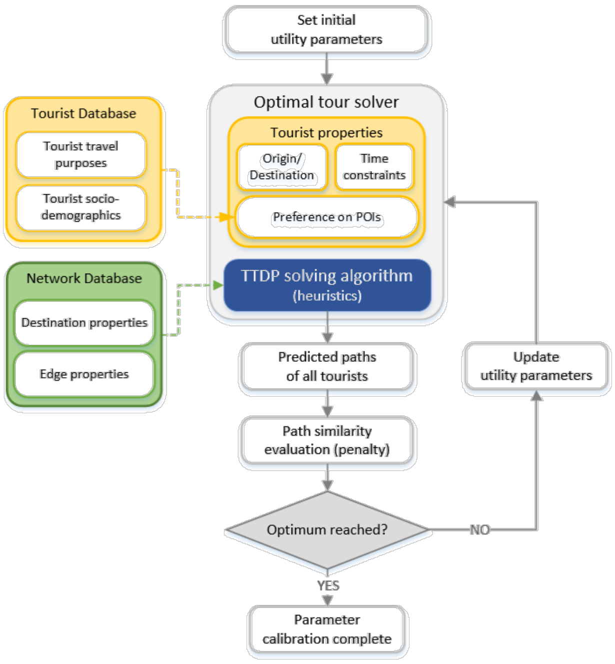

Fig.1 Parameter calibration framework.

With network database given as input, we solve the optimization problem and predict the most likely tour taken by each toursit under current parameters. Differences between the predicted and observed paths are then measured and taken as a penalty, which behaves like a numerical gradient that guides the solution towards the optimum. Eventually a set of parameters is derived that best describes the tourists' behavior in the model. This model allows us to simulate tourism management strategies under various scenarios. Following sections explain the key modules of the parameter calibration process.

(2) Preference prediction

Tourists' preference plays an important role in deciding where to visit when making tours. However, it is neither realistic to enumerate all available attractions nor convenient for respondents to answer in the survey.

Since tourists did not describe explicitly which type of sights they favor in the survey, we look for other information such as travel purposes to reflect travelers' tastes in attraction types. In the survey respondents were asked to choose up to three out of 17 options for their main reasons for coming to Kyoto. According to these answers, dummy variables are created presenting presence or absence of that attribute.

Dimension reduction is then performed to simplify the form of the preference vector, because as a model input we need to calibrate the intrinsic utility of attractions of the same length as the preference vector, thus a form with appropriate dimension size is preferred.

Table 1 Choice candidates for travel purposes.

Choice No.	Travel Purpose
1	shrine & temples
2	window shopping
3	night spots
4	cultural events, festival
5	leisure activities
6	red-leave tours
7	natural sceneries
8	gourmet, cuisine
9	shopping (souvenirs)

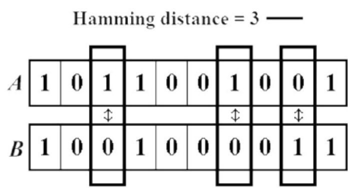

Fig.2 Illustration of Hamming distance.

We first reduced the number of choices from 17 to 9 by merging less chosen options into semantically similar and frequent equivalents. The choices merged are shown in Table 1. This is followed by a k-means clustering with K equal to 3 based on Hamming distance where centroids of choices are found and dominant choice patterns are extracted. In information theory, the Hamming distance measures the minimum number of substitutions required to change one string into the other[17)]. Figure 2 illustrates how the Hamming distance is calculated between two binary strings, that is, equivalent to the number of positions at which the corresponding bits are different.

We obtain following cluster centroids:

a) 1-6: red leave & temple shrines
b) 1-6-8: red leave & temple shrines & gourmet
c) 5: leisure activities

An ideal approach is to predict the tourists' preferences by obtaining information such as the socio demographics and trip related attributes, without having to ask or survey each time. Each tourist will be assigned to the corresponding cluster based on his answer to the travel purpose.

We then estimate a multinational logistic model in which the socio demographics of the visitors and their travel-related attributes are used as explanatory variables, while the preference label is used as the categorical dependent variable.

With significant factors from the socio-demographic data, we predict the probability of belonging to each of the cluster using multi-nomial logit regression. The probability of belonging to each cluster is taken as the weight under each preference type, which add up to 1. Such a vector is used as a preference vector for each tourist.

Since the multi-nomial logistic regressions ensure that the probability of each category adds up to 1, the tourists' preference vectors are naturally normalized to the same scale. We use them as input to the solver.

(3) Evaluation of attraction utilities

The intrinsic utility of destinations should also be estimated to run the optimal tour solving algorithm. In the survey, the destination is defined as a large area around one or several main sights as shown in Figure 3, which may include multiple different types of POIs.

Fig.3 Example of attraction areas in the survey.

In order to calculate the utility of the attraction areas we avoid using the frequency of visits to the attraction as an indicator as this would lead to endogeneity issues as the explanatory variable would be correlated with the error term. Instead, we suggest that the intrinsic utility of attractions in the corresponding dimensions can be roughly estimated based on guidebooks, user ratings and popularity. In addition, with Google Map and OpenStreetMap Place Query API, we conducted POI searches in each region. Intrinsic utilities in the same dimensions as the preference vector are evaluated by following metrics.

a) Red leave

The score on the 'red leave' dimension is assessed by the average number of popular destinations in each area that have been rated high by various websites and magazines over the past few years.

b) Temple and shrines

Kyoto, a city known for its temples and shrines. The city has the most traditional Buddhist culture in Japan. It is said that there are about 800 shrines and 1,700 Buddhist temples located around Kyoto. Both the quantity and popularity of temple and shrines are evaluated when assessing the score of an attraction in this dimension. Specifically, we give weights to each POI based on number of reviews and user ratings parsed from Google Map Place Query.

c) Gourmet

We evaluate the score on gourmet dimension in two ways: the ease of finding a place to eat, and the number of high-end restaurants in the area. Number of ordinary restaurants, bars and pubs as well as high-end restaurants are calculated by Google POI search.

d) Leisure activities

Scores in this dimension are evaluated by enumerating the number of facilities associated with leisure activities. Specifically, the number of shops, museums and art centers are counted and normalized respectively. An overall score is calculated by averaging above scores upon different categories.

(4) TTDP solving algorithm

Tourist tour design problem (TTDP) is NP-hard and can be formulated as an integer programming problem. Exact solutions based on branch-and-bound, branch-and-cut are only feasible for small-scale graph, whereas approximation algorithms are either too difficult to implement or have high execution time in practice[14].

Other than exact solutions, numerous heuristic rules were developed for solving optimization problems like TSP, OP and their variants. In comparison with exact solutions, heuristics have the advantages of being intuitive, easy to implement, and fast in terms of computational effort.

Since more than 3,000 optimal paths are to be solved for each set of parameters, developing a fast algorithm is a must. Specifically, we adopted a modified algorithm based on the OP solving heuristic by Chao[13] that includes 'exchange', 'improvement' and 2-opt 'clean up' steps to approximate the optimal solution. A general framework of the heuristic is illustrated below in pseudo code.

```
Input: edge/node database, origin/destination, preference, etc.
Step.1: Initialization     (insertion)
        set record and calculate deviation
Output: Path_op, Path_nop (set)
Input:
Step.2: Improvement
        A loop:               (reinitialization)
                B loop:        (improvement)
                        2-point exchange
                        1-point movement
                        2-opt clean up
                        if a better solution found:
                                set record and calculate deviation
                        else: end B loop
                end B loop
                reinitialization
        end A loop
```

Fig.4 Heuristic in pseudo code.

(5) Path similarity evaluation

Let R_n denote the route among the set of possible routes corresponding to $R_n = \{r \in R | \max U_n | o, d\}$.

Our goal is to find the set of parameters $\{\boldsymbol{P}_n, \boldsymbol{\beta}\}$ that minimise the spatial difference between observed routes $\hat{R}_n$ and estimated routes R_n.

Various kernel and distance functions as well as similarity coefficients are defined to compute pair-wise similarity between sequences[15]. Metrics like longest common subsequence (LCSS) have also been applied in the literature to estimate the similarity of tourist movement sequences[12].

Since the traveler may slightly change the order of visit between several neighboring nodes, the similarity metric needs be robust to noise. We suggest therefore a better way is to combine the assessment process with geographic interpretation. In order to so we initially use following metric to calculate the spatial difference between the predicted and observed visited points:

Let m_p be the number of visited points in the predicted path and $\widehat{m}_p$ be the true, observed number of POIs visited by person p. Further, let $u^T{}_{i_k\hat{i}_k}$ denotes the generalized cost of travel between the k-th POI i visited on the predicted and observed journey of person p. The difference D_p between two paths are computed as follows.

$$D_p = \sum_{k=1}^{\min(m_p,\widehat{m}_p)} u^T{}_{i_k\hat{i}_k} + \begin{cases} \displaystyle\sum_{k=\min(m_p,\widehat{m}_p)+1}^{\widehat{m}_p} u^T{}_{i_{\min(m_p,\widehat{m}_p)}\hat{i}_k} & \text{if } m_p < \widehat{m}_p \\ \displaystyle\sum_{k=\min(m_p,\widehat{m}_p)+1}^{m_p} u^T{}_{i_k\hat{i}_{\min(m_p,\widehat{m}_p)}} & \text{if } m_p > \widehat{m}_p \end{cases} \quad (4)$$

By accumulating the difference between the predicted and observed paths of each visitor, we derive the cost, or in other words, the fitness of the current set of parameters in describing the modeled behaviors of tourists. Since the problem does not have a closed form formulation, we use a heuristic method to calibrate behavioral paramters.

(6) Parameter update process

Since the fitness (cost) of each set of parameters is obtained by summing up the prediction errors between observed and estimated paths of all tourists, it does not appear to have a closed form formulation and hence no analytical gradients for the objective function. Thus, using an off-the-shelf solving algorithm like gradient descent will not be feasible.

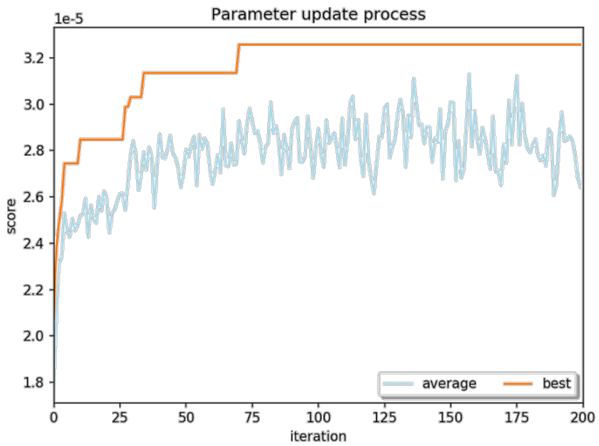

Fig.5 parameter update process in 200 iterations.

Therefore, we initially use a heuristic to solve the problem. Specifically, we first sample and compute the sample scores in the solution space at certain intervals to roughly grasp what the solution space looks like. This step allows us to see if ‘peaks’ exist and hopefully the ‘optimal’ solution will be unique. We then used a genetic algorithm-based framework to update model parameters.

Despite long computation time for a single set of parameters, the parallel evaluation nature of each parameter set allows us to apply multi-process programming, which speeds up the evaluation more than 10 times. The boost is determined by the number of threads the processor can provide.

The parameter update process using a genetic algorithm framework is illustrated in Figure 5.

5. Results and conclusion

(1) Tourist preference prediction

Through clustering we have three choice patterns that are dominant among tourists’ choice set. The cluster centroids correspond to:

a) 1-6: red leave & temple shrines
b) 1-6-8: red leave & temple shrines & gourmet
c) 5: leisure activities

In about 3,400 valid cases, 20% of respondents belong to cluster a, 63% belong to cluster b, and the remaining 17% belong to cluster c.

Multinomial logistic regression shows that the variables that have an important influence on determining clustering include age, travel peers, where they come from, frequency of visits to Kyoto as well as travel schedule. Table 2 shows some of the factors that have a significant impact on clustering.

Concretely, results indicates that when taking cluster a as reference a tourist is more likely to be a member of cluster c if he resides in Kyoto or visits Kyoto more frequently. This makes sense as for locals, they may not perceive these attractions as indispensable in the sense that some well known landmarks may just be someplace you see every day on the way to work. On the other hand, some experienced travelers also avoid visiting the city centre during the tourist season, especially during the red leaves and cherry blossoms seasons which Japan is famous for.

Table 2 Significant factors for clustering (details omitted for brevity)

Independent variable effects

with significant contributes (- negative, + positive)

Cluster (Ref. 1-6)

1-6-8	5
- group: individual	- group: couple
- children in group	+ living in Kyoto
+ group: friends/colleague	+ visit frequency
+ dwelling days	
+ monetary budget	

Model fit R-Square: a. Cox & Snell: 0.181 b. McFadden's: 0.110

In addition, if a visitor is traveling with friends or colleagues or has a longer schedule in staying Kyoto, he is more likely to also want to experience "gourmet" on this trip, whereas for individual travelers or families with children the probability of including "gourmet" in travel purpose is smaller.

(2) Behavioral model estimation

We calibrate behavioral parameters for travelers with different modal split respectively. For now only those coming to Kyoto by transit are considered. The best set of parameters for tourists using transit are as follows:

$$u_{ij}^{T} = -0.321 t_{ij} - 0.025 \varphi c_{ij} \tag{5}$$

$$\begin{aligned} & u_{n,\,k}^{P} \\ & = 5.0 * \boldsymbol{P}_n^T * (\boldsymbol{U}_{i_k} \circ exp(-0.215 \boldsymbol{A}_{n,k})) \\ & + 0.270 T_{i_k} \end{aligned} \tag{6}$$

We give β_1 an initial value at 5.0 as reference. Negative coefficients on travel time and monetary cost represent the negative utility of travel between attractions, which in most cases is a fact recognized by everyone. There are also situations where a journey becomes a part of sightseeing and hence brings positive utilities, e.g. walking and exploring around an area. However, as we do not consider attractiveness along with a travel in the model, it is beyond the scope of this paper.

In addition, note that φ equals to $1/\mathrm{VOT}$ which transfers the monetary cost into time. After converting into the same scale, it is revealed that travelers are far less sensitive to transit fares than time when they perceive the impedance on travels. Positive coefficient on dwell time at attractions also indicate that travelers enjoy more as they stay longer. Moreover, as each entry in the intrinsic utility vector of attraction has been normalized between 0 and 1, the coefficient -0.215 of β_2 indicates that cumulated satisfaction has a considerable negative effect on the benefit for visiting another place.

6. Future work

For remaining of this research we suggest a more flexible and noise-robust function to measure similarity between path sequences.

Moreover, we hope to develop a combined destination and duration choice model such that the duration at attractions are also predicted.

Data used for our model calibration are from tourism survey conducted by city government in 2006. Nevertheless, based on the fact that newer surveys are not available and that it has never been analyzed in a utility model, we use it as a foundation for using newer data. In particular, we have recently obtained GPS tracking trajectories of tourists which we hope to use instead in further work. Ultimately, we hope to use the behavioral models to simulate tourism management strategies for a variety of scenarios.

REFERENCES

1) Becken, S., Simmons, D., & Frampton, C. (2003). Segmenting Tourists by their Travel Pattern for Insights into Achieving Energy Efficiency. Journal of Travel Research,42(1), 48-56. Doi:10.1177/0047287503253938
2) Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. Annals of tourism research, 33(2), 403-423.ISO 690
3) Souffriau, W., Vansteenwegen, P., Vertommen, J., Berghe, G. V., & Oudheusden, D. V. (2008). A personalized tourist trip design algorithm for mobile tourist guides. Applied Artificial Intelligence, 22(10), 964-985.
4) Butt, S. E., & Cavalier, T. M. (1994). A heuristic for the multiple tour maximum collection problem. Computers & Operations Research, 21(1), 101-111.
5) Abbaspour, R. A., & Samadzadegan, F. (2011). Time-dependent personal tour planning and scheduling in metropolises. Expert Systems with Applications, 38(10), 12439-12452.
6) Vansteenwegen, P., Souffriau, W., Berghe, G. V., & Van Oudheusden, D. (2011). The city trip planner: an expert system for tourists. Expert Systems with Applications, 38(6), 6540-6546.
7) Lu, E. H. C., Lin, C. Y., & Tseng, V. S. (2011). Trip-mine: An efficient trip planning approach with travel time constraints. In 2011 IEEE 12th International Conference on Mobile Data Management (Vol. 1, pp. 152-161). IEEE.

8) Oh, J. S., Kim, H., & Jayakrishnan, R. (2012). Tourist activity simulation model for assessing real-time tour information systems. Journal of Intelligent Transportation Systems, 16(3), 118-131.
9) Sasaki, K. & Nishii, K. (2010). A process model of destination set choice in tourist excursion with combination utility of the destination set.
10) Kinoshita, Y., & Yokokishizawa, H. (2008). A tour route planning support system with consideration of the preferences of group members. In 2008 IEEE International Conference on Systems, Man and Cybernetics (pp. 150-155). IEEE.
11) Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place?. Tourism Management, 59, 267-280.
12) Zheng, Y. T., Zha, Z. J., & Chua, T. S. (2012). Mining travel patterns from geotagged photos. ACM Transactions on Intelligent Systems and Technology (TIST), 3(3), 56.
13) Chao, I. M. , Golden, B. L. , & Wasil, E. A. . (1996). A fast and effective heuristic for the orienteering problem. European Journal of Operational Research, 88(3), 475-489.
14) Gavalas, D. , Konstantopoulos, C. , Mastakas, K. , & Pantziou, G. . (2014). A survey on algorithmic approaches for solving tourist trip design problems. Journal of Heuristics, 20(3), 291-328.
15) Rieck, K. , Laskov, P. , & Sonnenburg, Sören. (2006). Computation of similarity measures for sequential data using generalized suffix trees. Advances in Neural Information Processing Systems, 22(4), 1177-1184.
16) Kyoto City (2017). Total survey on tourism in Kyoto. https://www.city.kyoto.lg.jp/sankan/cmsfiles/contents/0000222/222031/28chousa.pdf.
17) Hamming distance. (2019). Retrieved from https://en.wikipedia.org/wiki/Hamming_distance.