京都大学大学院工学研究科
都市社会工学専攻修士論文
令和２年２月

Master’s Thesis
Department of Urban Management
Graduate School of Engineering
Kyoto University
February 2020

A methodology for estimation of city tourism tours considering multiple satisfaction criteria: Case study with Kyoto survey data

（複数の満足度基準を考慮した都市内観光回遊の推定法：京都市の観光調査データを用いた事例研究）

京都大学大学院　工学研究科　都市社会工学専攻

交通マネジメント工学講座　交通情報工学分野

沈 凱

Abstract

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 adopt a similar objective function given in most literature, where tourists evaluate and rank the points of interest (POIs) according to their tastes and try to minimize their cost on traveling by planning an optimal route, considering time and other constraints. Compared to the ordinary utility-maximizing principles, we also include the following features: 1. attractions might be classified into several categories, their intrinsic utilities and tourists' preferences are evaluated over multiple dimensions. Therefore, the satisfaction that tourists get is determined by the degree of matching between the intrinsic utilities and tourists' preferences; 2. the choice of destinations will be "history-dependent" in that there is diminishing marginal utility gained by visiting additional POIs. As a result, the multi-dimensional objective function helped to explain the variance in people's behavior in choosing POIs and deciding visiting orders. As a result, the multi-dimensional objective function helps explain the variance in people's behavior when choosing POIs and deciding the visiting order. Estimated results also revealed that by introducing the diminishing marginal utility, our model can better predict the correct number of visits as well as tours close to the observed ones.

The non-aggregate tourist behavior model allows us to estimate the effects of policies and investments, e.g. if changes to the transport system or entrance fees to attractions are changed. Specifically, we simualted various scenarios with two types of strategies that modify the edge travel impedance and node attractiveness respectively. By looking at the shift in tourists' travel patterns as well as the attraction visit frequencies, we found that both strategies have a significant effect in changing people's travel patterns, while the latter has a more direct and intuitive effect on introducing tours and visits to the improved area(s).

Key Words: 1. Travel Demand Estimation, 2. Behavior Modelling, 3. Utility Maximization, 4. Tourist Preference Estimation, 5. Tourist Trip Design Problem, 6. Tourism Demand Management

Contents

Chapter 1　Introduction

1.1 Background

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. Since 2015, the term “over-tourism” has been used frequently to describe the negative impacts ascribed to tourism[1]. The World Tourism Organization (UNWTO) defines over-tourism as "the impact of tourism on a destination, or parts thereof, that excessively influences perceived quality of life of citizens and/or quality of visitor experiences in a negative way"[2]. In Kyoto, Japan, for example, the number of tourists exceeded 55million in 2016[3], and the travel demand in Kyoto city is still rapidly growing. 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.

To capture tourists’ travel behavior and identify the movements of inbound tourists, various surveys were proposed and carried out. One of the advantages of such ordinary surveys (questionnaire, interview or web-based surveys) is the capability of collecting individual attributes and socio-demographics. However, these surveys are often criticized for their high cost and inefficiency, not to mention that they are laborious to design, conduct, and respond to. Moreover, the requirement of ‘active participation’ limits the number of respondents and thus the size of data acquired. It is also difficult to evaluate the accuracy/reliability of the received answers.

To overcome such limitations, passive surveys using statistical approaches to predict tourists’ movement has recently become a research focus. Because data in these methods is automatically collected, they ensure high efficiency and do not put burdens on surveyors.

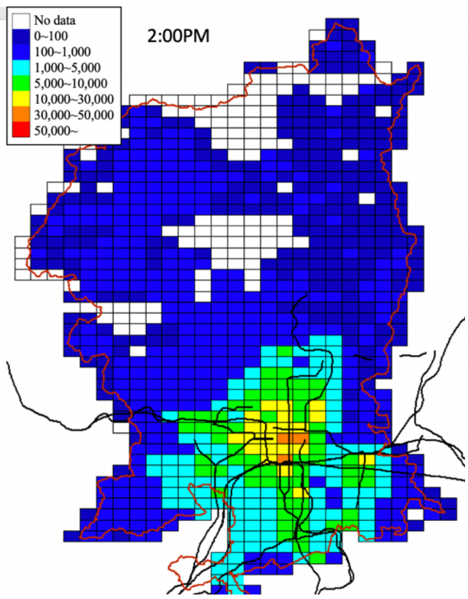


Figure 1-1 Mobile phone statistics

Figure 1-2 Trip sequence from Wi-Fi packet sensors

In the domain of passive surveys, popular studies include using mobile phone statistics to predict tourist flows and OD matrices[4]. 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 mining and predict the tourist's next location within a given attraction[5]. 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[6].

However, one limitation is that these data include none or few individual attributes or trip-related characteristics. Other drawbacks of the above studies include that they only give descriptive results which are of limited help for tourism "management" as they cannot 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 types of regulatory measures) are made.

1.2 Research aim

Although trip-based and activity-based models have been developed for ordinary travelers (work-based and home-based activities), there appears to be very limited literature describing the travel behavior of tourists inside cities, 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. Such behavioral model implies an abstraction of underlying motivation (driving force) that explains tourists' decision-making process as well as their sensitivity to environments, i.e. travel costs to a next location versus the potential satisfaction to be gained.

Specifically, we use non-aggregate methods to model the tourists' behavior in making tours, i.e. their decision-making process in choosing destinations and determining the appropriate order of visits, while taking into account their preferences for city attractions as well as a 'fatigue' factor that is history-dependent. An agent-based simulation system will then be developed based on such disaggregate representation of inbound tourists. Finally, this model allows us to simulate tourism management strategies under various scenarios. Our goal is to quantitively evaluate such TDM (tourism demand management) strategies, policies, and transit facility investments through this computational simulation system and propose cost-effective and promising solutions to alleviate urban congestion caused by growing tourism demand.

1.3 Methodology

We adopt a similar optimization principle for tourists from literature and formulate each decision-maker as rational who makes actions towards maximization of his satisfaction.

We calibrate the model using data such as socio-demographics, travel purposes as well as observed journey they actually made, such that it can best describe the key behaviors of most tourists. Figure 1.3 illustrates the calibration process of the utility parameters that describe tourist behavior when making tours. The general framework for estimating the behavioral parameters consists of two loops:

The outer loop generates parameters that describe tourist's behavior with respect to a set of observed tours. Metaheuristics are used for this to obtain the best fitting parameters. The inner loop enumerates all tourists from the survey and utilizes a problem-specific heuristic to predict possible tours. It compares the predicted paths against observed ones and finally sums up the prediction errors at each person.

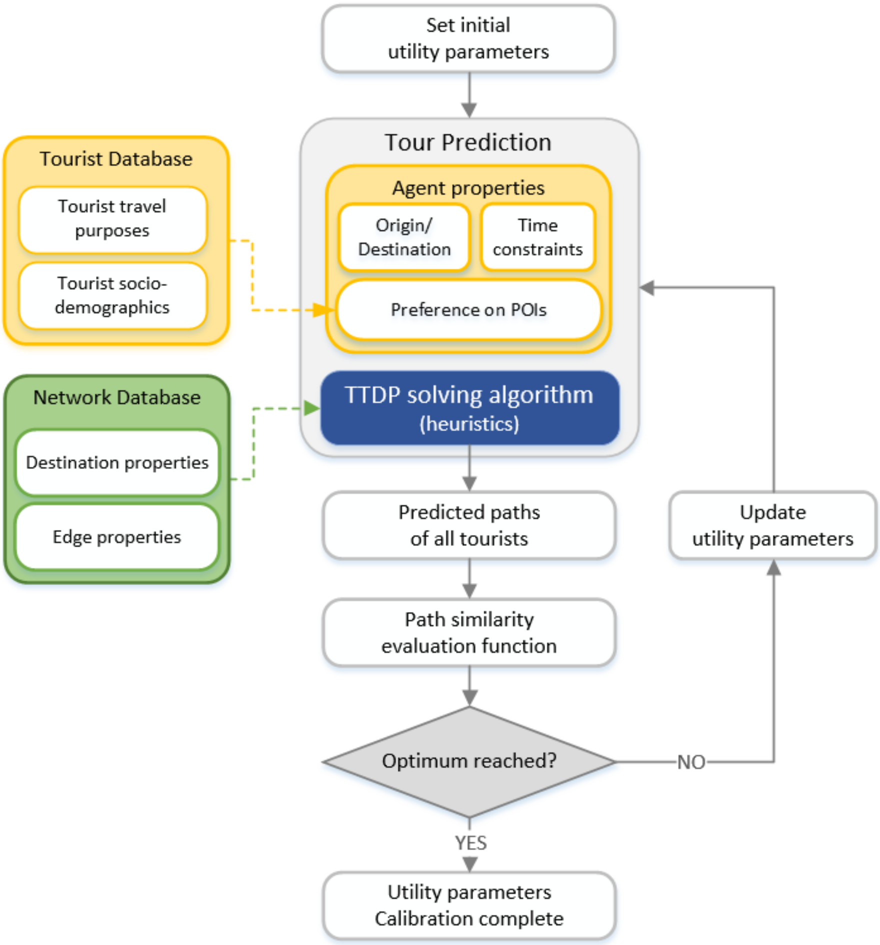


Figure 1-3 The Parameter calibration framework

We programmed a fast Tourist Trip Design Problem (TTDP) solving algorithm and adopted multi-processing to speed up the parameter update process.

As we mainly focus on the behavioral model, the intrinsic utilities of attractions and tourists' preferences 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 querying Google Maps API and are averaged from different periods throughout a day.

With the network database given as input, we solve a combinatorial optimization problem and predict the most likely tour taken by each tourist 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.

1.4 Thesis Outline

The organization of contents in this thesis is as follows: Chapter 1 explains the background and motivation for conducting this research, and briefly introduces the model methodologies. Chapter 2 studies previous attempts to model travel demand on both trip-based and activity-based approaches, discusses past studies concerning behavioral formulation in the Operations Research domain and explains why we choose a non-aggregate modeling approach to represent tourists' travel behavior in making tours. Chapter 3 presents the concepts and mathematical form of the tour activity schedule model and identifies important model design and features as innovation. Data processing and preparation for the model calibration are elaborated in Chapter 4. Chapters 5 describes the overall methodology, elaborates on key modules in the parameter calibration process and defines criteria for model performance evaluation. Chapter 6 presents the model estimation process and discusses on the estimated parameters. In Chapter 7, we use complaints data from the survey as an empirical reference for targeting problematic OD pairs. Such OD pairs are taken as targets at which we simulate the effectiveness of TDM strategies under various scenarios. Finally, Chapter 8 concludes the thesis and discusses specific ideas for future research to build on those conclusions.

Chapter 2 Literature review

A travel model is an analytical tool that provides a systematic framework for prediction of travel characteristics and usage of transport services, as well as for representing how travel demand changes in response to different input assumptions[7,8,9]. To date, two types of travel demand methods have been widely used: trip based and tour or “activity-based” travel demand forecasting models.

2.1 Trip based model

Trip-based travel models use individual person trips as the fundamental unit of analysis, leading to its nature of an aggregate approach. This well-known model often refers to as “four-step” models because it commonly includes four primary components. The four stages in this model are linked in a sequence and are largely independent. Trip-based models have several shortcomings which limited its performance in demand forecasting: 1) Since trips made by an individual are treated as separate, independent entities, this approach failed to consider the entire tour or relations between linked trips; 2) Direction of movement and other spatial-temporal constraints are neglected[10]; 3) It does not incorporate the time dimension and disregard time-space constraints. Only individual trips made during a peak hour is modeled[11]; 4) Individual behaviors in a household context as well as the intra-household interaction are not considered.

Despite the fundamental shortcomings in the trip-based model, their impacts were initially considered relatively small in applications for investment decisions of large-scale infrastructure[12]. Since the 1990s, however, calls for resolving transportation problems by better operating existing facilities instead of building new ones received great attention. This raised the voices to replace the classic four-step model by activity-based models, which adopted more disaggregate modeling approaches to make the model more sensitive to such TDM strategies and policies.

2.2 Activity-based model

Activity-based travel demand models suggest that travel is a derived demand from activities participation[13]. These models portray how people plan and schedule their daily travel. Activity-based travel demand models simulate the activity-travel decisions of households and individuals that collectively result in the activity patterns observed.

Compared to trip-based models that take trips in each aggregate zone as unit of analysis and base trip generation and distribution mechanism on social physics, the focus of activity models is on comprehensive, integral activity-travel patterns, of which the trip OD matrices are aggregated by derived trips from activity participation. Disaggregate approaches have always been used, representing the behavior at the level of the individual traveler. The development of comprehensive models incorporates the explicit representation of realistic constraints of time and space, and the linkages among activities and travel both for an individual and across multiple people in a household[10]. Such advances allow activity-based models to more closely replicate actual traveler decisions than aggregate trip-based models[14], and hold greater sensitivity to transportation, investment policies as well as TDM strategies. Among activity-based model systems, there are various conceptual frameworks and behavioral assumptions proposed. These include constraints-based models, utility-maximizing (discrete choice) models, structural equation models, and computational simulation models[8].

2.2.1 Constraints-based models

The constraints-based models represent the first type of activity-based models. A concept of 'anchor points' was first recognized by the modelers, which represent fixed activities (such as home and work) that require mandatory presence temporarily and geographically. On that basis, the model suggested that people allocate available time to access and conduct activities between these 'known' locations, in a space-time context, known as time geography[15]. Space-time prisms were adopted, defining the area that an individual can reach within a certain time window between two known locations, confining the destination and timing choices that an individual can make[10].

As illustrated below, Figure 2.1 shows the concept of space-time prisms, where the potential path area projected on the space dimension represents the reachable area given space-temporary and mandatory stay constraints. Figure 2.2 depicts a daily activity pattern where vertical lines represent activities whereas diagonal lines are travel episodes.

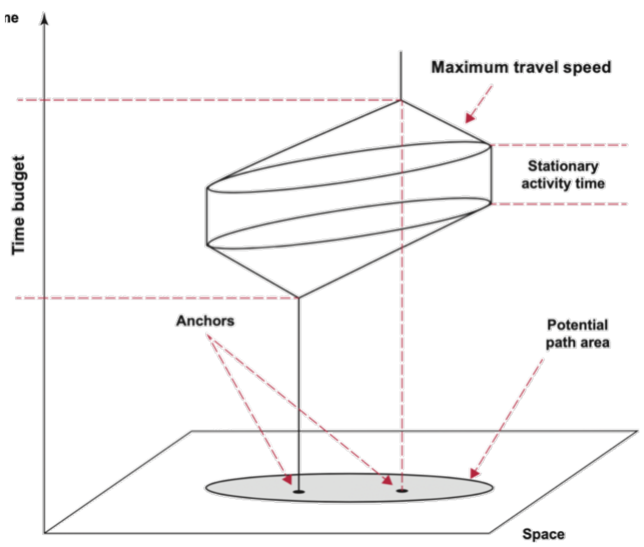


Figure 2-1 A space-time prism (taken from[15])

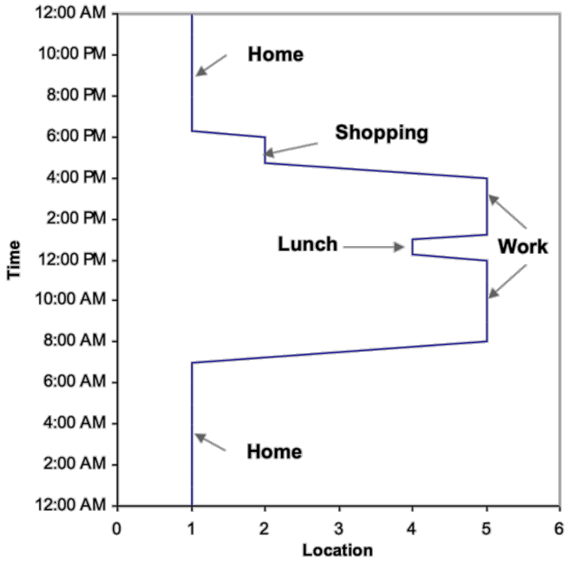


Figure 2-2 A daily activity pattern (adjusted based on[16])

As an early modeling approach, the aim of this model is not to predict individual or household activity patterns but rather to verify the feasibility of any possible travel itinerary (activity sequences) derived from combinatorial algorithms.

2.2.2 Utility maximizing (discrete choice) models

Compared to constraints-based models which emphasize the role of constraints in activity assignment, econometric (utility-maximizing) models focused on choice behaviors and described an individual's behavior in activity scheduling as a sequence of discrete choice problems. Empirically, utility maximization from an exhaustively determined feasible set of alternatives is the most frequently assumed protocol for modeling decisions.

As often the discrete choice models are extended as a nested logit model, e.g. Ben-Akiva and Bowman[14] developed a hierarchical structure for activity scheduling, in particular 'the daily activity schedule model', which became the very basis for many later models and empirical applications. They distinguished the key components involved in daily activity patterns as a multi-dimensional combination of primary activity, primary tour type and the number and purposes of secondary activities[17]. As illustrated in figure 2.3, a sequence of conditional choices comprises the nested logit model. Started with an activity pattern with choice of staying at home or travel, the nest is then followed with a system of conditional tours, which consists of four tiers: 1) primary tour time of day, 2) primary destination and mode, 3) secondary tour time of day and 4) secondary tour destination and mode. Activity types are categorized into home, work, school, and others.

Tour types can be categorized into:

- tour from home to work and back (h-w-h)
- tour with at least one additional stop for another activity (h-w-h+)
- tour with a work-based sub tour for another activity as well as any number (including zero) of additional stops for other activities (h-w + w-h)
- other tour categories such as mid-tour returns home, one with no additional activities (h-w-h-w-h) and another with one or more additional stops (h-w-h-w-h+).

Two purposes and three frequencies are distinguished for secondary tours: constrained and unconstrained purposes and 0, 1 and 2 or more secondary tours.

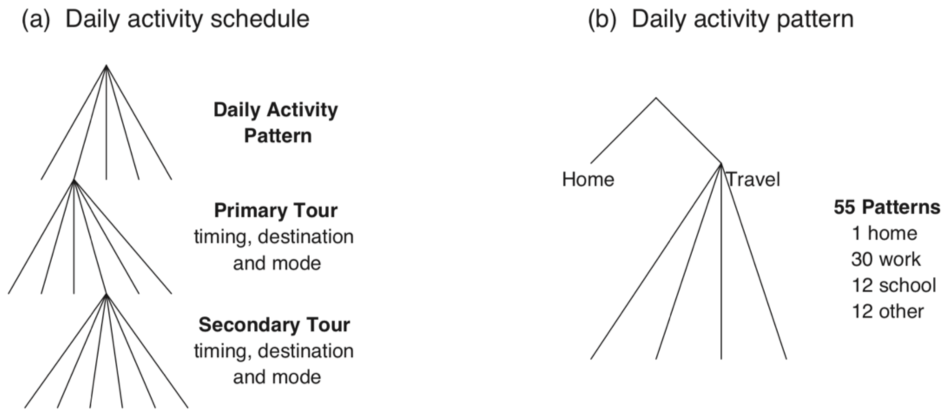


Figure 2-3 The daily activity schedule system and activity patterns (taken from [14])

A multinomial logistic model is often used to estimate the parameters in choice of tour time and destination, separately and conditionally for primary and secondary tours[18].

Another activity-based model, named 'The Prism-Constrained Activity Travel Simulator (PCATS)' was developed by Kitamura and Fujii[11]. This model differentiates the time of a day between open and blocked period. During the open periods, it incorporated space-time prisms to delineate the feasible area for travel and activities. Similarly, a nested logit model was used to predict activity-travel patterns, destinations, and mode choice. It also took into account the duration of activities, which is dependent on activity type, socio-demographics, time of day, location, prior and post-activity engagement. Available choice sets are generated by the computational process and activity choices are estimated in a probabilistic approach.

2.2.3 Summary

Trip-based models take individual person trips as the fundamental unit of analysis. Series of linked trips in a chain are modeled as separate, and spatial-temporal, as well as other plausible constraints like mode choice, are not considered. Although activity-based models share several similarities with traditional trip-based models, they incorporate several significant advances over four-step trip-based models. These advances include a more intuitive, consistent, and behaviorally realistic representation of travel than trip-based models, as well as the linkages among activities and travel both for a person and across multiple people in a household.

In activity-based models, rather than representing each trip as independent, for each traveler chains of trips (tours) are modeled as part of generating overall daily activity patterns. By functioning at the level of the individual traveler, activity-based models can represent greater variation across the population than aggregate trip-based models.

Due to the disaggregate nature, a key advantage of activity-based models is that they can incorporate new explanatory variables to represent a variety of policies and investments that are of interest to decision-makers. Effectiveness of such strategies and policies can be evaluated after implementation using a microsimulation framework, in which individuals and household choices are evaluated. Some key differences between the two models are illustrated in Table 2.1 as a complimentary comparison.

However, activity scheduling has not been of primary interest in empirical transport research. Criticisms include:

a) unrealistic behavioral assumptions, i.e. utility maximization in decision making by individuals
b) artificially restriction of activity scheduling to predefined choices
c) limitations on the timescale of analysis usually discretized to time-of-day periods
d) increased complexity and data requirements

As a result, there is still a lack of evidence that activity-based models indeed outperform four-step models. The need to supplement four-step models by activity-based models is still being questioned.

Table 2.1 A comparison between travel demand models

	Trip-based model	Activity-based model
Analysis unit	Independent trips	Activities in a day
Decision unit	Decision-maker isolated from a household context	Interactions between household members
Constraints	None	Space-time constraints, trip chaining, mode choice, etc.
Time dimension	None. Daily peak hour only.	Time-of-day variation, often discretized into periods
Resolution	Low, mostly aggregate level	High-resolution time and place
Spatial-temporal detail	Low-moderate	Moderate-high
Person/ household detail	Moderate	High
Policy sensitivity	Low	High

2.3 Travel demand forecasting of tourists

Despite the rapid progress of trip-based and activity-based models in the last several decades, it must be admitted that such models are mainly used in urban travel forecasting for ordinary travelers. There is limited work in literature discussing tourist behavior inside urban areas. Several differences should be acknowledged between modeling tourist and commuters:

2.3.1 Activity–travel pattern

As opposed to conventional travel demand forecasting where travel demand is derived from participation in daily activities, some of which might be fulfilled in various places, tourists' travel demand is derived from the desire to tour around (a set of) specific attractions inside a city.

There is evidence that for ordinary commuters, behavioral inertia is prevalent, and people tend to resist behavioral changes. In activity-based models, activities are sequentially connected in a continuous domain of time and space and are interrupted because of the existence of blocked periods (anchors). When modeling, a one-day schedule period is natural and common because of the daily rest period's regulating effect. Consequently, the daily activity pattern is described by a combination of primary and secondary activities from a predefined choice set. Considering interactions between household members and joint choices, modeler describes their activity scheduling process as to optimize the choice of destinations, starting and ending time as well as mode choices hierarchically and sequentially, usually with a nested logit model, where multiple access to a certain zone is allowed and rather common.

On the other hand, tourists have less behavioral inertia in comparison with individuals or households in the ordinary travel demand system. Compared with commuters who decide daily activity patterns before departure and from a fixed activity set, tourists optimize the activity themselves en route, evaluating the performance of location visited so far. To decide preferable destinations, tourists seek for more information and thus are more capable of discriminating among options to sort out a subjective optimum. Having multiple visits to the same place is rare unless for another purpose, e.g. sightseeing during the day and enjoy the food at night. Literature also prefers to depict tourists as rational decision-makers based on the fact that supplementary tools including travel books, trip planning software is easy to acquire. As a result, unlike ordinary travelers, tourists' behaviors are relatively simple (touring around attraction sites) while more dynamic due to the limited schedule to complete exploration. Moreover, for tourists with multiple-day stays, there might be completely different plans for different days, thus the time unit of analysis should also be the whole trip rather than a daily basis.

2.3.2 Partition of Traffic Analysis Zones

It is important to establish zone boundaries that are appropriate according to the purpose of the model.

For ordinary travel demand models, selecting appropriate TAZ boundaries is crucial and beneficial for quality control of demographics provided by the planning agency. Zones are usually aggregated to Census-defined units such as block groups, tracts, and census TAZs. For areas where partition might be flexible, boundaries are often drawn regarding topographic barriers such as large rivers or major railroad lines. Such procedures, especially in four-step models, are implemented for general zonal homogeneity (similar land use, density, socio-economic attributes, etc.) and trip generating potential.

On the other hand, it is hard to justify the adoption of a similar TAZ partition into tourist models. Tourists usually stay only several days due to limited schedule but might make significantly different travel plans across days. What's more, the massive inbound and outbound tourism movements keep replacing the individuals comprising the whole tourist population that is to be modeled. Finally, as sightseeing spots, as well as POIs related to tourism, contribute considerable attractions of visits, the presence and distribution of attraction areas should also represent a major concern in deciding the partition of TAZs.

2.3.3 Optimization principle

Researchers questioned the appropriateness of the optimization principle in that individuals have information that is partial and incomplete, and their decisions may not be internally coherent and consistently rational[19]. Moreover, Cullen and Godson[20] suggested individuals participate in activities in a context that is structured by his propensity and habits and thus form activity patterns. There is evidence that behavioral inertia is prevalent, and people tend to resist behavioral changes. Often represented as an extended discrete choice problem, the approaches elaborated above (*the daily activity schedule model* and PCATS) predict the activity and destination choices in a probabilistic way sequentially (usually in a nested logit framework).

Unlike ordinary travelers, literature prefers to depict tourists as rational decision-makers, who seek preferable destinations referring to supplementary tools including travel books, tourism websites, and trip planning software. Although the principle of optimization (e.g., minimization of adjustment costs, minimization of behavioral changes) has often been questioned in the case of ordinary travelers on a daily basis, tourists are believed to more or less behave towards an equilibrium where the 'satisfaction' is maximized. It is also revealed that with additional information, tourists are more capable of discriminating among options to sort out a perceptual optimum[21].

Our model adopts a similar temporal-spatial prism in the constrained-based model and utilized a utility-maximizing principle. Instead of using a nested logit model and predict activities sequentially, we describe individual tourist behavior through utility functions and see the decision of choosing a specific tour route as a process of solving an optimization problem.

To summary, our model for modeling tourists differs from the ordinary activity models in that:

1) In activity-based models, household members schedule daily activity patterns and decide pattern before departure. They are assumed to follow certain activity types from a fixed activity set (e.g. from 55 patterns defined in Ben Akiva's model[14]). More specifically, they optimize destination (TAZs) to visit, starting and ending time, mode choice, but usually not to optimizing the activity themselves;
2) In contrast, we assume tourists optimize the activity themselves dynamically, i.e. destinations to visit as well as activities to conduct in their decision-making process when making tours. Especially, they evaluate the satisfaction en route from the places visited so far, which will, in turn, affect the choice of next destination

2.4 Behavioral theory in Operations Research literature

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[21]. Factors that influence tourist movements identified in the literature generally include a set of destination characteristics, trip characteristics and a set of tourist characteristics[22, 23].

Operations research literature describing individual tourist behavior through utility functions sees the decision of choosing a specific tour route as a process of solving an optimization problem. As the problem studies the decision-making process of tourists, it is often named as tour route planning problem (TRPP) or tourist tour design problem (TTDP), referring to a route-planning problem for tourists interested in visiting multiple points of interest (POIs).

TTDP solvers derive daily tourist tours, i.e., ordered visits to POIs, which respect tourist constraints and POIs attributes. The main objective of the problem discussed is to select POIs that match tourist preferences, thereby maximizing tourist satisfaction, while taking into account a multitude of parameters and constraints (e.g., distances among POIs, visiting time required for each POI, POIs visiting days/hours, entrance fees, weather conditions) and respecting the time available for sightseeing on a daily basis. Figure 2.4 draws the main body of the algorithmic and operations research literature dealing with TTDP modeling and solving focuses on OP, TOP as well as their variants.

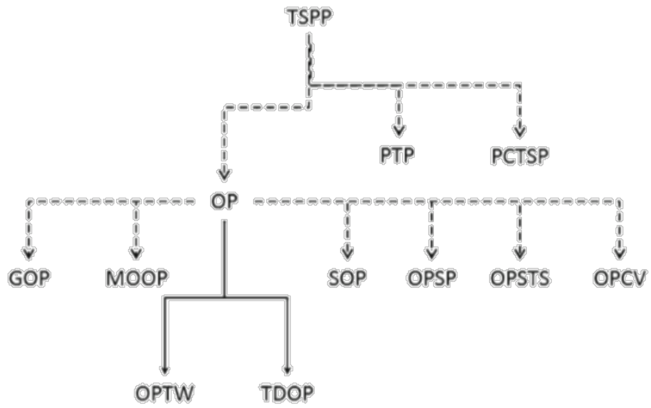


Figure 2-4 A hierarchy of TTDP and its variants

The TTDP can be modeled as a variant of the Orienteering Problem (OP), generalized from the TSP with additional goals that drive the collected profit to be maximized while maintaining the travel cost under a given constraint. As an NP-hard problem in combinatorial optimization, important in operations research, the traveling salesman problem (TSP) introduces a question where someone has to start from one city, find a tour with the shortest path in which all of the cities in the graph are to be visited exactly one time.

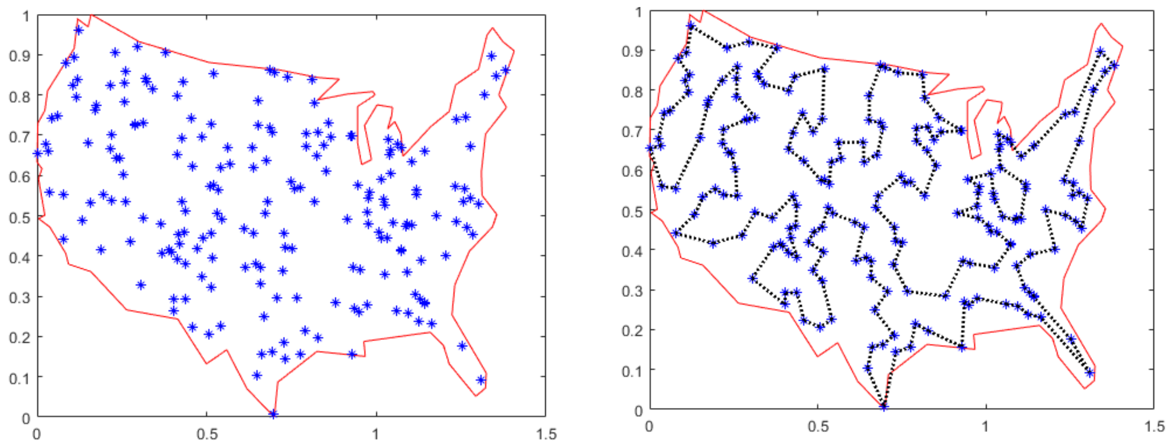


Figure 2-5 A solution of a large-scale TSP using exact approaches

TTDP has been a generic problem of customized tourist route generation which is mainly associated with the core functionality of personalized electronic tourist guides (PETs) and web/mobile tourist guides. Those personalized applications featuring route planning services typically offer following main functionalities: 1) a POI recommendation functionality that generates a list of preferable destinations for each different tourist profile; 2) a route generation functionality that employs an algorithm to generate personalized touring routes using the recommended POIs as well as other potential places; 3) a customization functionality which allows users to modify the generated personalized route (add/remove or reorder POIs) to better fit their preferences[24].

Recent studies showed a growing interest in the recommendation and route generation functionality among a broad range of topics covered by TTDP.

Specifically, studies on TTDP can be mainly classified into 2 fields:

1) Tourist preference evaluation and POI recommendation (Kinoshita et al., Huang Y et al., etc.)

2) Algorithms, heuristics for deriving an optimal visiting order (Vansteenwegen et al., Garcia et al., Oh, J. S. et al, Wu et al, etc.)

Additionally, the systems perspective like challenges for personal trip planners in online web/mobile applications are also studied by researchers in Geoinformatics and expert systems[25, 26, 27].

The algorithmic approaches for solving TTDP variants represent the most crowded field. Categories of solutions for TSP and its variants (OP, TTDP, EVRP, etc.) mainly include:

1) Exact algorithm: exact algorithms are algorithms that always solve an optimization problem to optimality in a finite amount of time. However, their time complexity increases exponentially respect to the dimensions of the problem so that they are only efficient when the dimension of the problem is not huge. A typical approach is branch-and-bound which is also used as a base for various heuristics;
2) Approximate approaches
3) Problem-dependent heuristics: they are problem-dependent and designed to be greedy, i.e. to take advantage of the specificity of their problems at hand. However, this nature also leads them to easily get trapped in a local optimum. Specifically, in the field of OP and TTDP, problem-dependent heuristics can be further divided into:
    - Tour building algorithms, where popular techniques include Nearest neighbor algorithm, Clarke and Wright Savings and Insertion algorithm;
    - Tour improvement algorithm, known as K-opt/Or-opt;
    - Integrated procedure. An example is to integrate the Insertion Algorithm with 2 or 3-opt and thus become a two-stage method.
4) Problem independent (meta) heuristics: These are problem-independent techniques. As such, they do not take advantage of any particularities of the problem and have a problem-independent search space. They may even accept a temporary worse solution during the optimum search process to explore hopefully better results. Sometimes the term metaheuristic is also used to refer to a high-level problem-independent algorithmic framework that provides a set of guidelines or strategies for a problem specific implementation of a heuristic optimization algorithm. Popular techniques include simulated-annealing, Genetic Algorithm, Tabu search, Ants routing and so on.

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[28,29,30,31,32]. With improvements in heuristics, complicated variables are introduced to simulate more realistic TTDP scenarios. Abbaspour et al[30] introduced a multimodal transportation network with time-dependent travel time, while Lu et al[28] developed a system for integration of tourist packages and single attractions with time-dependent interest levels.

Despite the advances above, applying the TRPP or TTDP requires first to understand the choice parameters. Although above optimization approach for modeling 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 behavior 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[33] summarized sightseeing facilities in Kyoto, Japan into the three categories “Downtown”, “Shrine and Temple” and “Natural beauty”. Yuichiro et al[34] 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[23] used factor analysis to reduce attraction categories into 5 dimensions (factors).

In our research, we follow the three-category approach and suggest attractions have different intrinsic utilities over these categories, which might consist of, for example, “Natural and Scenery”, “Cultural and Art”, “Leisure and Entertainment” etc. We evaluate both intrinsic utilities of POIs and tourists’ preferences for these dimensions.

Chapter 3 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…v_n }, and the origin and destination set S = {v_{n+1},v_{n+2},…,v_{n+s}}. The edge set E = {(v_i, v_j): v_i,v_j ∈ 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 tourist websites, 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, although a simple additive optimization principle for modeling 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 behavior 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 throughout 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.

Thus, in line with the above discussion, instead of making the total utility achieved by tourists additive to each 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 (3.1):

$$\max U_n | o, d = \max\ u_{oi_1} + \sum_{k=1}^{m_p - 1} \left(u_{ni_k}^P + u_{i_k i_{k+1}}^T \right) + u_{ni_m}^P + u_{i_m d}^T \tag{3.1}$$

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 (3.1) 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 traveling from *i* to *j*. These utilities can be further specified as follows:

3.1 Utility of travel

The utility of traveling on each edge is assumed to have a linear time and cost function as in (3.2) where $\alpha_1 = 1/\mathrm{VOT}$ (value of time) estimated and transfers the monetary cost (1 JPY) into time (min):

$$u_{ij}^T = -(t_{ij} + \alpha_1 c_{ij}) \tag{3.2}$$

3.2 Utility of visit

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:

$$u_{n,\,k}^P = \beta_1 \boldsymbol{P}_n^T * (\boldsymbol{U}_{i_k} \circ (1 - F(\boldsymbol{A}_{n,k}; k,\, \theta)) \tag{3.3}$$

$$\text{where } F(x; k,\, \theta) \sim \Gamma(k,\, \theta) \equiv \mathrm{Gamma}(k,\, \theta) = \int_0^x f(u; k,\, \theta) du = \frac{\gamma(k, \frac{x}{\theta})}{\Gamma(k)}$$

The '$\circ$' mark in (3.3) stands for the entry wise 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. The "discount factor" $F(x; k,\, \theta)$ is the cumulative distribution function of a gamma probability distribution where k and θ represent the shape of the diminishing marginal utility for visiting the next places. It is assumed to be negative since more cumulated satisfaction is reducing the benefit to visit another place. If the curve is very smooth, e.g. flat with a constant, then the travel history does not have any influence on a person's subsequent decision.

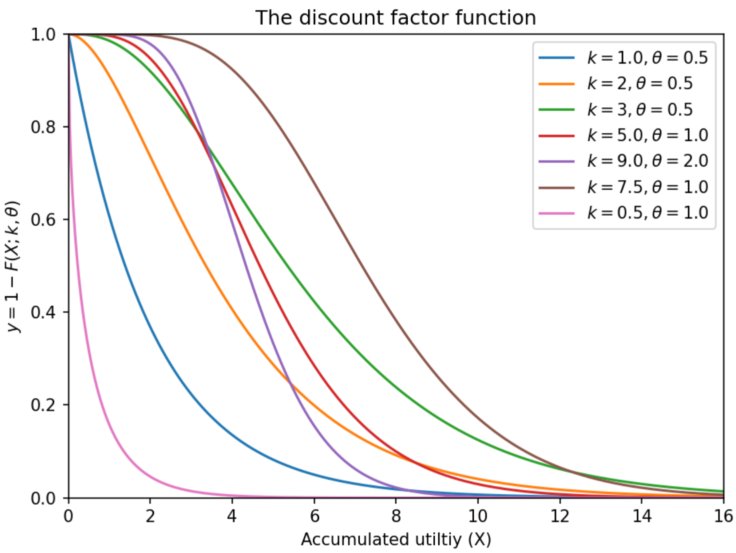


Figure 3-1 Function shapes at different combination of coefficients

Specifically, the reason for adopting a gamma form distribution is that it represents a universal and flexible form of the discount factor we hope to introduce. With the skewness, offset and sharpness coefficients that control the shape of such utility function, it can vary from the normal exponential form which drops sharply in the very beginning, to a rather general logistic form where the gradient reaches most negative in the middle.

Constraints for the problem are:

$$\sum_{i=1}^{n} x_{id} = \sum_{j=1}^{n} x_{oj} = 1 \tag{3.4}$$

$$\sum_{i=1}^{n-1} \sum_{j=2}^{n} x_{ij} \leq 1 (i \neq j) \tag{3.5}$$

$$\sum_{i=0}^{n} \sum_{j=1}^{n+1} x_{ij} t_{ij} + \sum_{i=1}^{n} \delta_i T_i \leq t_{end} - t_{start} \tag{3.6}$$

$$x_{ij} = \begin{cases} 1 \text{ if going from node } i \text{ to node } j; \\ 0 \text{ otherwise} \end{cases} \tag{3.7}$$

$$\delta_i = \begin{cases} 1 \text{ if node } i \text{ is visited;} \\ 0 \text{ otherwise} \end{cases} \tag{3.8}$$

Constraints (3.4) ensure that the tour route starts from the origin and ends at the destination node d. Constraint (3.5) implies that no attractions are visited more than once, and tourists could travel at most once among two nodes. Constraint (3.6) guarantees that the time budget of the tour is satisfied, whereas node number 0 and n+1 represent for origin and destination respectively. Constraints (3.7) and (3.8) define binary dummy variables.

To summarize, we model tourists' behaviors in making tours given a time budget, taking consideration of key features that utility (satisfaction) is achieved in multiple dimensions, and people encounter a diminishing marginal utility along with more utilities being achieved. Since the time-dependent variation of node utilities and transport network properties do not belong to the main focus of our research, the time dimension is omitted to simplify the formulation and reduce computational cost. Constraints on open and close time of attractions are also released since they do not necessarily improve our model accuracy, we initially dismiss those constraints and leave them to the future work.

Chapter 4　Data development

4.1 Overview

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, gender, occupation, ownership of vehicles, home city, etc.;

b) tour related attributes such as travel purpose, schedule, travel group, frequency of visiting Kyoto city, comments and discontents, etc.;

c) a trip diary of detailed trip chains that consists of destinations, travel time and mode choice.

Table 4.1 Details of the Kyoto city tourism survey

Date	Nov.19 (Sat.) and 25 (Sun.), 2006
Survey method	Hand out manually, return by mail
Distribution locations	29 main attraction spots and 6 train terminals
Amount	Around 20,000 copies distributed
Survey items	Subject demographics, trip-related attributes (day-of-schedule, travel frequency), trip chain, comments and advice about Kyoto city, etc.
Respondents	3,456 (recover rate: 18%)

There were about 3,400 valid questionnaires received. We note that there are also two complimentary PT surveys conducted in the 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. Figure 4.1 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.

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

Figure 4-1 An example of observed trip chain from the travel diary

As shown in figure 4.2, the tour prediction module utilizes a TTDP heuristics to predict possible tours under each set of behavioral parameters. In line with our model formulation, tourists' demographics and trip-related attributes are processed in this chapter to prepare the data needed for tour prediction, including origin, destination, starting and ending time as well as the tourist preference, which is one of the key features in our behavioral model.

We suggest attraction utilities, as well as the tourists' preference, 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. Accordingly, 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 throughout his tour.

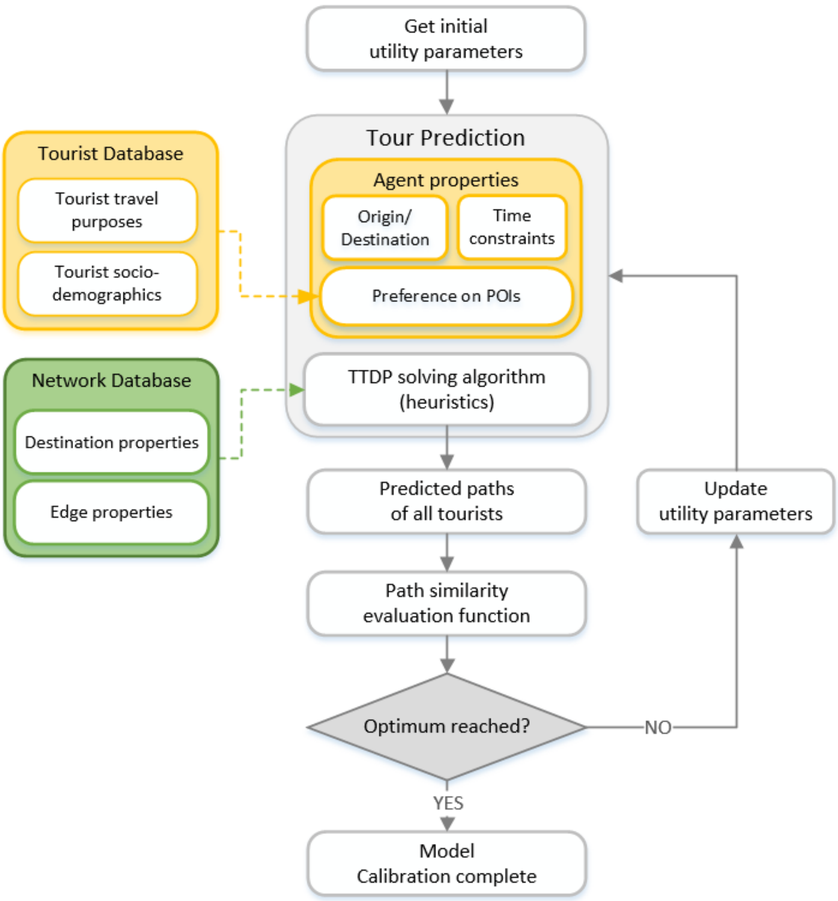


Figure 4-2 Data preparation in the model calibration framework

As we mainly focus on the behavioral model, the intrinsic utilities of attractions and tourists' preferences 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 querying Google Maps API and are averaged from different periods throughout a day.

4.2 Tourist 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 the 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 4.3 Choice candidates for travel purposes

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

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 4.3. 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[35]. Figure 4.3 provides an image of 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.

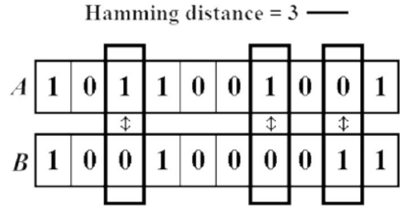


Figure 4-3 An illustration of Hamming distance

Details of the clustering result are illustrated in table 4.4 below.

Table 4.4 Clustering results

Cluster #	Choice of purposes	Interpretation	# of observations	Proportion
a	1-6	red leave & temple shrines	668	19.3%
b	1-6-8	red leave & temple shrines & gourmet	2142	62.0%
c	5	leisure activities	646	18.7%

The clustering result revealed a dominant proportion of tourists with interest mainly in red leave and temple shrines. This could somewhat be explained by the fact that the survey was conducted in November, during the foliage season when Kyoto city is gorgeous with its temples and relics decorated with lovely maple tree leaves.

With revealed clustering result regarding respondents' travel purpose, an ideal approach is to predict the tourists' preferences by obtaining information such as the socio-demographic 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 clusters using multinomial logit regression, taking the first cluster a) 1-6 as reference. The probability of belonging to each cluster is taken as the weight under each preference type, which adds up to 1. Such a vector is used as a preference vector for each tourist.

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 4.4 shows the variables and their effects on determining the probability of being either one of the two other clusters in comparison with the default cluster: a) 1-6.

Concretely, results indicate 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 center during the tourist season, especially during the autumn foliage season and cherry blossoms season which Japan is famous for.

Besides, 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. Moreover, for the people with less budget on food, the probability of being in cluster B or C will be smaller, as both gourmet and leisure activities can be seen as extra expenses apart from the main purpose of sightseeing.

Table 4.5 Explanatory variable effects (with significant contributes)

Cluster		(b) 1-6-8			(c) 5		
Variable	Level	Parameter	Std. Error	Sig.	Parameter	Std. Error	Sig.
Intercept		-1.797	1.048	0.086	-17.506	6295.287	0.998
Travel group composition							
	Single	-0.710*	0.349	0.042	0.242	0.273	0.375
	Couple	0.112	0.287	0.697	-0.723**	0.260	0.005
	Family	0.196	0.289	0.497	-0.199	0.246	0.419
	Friends/Colleague	0.656*	0.283	0.021	-0.376	0.252	0.135
	Sightseeing group	0			0		
Visit frequency							
	first time	-0.247	0.215	1.330	-0.598**	0.226	0.008
	2~3 times in 5 years	-0.144	0.191	0.569	-0.421*	0.182	0.021
	every year	-0.190	0.205	0.862	-0.775***	0.207	0.000
	2~3 times per year	-0.110	0.185	0.350	-0.449**	0.163	0.006
	> 4 times per year	0			0		
Living in Kyoto dummy							
	yes	0.502	0.810	0.384	-1.162*	0.639	0.069
	no	0			0		
Length of trip							
	day trip	-0.625*	0.246	0.011	0.186	0.308	0.366
	2-day trip	-0.287	0.213	0.178	0.179	0.287	0.390
	3 days trip	-0.408*	0.217	0.060	-0.209	0.301	0.482
	4 days and more	0			0		
With children dummy							
	no	1.217***	0.347	0.000	-0.138	0.233	0.351
	yes	0			0		
Budget on food (Japanese yen)							
	0-2k	-1.329***	0.274	0.000	-0.789*	0.313	0.012
	2-5k	-0.709**	0.270	0.009	-0.695*	0.314	0.027
	5-8k	-0.148	0.300	0.620	-0.454	0.361	0.208
	8-15k	0.281	0.318	0.377	-0.091	0.385	0.814
	15k and above	0			0		
Model Fitting Pseudo R-Square		a. Cox and Snell: 0.181; b. McFadden: 0.110					
Sample size		2818 valid, 638 missing					

Significance level: *0.05 **0.01 *0.001

Finally, since the multinomial 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.

4.3 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 4.6, which may include multiple different types of POIs.

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. Besides, 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 the following metrics.

Figure 4-4 Example of attraction areas

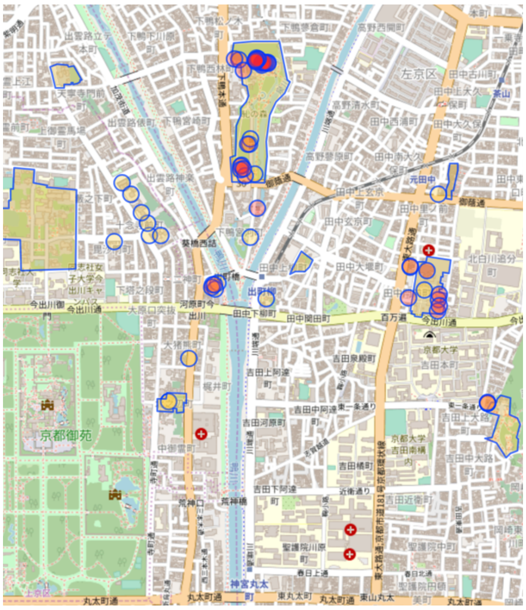


Figure 4-5 Parsed result of "temple or shrines"

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 the temple and shrines are evaluated when assessing the score of an attraction in this dimension. Specifically, we give weights to each POI based on the 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. The number of ordinary restaurants, bars, and pubs, as well as high-end restaurants, are calculated by Google POI search.

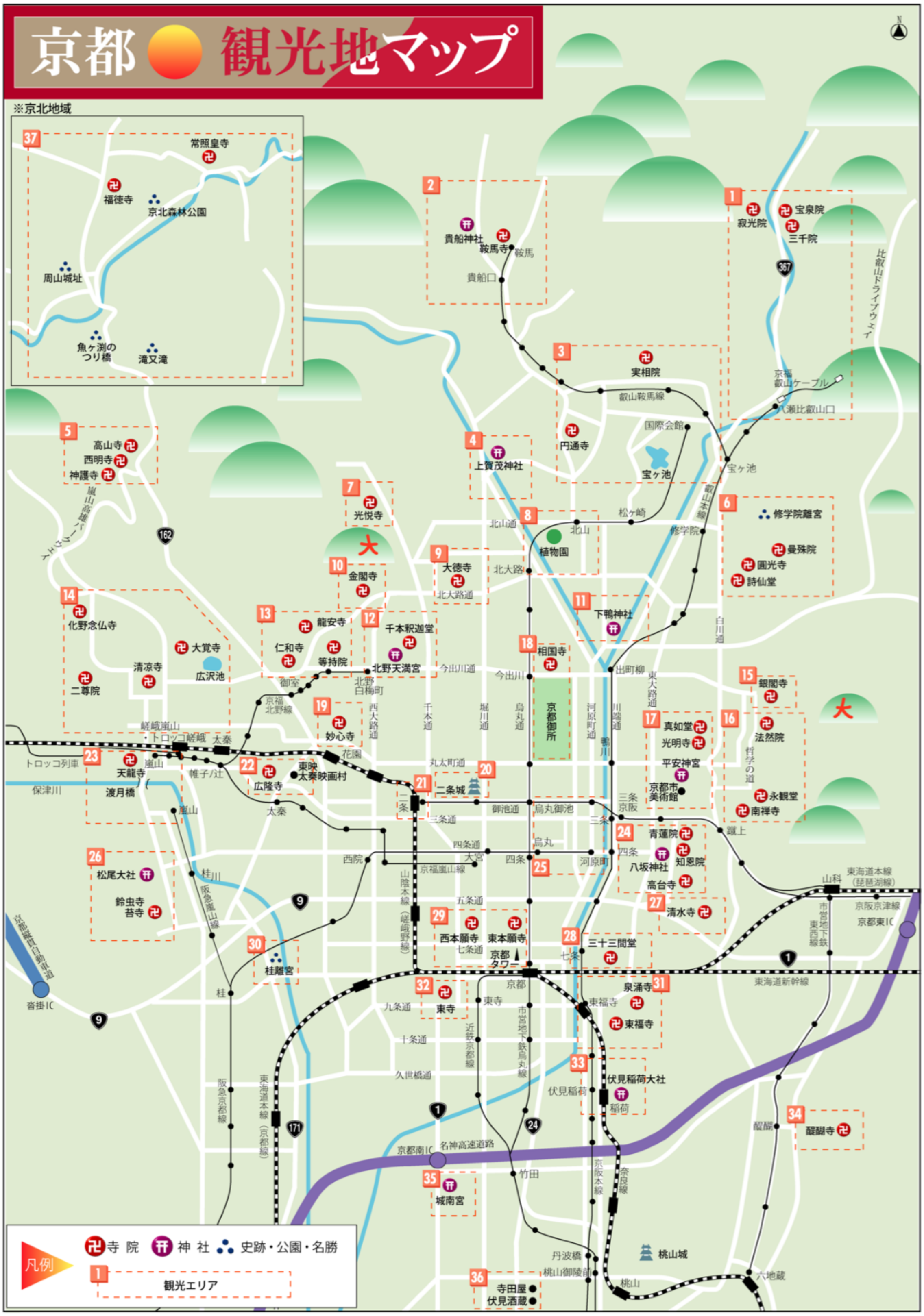
d) Leisure activities

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 the above scores upon different categories.

Finally, the tourists' preference vectors and attraction intrinsic utilities are normalized such that all elements in vector are scaled to have a value between 0 and 1. Figure 4.7 illustrates the attractions' intrinsic utilities in bars stacked by values in the three dimensions respectively.

4.4 Network edge properties

We then linked the attraction areas with arcs to each other, which altogether comprises of a complete and undirected graph network. To evaluate the attributes such as time, distance as well as the cost for traveling on each edge, we utilized the Google Maps services: Direction API for evaluating the mode-specific (which is transit for the current model) travel time, distance and transit fare between any two areas. We defined the nodes for generating and absorbing traffics as area centers or main transit entrances. Matrices of travel time, distance and fare are measured and averaged from different periods (7:30, 12:30 and 17:30) throughout a day.



1 大原・八瀬方面
2 鞍馬方面
3 宝ヶ池方面
4 上賀茂神社周辺
5 高雄方面
6 修学院・詩仙堂周辺
7 光悦寺周辺
8 北山通周辺
9 大徳寺周辺
10 金閣寺周辺
11 下鴨神社周辺
12 北野天満宮周辺
13 衣笠・御室方面
14 嵯峨野方面
15 銀閣寺周辺
16 哲学の道周辺
17 平安神宮周辺
18 京都御所周辺
19 花園方面
20 二条城周辺
21 二条駅周辺
22 太秦方面
23 嵐山方面
24 祇園方面
25 河原町・新京極方面
26 松尾大社周辺
27 清水寺周辺
28 三十三間堂周辺
29 京都駅周辺
30 桂離宮周辺
31 東福寺周辺
32 東寺周辺
33 伏見稲荷大社周辺
34 醍醐寺周辺
35 城南宮周辺
36 伏見方面
37 京北方面

Figure 4-6 Definition of attraction areas in the survey

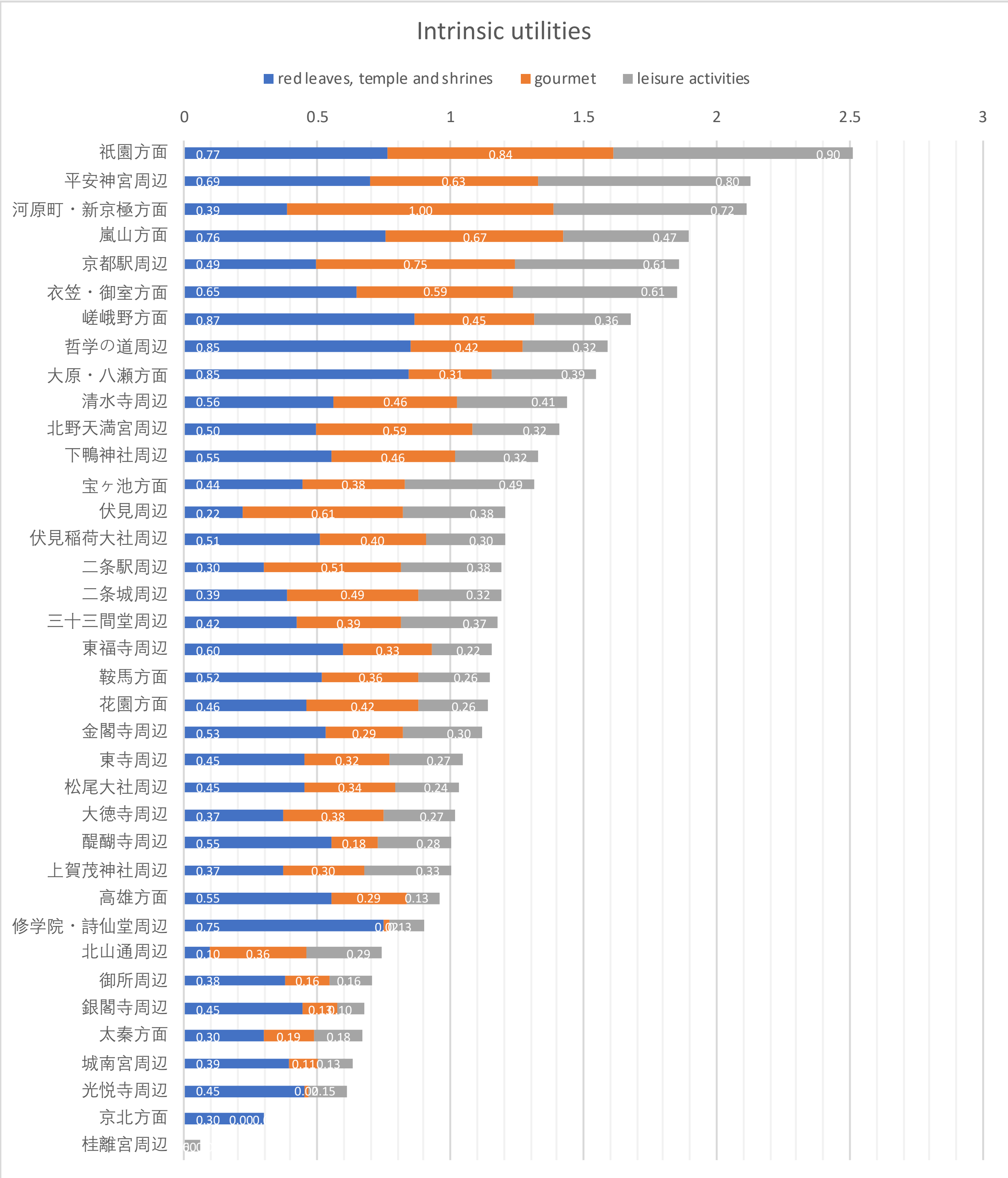


Figure 4-7 Obtained intrinsic utilities of attraction areas

Chapter 5 Methodology

5.1 Overview

In the last chapter, tourists' attributes are extracted and estimated to present the constraints and other data needed as agent properties for the simulation. Meanwhile, node intrinsic utilities and edge properties are also calibrated and taken as input. We utilized a framework illustrated in figure 5.1 to calibrate the model proposed.

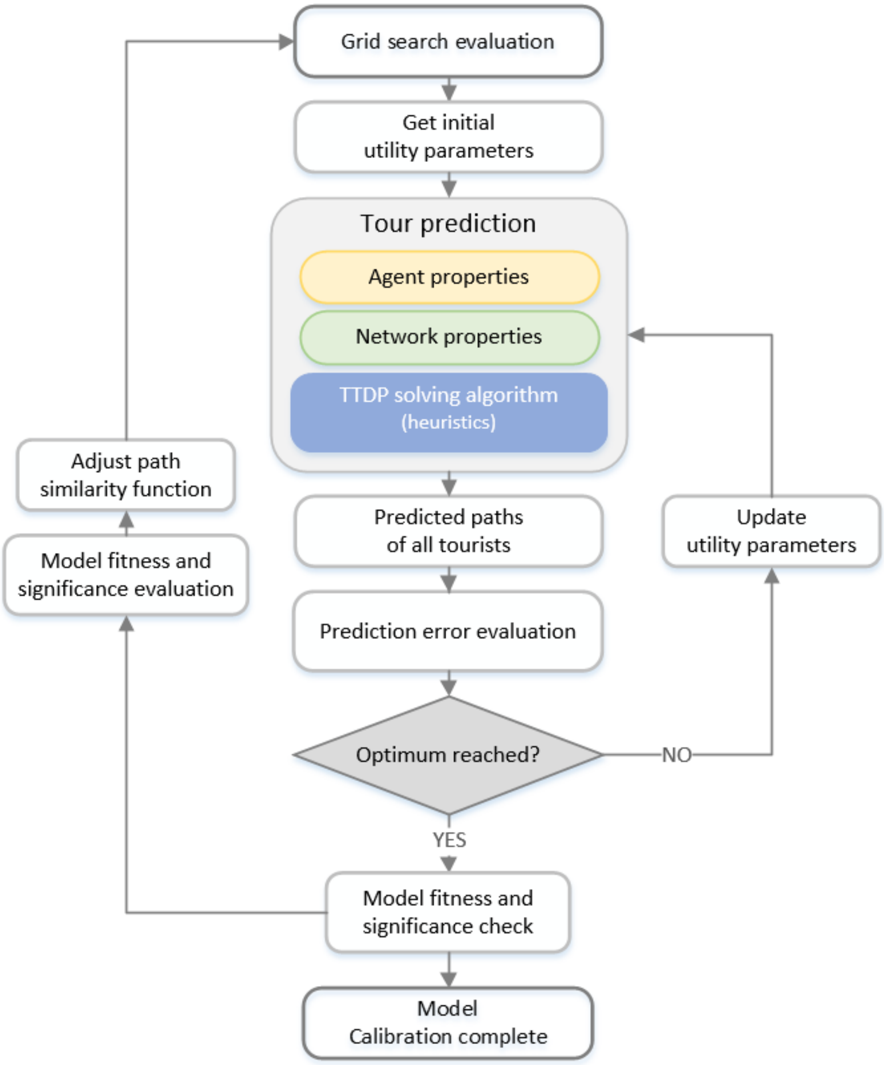


Figure 5-1 the parameter calibration framework

There are two loops in the parameter update framework:

1) **The inner loop**:

   Embedded in the tour prediction module, this inner loop enumerates all tourists from the survey and utilizes a problem-specific heuristic to predict possible tours. It compares the predicted paths against observed ones and finally sums up the prediction errors at each person.

2) **The outer loop**:

The outer loop generates parameters that describe tourist's behavior with respect to a set of observed tours. 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. Metaheuristics are used for this to obtain the best fitting parameters.

This model allows us to simulate tourism management strategies under various scenarios. Sections in this chapter explain the key modules of the parameter calibration process.

5.2 Solution heuristic: a 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[36].

Other than exact solutions, numerous heuristic rules were developed for solving optimization problems like Traveling Salesman Problem (TSP), Orienteering Problem (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 developed and examined the performance of two heuristics in the pursuit of the global optimum.

Algorithm 1: the OP based heuristic in pseudo code

```
Input: edge/node database, origin/destination, preference, etc.
Step.1: Initialization
        perform Insertion
        set record and calculate deviation   <- to accept an inferior solution
Output.1: Path_op, Path_nop (set)
Step.2: Improvement
        A loop:  (reinitialization)
               B loop:  (improvement)
                     perform 2-point exchange
                     perform 1-point movement
                     perform 2-opt clean up
                     if Solution better than bestFoundSolution:
                           update bestFoundSolution
                           set record and calculate deviation
                     else: end B loop
               end B loop
               reinitialization
           end A loop
Output: bestFoundSolution
```

Figure 5-2 The OP-based heuristic in pseudo code (defined in[37])

The first heuristic is a modified algorithm based on the OP solving heuristic by Chao[37] 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.

Most TTDP solving algorithms base on heuristics for OP. As our problem introduces a diminishing marginal utility for the satisfaction gained in history, ordinary OP-based solution may not guarantee its performance in our problem. Thus, we adopted a modified heuristic based on a more intuitive heuristic: the Iterated Local Search and performed a benchmark test between the ILS and ordinary OP-based algorithm.

This basic form of iterated local search heuristic consists of two steps, the Insertion, and the Shake. The insertion step aims to add one after another all possible visits on a tour while simultaneously respecting the time budget available. The insertion step is finished when there are no more possible insertions available, yielding a local optimum. After that, a shake step is introduced to escape from local optima. In this step, a certain number of visits will be deleted from the solution. The number of visits that gets deleted is denoted by R. And the starting node for the deletion is denoted by S. After the visits are deleted from the trip, the disconnected visits of S and $S+R$ are re-connected by re-insertion in the pursuit of optimality. A general flow of the heuristic is illustrated in pseudo-code (Fig.5). For more details of the ILS heuristic, see Σανιδάς, Γ. ,2019[38].

Algorithm 2: the ILS heuristic in pseudo code

```
Input: edge/node database, origin/destination, preference, etc.
S=R=1
InitialSolution = [o, RandomVisit, d]
while NoImprove < 50
        while not LocalOptima:
                perfom Insert;
        end
        if currentSolution better than bestFoundSolution:
                update bestFoundSolution
                R=1; NoImprove=0
        else:
                NoImprove = NoImprove + 1
        S = random(1 to length(Solution)-2);
        R = random(1 to length(Solution)-S);
        perfom Shake
end
return bestFoundSolution
```

Figure 5-3 The ILS heuristic in pseudo code (defined in[38])

We compared the two heuristic approaches in terms of their execution time (efficiency) as well as the quality of solution (scores) to see how close the solutions are to the global optimum. As the ILS heuristic has the opportunity to achieve the global optimum, we run multiple iterations and use the best solution found to approximate the global optimum. The results are illustrated in Fig. 5.2 and 5.3.

As revealed, the steps ('exchange', 'improvement' and 2-opt 'clean up') in the OP based heuristic are well-tuned to reduce execution time, but the heuristic yields relatively inferior solutions than the Iterated Local Search. It is because the algorithm is more problem-specific and did not fit well for approaching the optimal defined in our problem. In comparison, the solution time of ILS grows faster than the OP over time, probably due to the Shake step which occurs with fixed iterations each time a better solution is found. Still, it is more sensitive to the constraints and calculates better solutions every time the time budget constraint is relaxed.

Since we presume an optimization principle among the tourists who tend to maximize their utilities, we adopt the ILS algorithm which is more likely to find the global optimum to predict most possible tours people would take under given behavioral parameters.

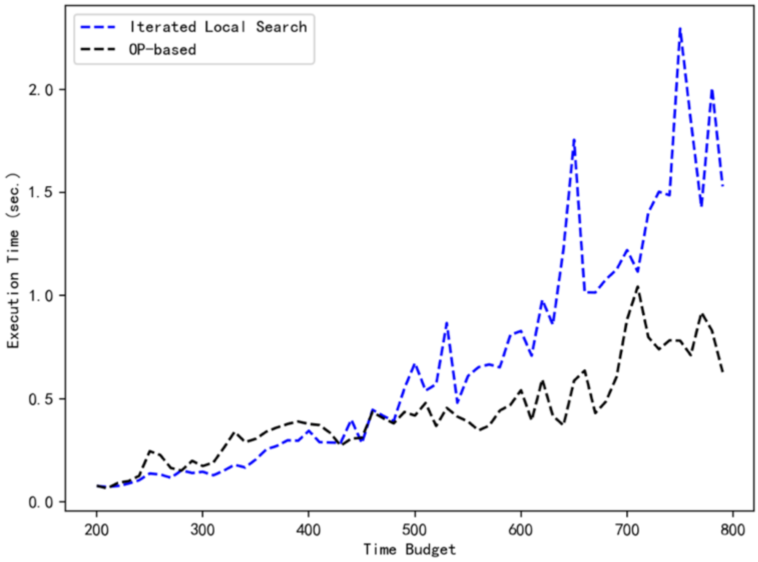


Figure 5-4 Execution time test

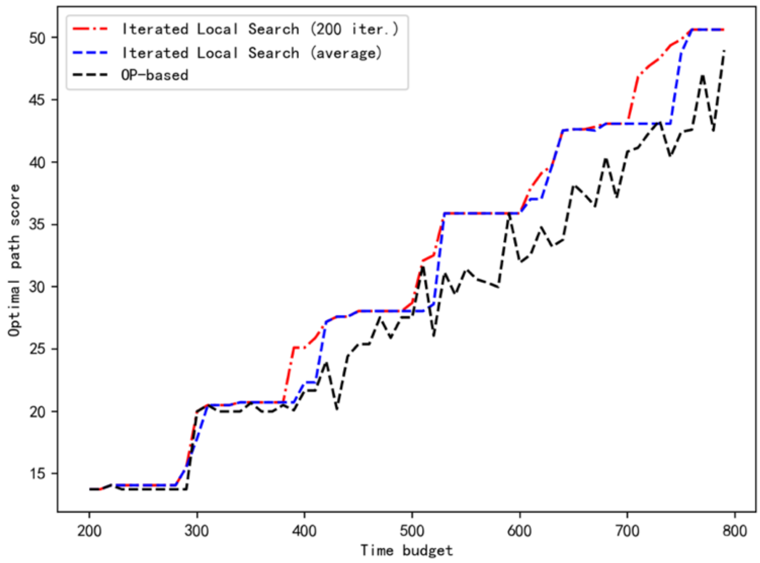


Figure 5-5 Optimum approximation test

5.3 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 $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ that minimize 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 the pairwise similarity between sequences[39]. Metrics like the longest common subsequence (LCSS) have also been applied in the literature to estimate the similarity of tourist movement sequences and as patterns for grouping travelers[40]. Since the traveler may slightly change the order of visit between several neighboring nodes, the similarity metric needs to be robust to noise. We suggest therefore a better way is to combine the assessment process with geographic interpretation, which not only measures the number of different entries among the two sequences but also to what degree the entries are different.

Let m_p be the number of visited points in the predicted path and $\hat{m}_p$ 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. A possible metric for determining the difference D_p between two paths can be computed as follows:

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

To measure to what degree are two sequences similar, we exploit the Levenshtein Distance, one of the best-known string metrics widely used in areas like computer science, as a similarity metric or a measure for the "distance" of strings[41]. The algorithm defines three basic editing operations to transform one sequence into the other, the Insertion cost, Deletion cost, and Substitution cost. Although it is most common to set the cost of all three operations to 1, we can assign different weights or costs to each editing operation. A common form of the Levenshtein Distance is formulated in (5).

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & if \min(i,j) = 0, \\ min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_i)} \end{cases} & \text{otherwise}. \end{cases} \tag{5}$$

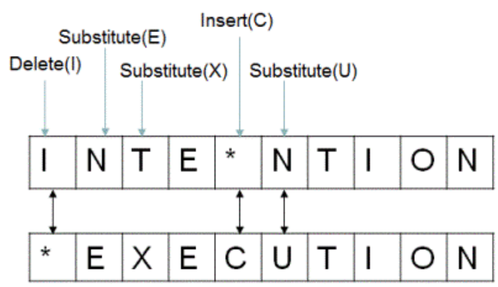


Figure 5-6 Three types of operations in Levenshtein Distance

1) Substitution cost:

In many cases, travelers tend to choose between "alternative" destinations that belong to the same category, are close to each other, and are similar in terms of the utility to be obtained, while keeping the additional travel impedance within acceptable limits. Intuitively, the penalty for replacing a destination to another should depend on the spatial difference between them. Therefore, we define the substitution cost as the geographical distance between the center nodes of any two attraction areas.

2) Insertion cost / Deletion cost:

Although the substitution cost can be calculated intuitively by referring to the two nodes that are replaced, determining the insertion and deletion costs will be tricky because the program will not know which characters to refer to for the insertion cost. Especially in the cases when the insertion is at the beginning or end of the sequence. Because the algorithm cannot dynamically determine the two nodes to be referenced for insertion, we modified the cost by fixing the reference 'node' as the center of weight of the observed route, and compute the insertion cost as the difference (offset) between the inserted node and the center of the observed route, as we hypothesize that it is more likely to add or drop a visit near the middle of the nodes visited than at the edges.

Finally, we implemented the algorithm through bottom-up dynamic programming. We determine the difference between two visited routes/sequences that takes into account geographic interpretation by calculating the number of substitutions, insertions, and deletions needed to transform one string into another one.

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 parameters.

Algorithm 3: a Modified Levenshtein Distance in pseudo code

```
Input: strings (s, t), coordinates and distance matrix
rows = length(s)+1
cols = length(t)+1
dist = matrix of zeros (rows*cols)
for i in range(1, rows):
    dist[i][0] = i
for i in range(1, cols):
    dist[0][i] = i
for col in range(1, cols):
    for row in range(1, rows):
        dist[row][col] = min(dist[row-1][col] + del_cost(s[row]),
                             dist[row][col-1] + ins_cost(t[col]),
                         dist[row-1][col-1] + sub_cost(s[row], t[col]))
return dist[row][col]

*comment:
a. del_cost(x) / ins_cost(x):
    cost = geographical_distance(center_of_observed, x)
b. sub_cost(x, y):
    if x == y:
        cost = 0
    else:
        cost = geographical_distance(x, y)
```

5.4 Behavioral model estimation

To evaluate how accurate each set of parameters describes the actual behavior of tourists when making tours, the program executes the "inner loop" shown in Figure 5.1, checks all tourists, and accumulates the prediction errors between the observed and estimated paths. After enumerating all tourists, the cumulative total error is taken as the fitness (cost) of each set of parameters.

During the prediction, each tourist is modeled to optimize the activities by picking a route with the highest utility among the set of alternative paths. After that, the objective function sums up the difference with geographic interpretation between the observed predicted paths, which eventually comprises the objective value of each set of parameters. Therefore, it does not appear to have a closed-form formulation between the fitness (objective value) and the coefficients representing expected utilities and hence does not seems to be analytical gradients for the objective function. Thus, using an off-the-shelf solving algorithm like gradient descent will not be feasible.

To overcome the flaws in the nature of the model, we adopted a GA like meta-heuristic to search for the optimal solutions over generations. The initial solutions (first generation) for further optimization is generated from a “grid search”: before initializing the optimum search, we first sample and compute the sample scores in the solution space at certain intervals to roughly grasp what the solution space looks like. Specifically, we adopted a half-logarithmic increment to sample from a reasonable range for each utility coefficient, e.g. from 0.01- 10^2 resulting in 7 values 0.01, 0.03, 0.1, 0.3, 1, 3, 10. Grid points are determined by an exhaustive permutation of all the values. As a result, this step allows us to see if ‘peaks’ exist and hopefully the ‘optimal’ solution will be unique.

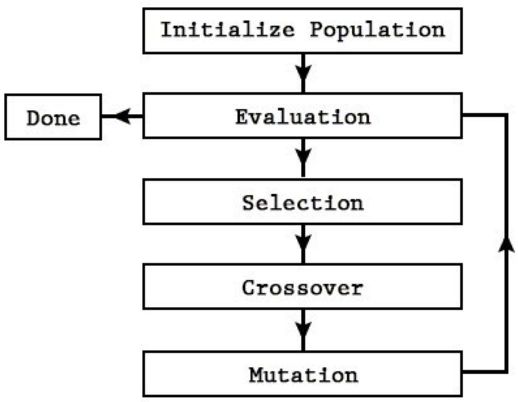


Figure 5-7 A flow chart of the Genetic Algorithm

After that, parameters with relatively low prediction penalty are taken as the first population that will reproduce offspring through a loop of three fundamental phases: selection, mating/crossover, and mutation. Specifically, because in our problem each gene (individual) representing the behavioral parameters is encoded as a vector of real numbers (in genetic algorithms individuals should be encoded as integers), we have adopted an Evolutionary Strategy framework that accepts individuals with real number entries. Scores (fitness) of parameters are calculated by taking the inverse of the prediction error powered to a large number, to increase the selection probability of a slightly better solution.

Through repetitions and iterations, a set of parameters is derived that best describes all tourists' behavior in the model eventually. 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 overall evaluation finishes in an acceptable time as we programed a fast TTDP solving algorithm and adopted multi-processing to speed up the optimum search.

5.5 Solution fitness and confidence

To evaluate the goodness of fit and how confident we are about the model estimated, we firstly adopted McFadden's Pseudo Rho-squares for the goodness of fit which is defined in (6), where $\hat{L}(M_{Full})$ is the log-likelihood value for the fitted model and $\hat{L}\left(M_{Intercept}\right)$ is the log-likelihood for the null model which includes only an intercept as a predictor.

$$\text{McFadden's } R^2 = 1 - \frac{\ln \hat{L}\,(M_{Full})}{\ln \hat{L}\left(M_{Intercept}\right)} \tag{6}$$

$$\text{Model's } R^2 = 1 - \frac{Q(\boldsymbol{\theta} mod)}{Q(\boldsymbol{\theta} null)} \tag{7}$$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\mathrm{s.e.}\left(\hat{\beta}\right)} \tag{8}$$

We adopted the idea of comparing the error explained by the prediction model with the error in the null case, and modified the calculation method of pseudo R-square as shown in formula (7), where $Q(\boldsymbol{\theta}) = \sum_{i=1}^{N} D_p(w_i, \boldsymbol{\theta})$ which sums up the prediction error between estimated and observed routes of all tourists under each behavioral principle. The $Q(\boldsymbol{\theta} mod)$ as well as $Q(\boldsymbol{\theta} null)$ represent the model prediction error under the estimated parameters and the null case respectively. As a result, the aforementioned pseudo R-squared will explain how much system error the model explains, and thus represent the goodness of fit of the estimated model.

Apart from the model fitness, we calculate the t-statistics (8) of the current solution which tells how confident we are about the model estimated. Specifically, the t-statistics calculates the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error. And by default, statistical packages report t-statistic with $\beta_0 = 0$.

$$\begin{aligned} \mathrm{var}\left(\hat{\beta}\right) &= \left[I\left(\hat{\beta}\right)\right]^{-1} \\ &= \left(-E\left[H\left(\hat{\beta}\right)\right]\right)^{-1} \\ &= \left(-E\left[\frac{\partial^2 \ln L(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'}\right]\right)^{-1} \end{aligned} \tag{9}$$

We adopted formula (9) in maximum likelihood estimation using which we calculate the standard errors of the estimator, $\hat{\beta}$ by taking the square roots of the diagonal terms in the variance-covariance matrix. In (10), the variance of an ML estimator $\hat{\beta}$, is calculated by the inverse of the Information matrix $I\left(\hat{\beta}\right)$, which is the negative of the expected value of the Hessian matrix. As our model does not adopt a probabilistic approach, e.g. choice models in estimating the parameters, the Hessian matrix will be deterministic and thus we take the numerical Hessian of parameters for calculating the t-statistics.

Finally, the combination of the pseudo R-square and pseudo t value will represent the goodness of fit as well as the significance of the current solution, which in other words is the confidence of the estimated model. We utilized these two indicators to prove the performance of the model we calibrated to justify the results in the simulation of TDM strategies.

5.6 Summary

All the aforementioned modules comprise the structure for estimating the behavioral model parameters.

Usually, in discrete choice problems, model parameters are estimated by distinguishing between the probability of choosing different paths and maximizing the probability of choosing the observed one, e.g. in maximum likelihood estimation, the logarithm of the probability for choosing the "right" path is maximized. However, since we describe tourist behavior as a non-discrete choice problem, there is no metric such as the probability to distinguish between alternatives other than the path identical to the observed one. In other words, predicting either the right or "wrong" path does not measure how close we are to the ground truth. Thus, finding the criteria to determine the prediction error becomes tricky.

In the model calibration, we adopt different methods to evaluate the error between the predicted and observed path. There are different forms of penalty functions depending on which error we emphasize. For example, we can emphasize the geographical error, in which case tourists behave conservatively by visiting only sights near the city center. As a result, they would rather stay where they are than visit places that might be "wrong". Similarly, we can emphasize the cost of dropping or adding new visits compared to the observed routes to predict the correct number of visits.

Therefore, there is no such problem as "right or wrong". Instead, we are just making trade-offs by adjusting the criteria of prediction errors, based on what type of travel behavior we want the model to reflect, which kind of conclusion we want to draw, or what objective we want to achieve.

Chapter 6 Model estimation and discussion

6.1 Model estimation result

We update the parameters and estimate the model using an iterative Genetic Algorithm based framework elaborated in Chapter 5. Figure 6.1 shows the parameter update process in 200 iterations, in which the y-axis is the score (fitness) of the best set of parameters in each iteration, which is calculated by taking the inverse of the prediction error powered to a large number as formulated in (6.1).

$$score(\boldsymbol{\theta}) = \left(10^4 \times \frac{1}{Q(\boldsymbol{\theta})}\right)^{20} \tag{6.1}$$

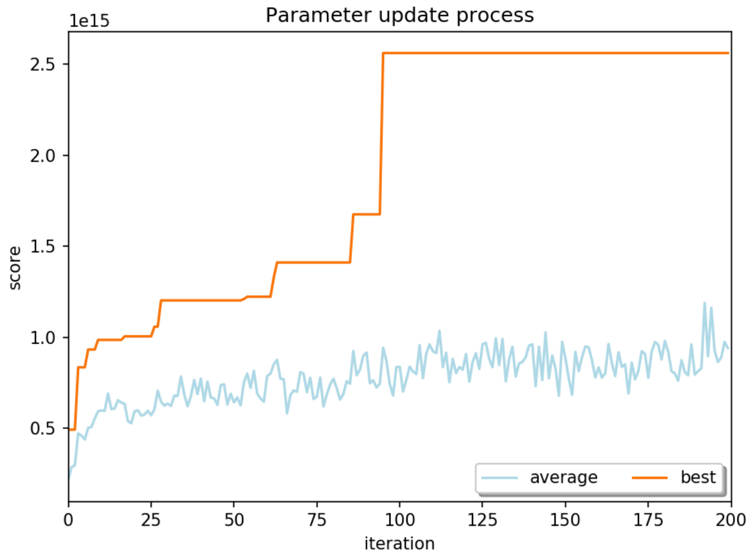


Figure 6-1 Model calibration process (in 200 iterations)

The best set of parameters having a maximum score was found near the half of the iterations. At the same time, after the random search started, the average score of each generation also showed an upward trend but will not converge due to the random search nature.

It should be noted, that the model estimation does not run a single iteration but several iterations in which we update the metric for assessing prediction errors depending on what kind of error we want to emphasize, by looking at the travel patterns displayed under the estimated parameters. Specifically, we adjusted the penalty function to emphasize the ability of the model to predict the correct number of visits, while maintaining the geographical interpretation between two paths.

Finally, we calibrate behavioral parameters for travelers with different modal split respectively, and only those coming to Kyoto by transit are considered. Besides, we exclude tourists who are not on the first day of travel, as tourists may visit completely different destinations on different days and are unlikely to go to the same place again. Although it is a feasible way to predict the path of the next day by excluding places that have already been visited, the survey data does not include the places visited previously thus to validate the multi-day trip prediction will be impossible. For now, we exclude tourists with multi-day travels and predict a one-day travel itinerary for all tourists. The best set of parameters for tourists using transit is illustrated by (6.2).

$$u_{ij}^{T} = -(t_{ij} + 0.013c_{ij}) \tag{6.2}$$
$$u_{n,k}^{P} = 393.72 \times \boldsymbol{P}_n^T \cdot (\boldsymbol{U}_{i_k} \circ (1 - F(\boldsymbol{A}_{n,k}; 0.859,\ 0.391))$$

where $\circ$: entrywise product; $\boldsymbol{P}_n$:preference of tourist n; $\boldsymbol{U}_{i_k}$: intrinsic utility of k^{th} POI;

$\boldsymbol{A}_{n,k}$: accumulated utility; T_{i_k}: activity time

To measure how well the estimated behavioral model accounts for the error in tourists' decision-making process, we then calculate the pseudo R square by comparing the prediction error of the model estimated with that of the null case, a case where the problem degenerates into a typical Orienteering problem, where the attraction scores are fixed (sum of intrinsic utilities) and agents maximize their total scores while maintaining the total travel time under the time budget. The null case represents a scenario where tourists do not have personal taste over different types of destinations, who discriminate attraction sites and make tours by only referring to an identical interesting level of attractions (see in 6.3).

$$u_{ij}^{T} = -t_{ij} \tag{6.3}$$
$$u_{n,k}^{P} = c\left\|\boldsymbol{U}_{i_k}\right\|_1$$

We then calculated the model fitness in contrast to the null case illustrated by (6.3) and evaluated the significance of the solution by a numerical approach.

$$\text{McFadden's } R^2 = 1 - \frac{\ln \hat{L}\,(M_{Full})}{\ln \hat{L}\,(M_{Intercept})} \tag{6.4}$$

$$\text{Model's } R^2 = 1 - \frac{Q(\boldsymbol{\theta} mod)}{Q(\boldsymbol{\theta} null)} = 1 - \frac{1736.47}{6759.52} = 0.743$$

It is revealed in (6.4) that our non-aggregate approach has a pretty high pseudo Rho-square of 0.743 in comparison with the null case where tourists perceive an identical interesting level of attractions. Our model can better describe tourists' behavior in choosing destinations and deciding the visiting order.

$$\boldsymbol{t}_{\hat{\boldsymbol{\beta}}} = \frac{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0}{\text{s.e.}(\hat{\boldsymbol{\beta}})} = \frac{\hat{\boldsymbol{\beta}}}{\sqrt{\text{var}(\hat{\boldsymbol{\beta}})}} = [0.008, 66.965, 1.468, 0.587] \tag{6.5}$$

Where

$$\begin{aligned} \text{var}(\hat{\beta}) &= [I(\hat{\beta})]^{-1} \\ &= \left(-E[H(\hat{\beta})]\right)^{-1} \\ &= \left(-E\left[\frac{\partial^2 \ln L(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'}\right]\right)^{-1} \end{aligned} \tag{6.6}$$

As we do not have the probability of the estimated parameters, we utilized a numerical differentiation tool compiled in Python to calculate the numerical Hessian, which replaces the estimated Hessian and calculates the pseudo t value. The pseudo t value is not the real t statistics, instead it indicates the change in the gradient with respect to the objective function, which is exactly the R square in our case. As revealed by the pseudo t value, we are confident about the estimated intercept on node utility and the shape of the discount factor that decides after how many visits people begin to feel fatigued. On the other hand, we are not so confident in tourists' sensitivity to transit fare, probably because they do not care much about the trivial difference in the fare compared to the time it takes.

From the estimated parameters we have some interesting findings as explained below.

First, 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. Besides, note that α equals the inverse of the value of time (VOT) that transfers the monetary cost into time. This coefficient identifies how people are sensitive to the monetary cost against temporary cost when measuring the impedance of travel. It also assesses how much people are willing to pay for a unit of saving in time while traveling. As revealed in literature, the value of road travel time of private vehicle users ranges from 36~57 JPY/min, depending on the travel purpose and type of cars[42]. In contrast to that, our model estimated VOT as 76.92 JPY per minute for tourists using transit, which is higher than that of car users given in the literature. This may indicate that travelers are not sensitive to a little increase in transit fares but emphasize the directness of travel, e.g. some might take a taxi for convenience, which resulted in a higher VOT estimated.

Second, the intercept of the node utility visit is calibrated as 393.72 (min). This represents an upper boundary of the utility tourist can gain from visiting nodes, as each entry in the intrinsic utility vector of attraction has been normalized between 0 and 1. Due to the diminishing marginal utility gained by visiting additional POIs throughout the tour, the utility for visiting extra destinations gradually decrease and will eventually be surpassed by the negative utility (impedance) for traveling there. 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. Specifically, as revealed by the calibrated parameters, the coefficient on the shape and scale of the diminishing marginal utility shows that people perceive a higher utility in their first and second visits. The gains of utility for visiting next place are history-dependent which drops sharply after the accumulated utility reaches an expected value of $k \times \theta = 0.34$. The cumulated satisfaction has a considerable negative effect on the benefit of visiting an additional place after two or three visits. Such a result on the discount factor is predictable as we give more weight on the error that measures whether the algorithm predicts the correct number of visits in the penalty function.

6.2 Predicted travel patterns

Next, we looked into the frequency and variance of travel patterns that are predicted by the model. Specifically, when enumerating the predicted trip chain of each tourist, it is revealed that areas on top of the attractiveness ranking such as 24: "祇園方面", 25: "河原町・新京極方面" and 29: "京都駅周辺" are quite dominant destinations, and their combinations will be most probably included in the tours. After getting the visit frequency of each destination by aggregating the trips to each attraction area, we also found that the three areas are the most frequently visited ones which are consistent with the observed.

Table 6.1 Frequent trip patterns (O and D excluded)

Trip Pattern	Pattern in Japanese	Count
24-25-29	祇園方面-河原町・新京極方面-京都駅周辺	467
24-25	祇園方面-河原町・新京極方面	171
24-29	祇園方面-京都駅周辺	99
24	祇園方面	58
24-25-28-29	祇園方面-河原町・新京極方面-三十三間堂周辺-京都駅周辺	28
	…	
24-16-17-25-29	祇園方面-哲学の道周辺-平安神宮周辺-河原町・新京極方面-京都駅周辺	10
24-25-20-21-29	祇園方面-河原町・新京極方面-二条城周辺-二条駅周辺-京都駅周辺	8
25-24-29	河原町・新京極方面-祇園方面-京都駅周辺	5
29-28	京都駅周辺-三十三間堂周辺	4
24-25-14-23-29	祇園方面-河原町・新京極方面-嵯峨野方面-嵐山方面-京都駅周辺	3
29-34	京都駅周辺-醍醐寺周辺	3
	…	

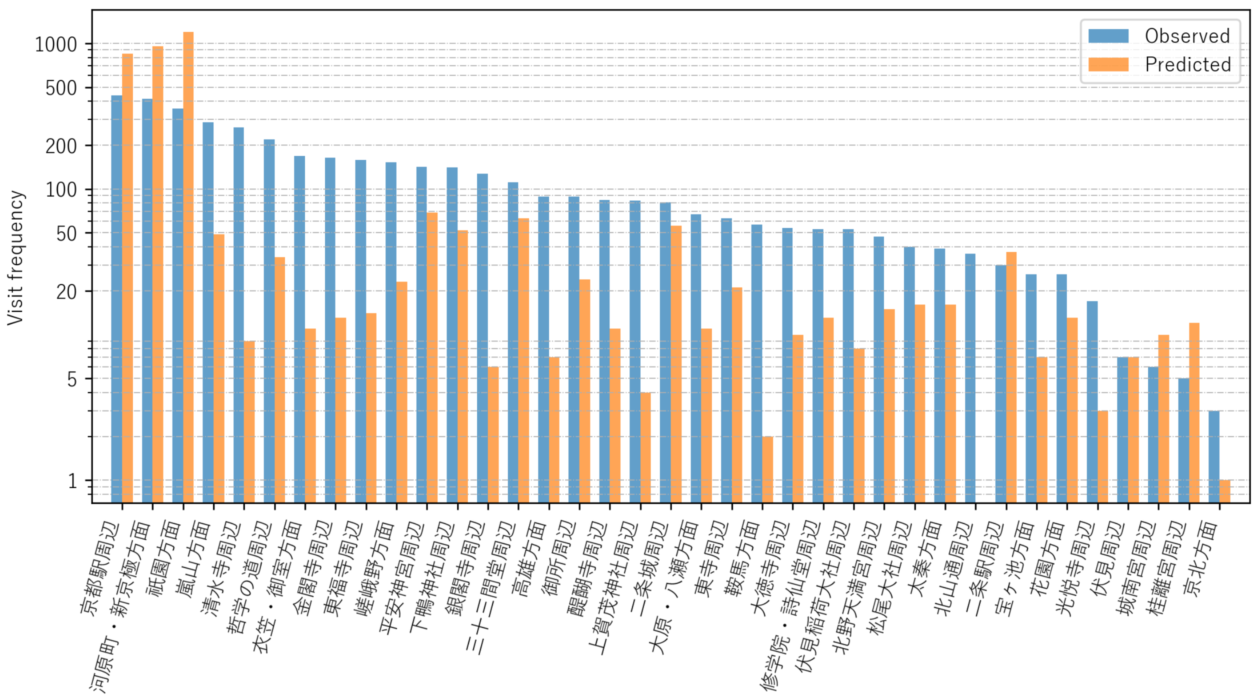


Figure 6-2 Observed and predicted visit frequency of attraction areas

Specifically, the area 24: “祇園方面”, which estimated to have the highest intrinsic utilities over all attractiveness dimensions, appeared in most frequent trip patterns predicted. Apart from area 24 with which most tourists are predicted to kick-off their tours, people behave differently in choosing the rest of the destinations and deciding the visit order, since the other attributes like preference and time budget affect their decision-making process as well. As a result, various choice patterns in making tours are reflected in the table of travel patterns.

Further, although it is not realistic to hope the algorithm predicts the exact path as observed every time, it indeed, usually predicts an alternative that is good enough in the sense of the closeness to the ground-truth and the total utility to be obtained. This is because we combined the geographical difference in the path penalty function when measuring the prediction error.

For instance, as illustrated by figure 6.3, the estimated model predicts an outgoing tour as 24-17 for a tourist touring from the “Keihan” station, which appears to be very close to the path of 16-15 geographically that he took. However, the tourist was predicted to pick up another path that consists of area 24 and then area 17 because both sites are estimated to be more attractive than those in the observed path. The utility of the observed route should be higher if we consider the fact that area 15 contained one of the famous landmarks of Kyoto.

As we estimate the attraction utilities before the parameter estimation and take that as input for the model, we appeared to over-estimate the attractiveness of such areas which caused a skewed trip distribution to those dominant areas. Moreover, we estimated the attractiveness of different areas independently, which appears to miss the covariance between the attraction areas. Many attraction areas are clustered, either because they are close to each other or they have complimentary sightseeing spots, as recommended by travel books and other resources on tourism. The existence of attraction clusters is empirically acknowledged, e.g. we can see travel agencies often cluster several attractions as combinations in promoting travel packages.

Figure 6.3 A closely predicted example

6.3 Observed and predicted trip matrices

We also compared the predicted and observed trip distribution matrices aggregated from each tourist as illustrated in Figure 6.4. As a result, the predicted trip distribution has less variation and tend to concentrate on some popular sightseeing sites. Specifically, we over-estimated the number of visits to the attraction like 24: "祇園方面", 25: "河原町・新京極方面" and 29: "京都駅周辺", which are dominantly "attractive" in terms of intrinsic utilities. This is because we assume tourists are one-day travelers and we predict a one-day travel itinerary for the tourists. This is inevitable as the model tends to underestimate the variance in people's taste in different attractions and choice behaviors, suggesting a "reasonable" logic that most tourists will include the top attractions in his tour.

Moreover, the survey did not confine the tourists to be a one-day traveler or on their first day of travel. Instead, they had probably visited several areas in previous days which were not reflected in the survey, resulting in the observed trip matrix to be more diverse in trips distribution and attraction visits. We acknowledge that we appeared to over-estimate the attractiveness of such areas as we estimate the attraction attractiveness independently before the parameter estimation and take that as input for the model, but the OD matrix fit is only partially a good indicator of how good the model performs.

Nevertheless, as shown by Figure 6.2, after converting the number of visits into the logarithm scale, we can still find an acceptable variation in tourists' destination choices among the rest of the attractions excluding the abovementioned dominant areas.

Observed trip distribution

Origin \ Dest.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
大原・八瀬方面	1	1	1	1	0	0	3	0	0	0	0	4	0	0	0	3	3	0	0	0	0	0	0	1	5	6	0	0	2	4	0	0	0	0	0	0	1	0
鞍馬方面	2	4	0	2	2	0	6	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	4	10	0	3	0	3	0	0	0	0	0	0	0	0
宝ヶ池方面	3	0	2	0	2	1	1	0	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
上賀茂神社周辺	4	0	5	2	1	0	4	1	7	5	2	21	2	0	0	0	1	0	0	0	1	0	1	0	1	4	0	0	0	1	0	0	0	0	0	0	0	0
高雄方面	5	0	0	0	0	0	1	0	0	0	0	0	0	19	13	0	0	0	0	0	1	1	1	3	0	4	1	2	1	7	0	1	1	0	0	0	0	2
修学院・詩仙堂周辺	6	3	1	2	0	0	0	0	1	0	1	3	0	0	0	3	2	3	0	0	0	0	0	0	6	6	0	1	0	3	0	2	0	0	1	0	0	0
光悦寺周辺	7	0	0	0	3	0	0	0	0	2	4	0	0	2	0	0	0	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0
北山通周辺	8	1	0	0	5	0	0	0	0	0	0	7	0	0	0	1	0	0	0	0	0	0	0	1	2	2	0	1	0	1	0	0	0	0	0	0	0	0
大徳寺周辺	9	0	0	0	4	0	0	1	0	0	9	6	0	3	0	0	0	0	1	1	1	1	0	2	2	3	0	0	0	2	0	0	1	0	0	0	0	0
金閣寺周辺	10	1	0	0	3	1	0	0	0	5	0	1	9	45	2	8	4	6	1	1	5	2	2	3	6	6	1	1	0	6	0	0	1	0	1	0	0	0
下鴨神社周辺	11	1	3	1	6	0	2	0	1	1	2	0	3	0	1	6	6	5	9	0	0	0	0	0	7	18	0	1	1	7	0	2	1	1	0	0	0	0
北野天満宮周辺	12	0	0	0	1	0	0	0	0	0	1	0	0	3	1	3	1	0	1	1	0	1	0	5	1	1	0	2	0	2	0	0	0	0	1	0	0	0
衣笠・御室方面	13	0	0	0	3	2	1	1	0	1	11	2	10	0	3	4	2	0	3	16	6	2	6	12	8	9	0	2	0	11	0	0	0	0	0	0	0	0
嵯峨野方面	14	0	0	0	0	4	0	0	0	1	3	1	1	9	1	0	1	1	3	0	1	1	2	41	2	12	0	1	0	9	0	2	1	0	0	0	0	0
銀閣寺周辺	15	0	0	0	0	0	0	0	0	1	9	0	0	0	0	0	40	10	3	0	0	0	0	0	7	3	0	9	0	8	0	0	0	0	0	0	0	0
哲学の道周辺	16	0	0	1	0	0	1	1	0	0	1	3	0	0	0	33	0	19	3	0	4	1	0	2	20	19	0	4	1	14	0	5	1	1	2	0	0	0
平安神宮周辺	17	0	0	1	1	0	0	0	0	1	0	1	0	0	0	4	27	0	3	0	0	0	0	0	13	11	0	9	3	7	0	2	0	0	0	0	0	0
御所周辺	18	1	2	0	0	0	1	0	2	4	2	7	0	0	0	3	1	4	0	0	4	1	0	1	3	7	0	2	2	9	1	0	0	0	1	0	0	0
花園方面	19	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	1	0	0	0	0	2	1	0	1	0	0	0	2	0	3	0	0	0	0	0	0
二条城周辺	20	0	0	0	0	1	0	0	1	3	5	2	1	0	0	0	0	1	4	0	0	3	0	0	8	12	0	5	0	12	0	0	0	0	0	0	0	0
二条駅周辺	21	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	2	0	2	0	0	3	0	1	0	1	0	6	0	0	0	0	0	0	0	0
太秦方面	22	0	0	0	1	0	0	0	0	0	2	0	2	2	1	0	0	0	0	1	1	0	0	6	0	2	0	2	1	2	0	0	1	0	0	0	0	0
嵐山方面	23	2	0	0	0	3	0	0	1	0	8	2	1	15	62	1	2	1	0	1	2	2	6	8	9	25	10	2	1	13	1	3	0	1	2	0	0	0
祇園方面	24	0	1	0	1	1	1	0	0	1	6	4	0	1	0	6	13	10	2	0	2	1	0	0	6	46	0	59	4	26	0	1	2	0	1	0	0	0
河原町・新京極方面	25	1	0	1	2	0	2	0	2	0	0	5	0	1	3	1	10	6	1	0	5	1	0	1	32	10	0	9	1	45	1	2	1	3	4	0	0	0
松尾大社周辺	26	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	1	0	0	0	0	0	1	16	1	6	0	1	1	2	0	0	0	0	0	0	0	0
清水寺周辺	27	0	1	0	0	0	0	0	1	1	3	0	0	0	1	5	8	2	2	0	0	0	0	5	80	21	1	1	13	12	0	7	0	1	0	0	0	0
三十三間堂周辺	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	1	0	0	0	0	0	7	6	0	22	0	13	0	4	2	0	0	0	0	0
京都駅周辺	29	0	0	2	1	2	0	3	1	2	3	1	0	2	5	1	2	3	3	0	5	1	1	4	7	16	3	14	10	17	1	7	8	4	1	1	1	0
桂離宮周辺	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
東福寺周辺	31	2	2	1	1	0	3	1	0	0	3	1	0	0	0	2	7	1	0	0	3	1	0	0	10	6	0	17	15	15	0	1	3	10	4	1	0	0
東寺周辺	32	1	0	0	0	0	0	0	0	0	1	1	0	1	0	2	0	2	1	0	0	0	0	0	2	4	0	1	0	14	0	3	1	0	1	1	0	0
伏見稲荷大社周辺	33	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	3	2	0	0	0	9	0	8	2	0	2	0	2	0
醍醐寺周辺	34	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	5	0	0	2	1	0	0	1	4	14	0	1	0	5	0	5	0	0	0	0	1	0
城南宮周辺	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
伏見周辺	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0
京北方面	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Predicted trip distribution

Origin \ Dest.		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
大原・八瀬方面	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
鞍馬方面	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
宝ヶ池方面	3	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0
上賀茂神社周辺	4	0	1	0	1	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
高雄方面	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
修学院・詩仙堂周辺	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
光悦寺周辺	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
北山通周辺	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
大徳寺周辺	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
金閣寺周辺	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	1	1	0	0	0	4	0	0	0	0	0	0	0	0
下鴨神社周辺	11	0	1	3	2	0	0	0	0	0	0	2	0	0	1	0	0	4	0	0	0	0	0	3	1	10	0	1	0	11	0	0	0	0	0	0	0	0
北野天満宮周辺	12	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	2	0	0	2	0	0	0	0	3	0	0	0	0	0	0	0	0
衣笠・御室方面	13	0	0	0	0	3	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	5	0	0	0	0	0	0	0	1
嵯峨野方面	14	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	12	0	3	0	0	0	0	0	0	0	0	0	0	0	0
銀閣寺周辺	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0	0	0	0	0
哲学の道周辺	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	26	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0
平安神宮周辺	17	0	0	1	1	0	0	0	0	0	0	7	0	0	2	1	1	1	0	0	4	0	0	0	1	22	2	0	0	6	0	0	0	0	0	0	0	0
御所周辺	18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	4	0	0	0	3	0	0	0	0	0	0	0	0
花園方面	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
二条城周辺	20	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0	0	0	0	2	1	7	0	0	0	10	0	0	0	8	0	0	0	0	0	0	0	0
二条駅周辺	21	0	0	0	0	0	0	0	0	0	1	0	1	3	0	0	0	0	0	1	5	3	1	0	0	2	0	0	0	12	0	0	0	0	0	0	0	0
太秦方面	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
嵐山方面	23	0	0	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	2	0	6	1	4	4	0	0	11	1	0	0	0	0	0	0	0
祇園方面	24	6	1	7	7	5	0	0	0	6	3	12	5	7	4	9	22	30	10	7	5	7	1	7	11	796	2	15	7	140	1	8	4	0	1	2	0	0
河原町・新京極方面	25	1	0	2	0	0	0	0	0	1	5	14	7	7	6	0	7	4	2	5	21	6	1	5	31	15	0	1	39	508	1	0	0	0	9	0	0	0
松尾大社周辺	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
清水寺周辺	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	2	4	2	0	0	0	0	1	0	0	0
三十三間堂周辺	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	52	0	2	0	0	0	0	0	0
京都駅周辺	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	3	1	0	10	16	0	2	2	23	0	0	20	6	1	2	0	0
桂離宮周辺	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	0	0	0	0	0	0	0
東福寺周辺	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	4	0	0	0	0	1	0	1	0	0
東寺周辺	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
伏見稲荷大社周辺	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
醍醐寺周辺	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	5	0	0	0	0	0	0	0	0
城南宮周辺	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0
伏見周辺	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
京北方面	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 6-4 Observed and predicted trip distribution

Chapter 7　Simulation of TDM strategies

In this chapter, we test the effectiveness of various Travel Demand Management (TDM) strategies by looking at how tourists change their tour routes under the behavioral principles estimated. Specifically, we simulate the strategies that will result in a change of the network properties, e.g. node attractiveness and edge travel time, to see how people adapt to the changes and update their trip chains, which will eventually lead to changes of the distribution of tourists among all attraction areas as well as the volume of travel flows in between. Our main interest in the TDM strategy includes changes in the public transport network (which may alleviate the problem) and changes in the attractiveness of certain sightseeing spots to see if this reduces or increases the traffic to the most crowded one if nearby attractions become more interesting.

To do this, we first look at the OD pairs with the most frequent traffic and consider where there is traffic congestion. By aggregating the individual's trip chains observed from the survey, the travel frequency table shows the crowd's movement and indicates the OD pairs with heavy traffic. However, because the capacity to hold tourists varies by areas and links, large traffic volume does not necessarily mean traffic jams, another way is to confirm from other data sources that somewhere was suffering from traffic jams, or that heavy traffic was about to happen.

7.1 Identifying problematic OD pairs

As a few extra questions, tourists were asked to reflect on their own trips and select and respond with whatever they were not satisfied with each trip. We use this "complaint" data to locate problematic OD pairs, as frequent complaints against certain OD pairs indicate actual traffic congestion and other issues, so we can refer to this information to identify transit route pairs that need improvement and apply them with the TDM strategies.

利用交通手段の不満点

→自家用車・レンタカーの不満点

A．道路が混雑
B．目的地への経路がわかりにくい
C．道路が細い
D．駐車場がない・場所や入り口がわかりにくい
E．駐車場の待ち時間が長い
F．駐車場代が高い

→市バス・民バスの不満点

G．バスの本数が少ない
H．バス運賃が高い
I．道路が混雑
J．乗り換えがめんどう
K．どのバスに乗るのかわかりにくい
L．バス停がどこにあるのかわかりにくい
M．バスの運転が荒い
N．バス車内が混雑

→ＪＲ・その他私鉄・地下鉄の不満点

O．電車の本数が少ない
P．鉄道運賃が高い
Q．乗り換えがめんどう
R．駅のバリアフリー化が不十分
S．駅がどこにあるのかわかりにくい
T．電車内が混雑

→鉄道～バス間の不満点

U．鉄道～バス間の乗り換えがめんどう

Figure 7-1 Categories and choices of complaints

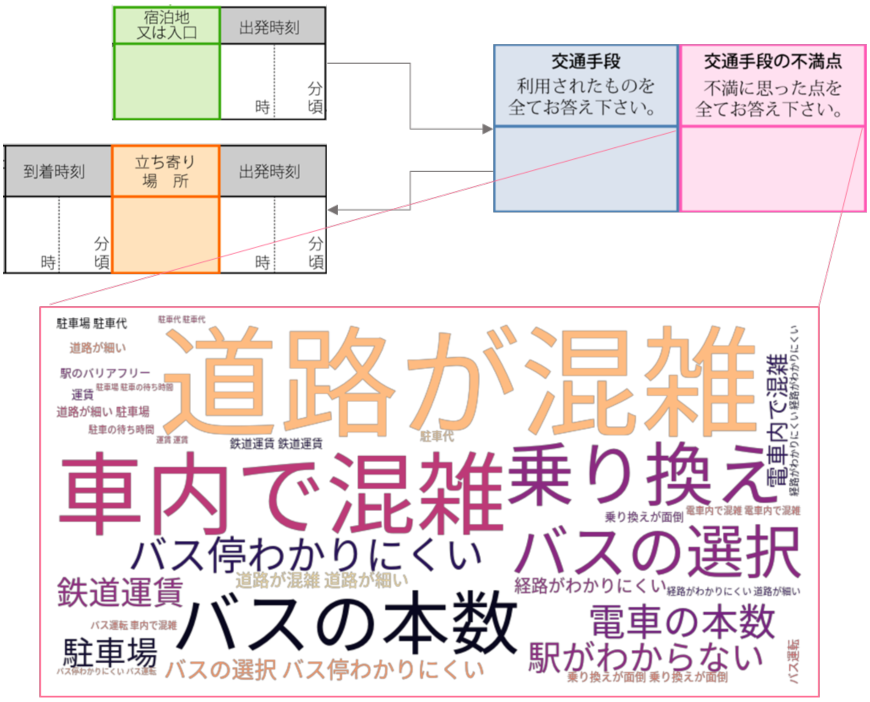


Figure 7.2 A visualization of frequent complaints

We calculated the complaint rate by dividing the number of complaints by total trips observed between each OD pair and found the OD pairs that have been the most complained about in areas such as road congestion, congestion in cars, transit fares or operating frequencies. For example, as illustrated by Figure 7.3, tourists felt unpleasant about road congestion in among 88% of the trips traveled from “Sanjusangen-do Temple” to “Gion”, 68% of the trips traveled from “Ginkakuji area” to “Arashiyama region”, and 48% from “Kyoto station area” to “Kiyomizu Temple area”. We identify above OD pairs with heavy road traffic as the ones need to be improved in terms of travel time. Similarly, we identified problematic OD pairs in terms of:

a) cabin congestion
  - from “Arashiyama region” to “Gion”

b) operation frequency
  - from “Ginkakuji area” to “Gion”
  - from “Kyoto station area” to “Sanjusangen-do Temple”

c) transit fare
  - from “Sagano Region” to “Kawaramachi area”

We then simulate TDM strategies by improving the level of service of identified OD pairs corresponding to each category of complaints and measure the effectiveness of the above strategies.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
大原・八瀬方面	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
鞍馬方面	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
宝ヶ池方面	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上賀茂神社周辺	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
高雄方面	5	0	0	0	0	0	0	0	0	0	0	0	0	0	6.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
修学院・詩仙堂周辺	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
光悦寺周辺	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
北山通周辺	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
大徳寺周辺	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
金閣寺周辺	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
下鴨神社周辺	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.6	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0
北野天満宮周辺	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
衣笠・御室方面	13	0	0	0	0	0	0	0	0	0	0	0	6.2	0	0	0	0									8.7	0	0	0	0	0	0	0	0	0	0	0	0
嵯峨野方面	14	0	0	0	0	0	0	0	0	0	0	0	0	0	3.8	0	0								0	25	0	0	0	0	0	0	0	0	0	0	0	0
銀閣寺周辺	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.8	0	0	0	0	0	0	0	68.4	0	0	0	0	0	0	0	0	0	0	0	0	0
哲学の道周辺	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11.9	0	0	0	0	0	0	1.6	9.4	0	0	0	0	0	0	0	0	0	0	0	0
平安神宮周辺	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	29.2	0	0	0	5.6	0	0	0	0	0	0	0	0
御所周辺	18	0	0	0	0	0	0	0	0	0	0	10.5	0	0	0	0	0	0	0	0	0	0	0	0	0	5.9	0	0	0	15.8	0	0	0	0	0	0	0	0
花園方面	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
二条城周辺	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
二条駅周辺	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
太秦方面	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
嵐山方面	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.4	0	0	0	0	0	0	0	0
祇園方面	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7.7	0	0	0	0	0	0	4.3	0.9	0	3.3	0	0	0	0	0	0	0	0	0	0
河原町・新京極方面	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15.8	0	0	0	0	0	0	0	2.9	9	0	0	0	3.4	0	0	0	0	0	0	0	0
松尾大社周辺	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
清水寺周辺	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.5	0	0	0	8									
三十三間堂周辺	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88.2	0	0	0	18.4	0	0	0	0	0	0	0	0
京都駅周辺	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									0	10	0	0	48	0	0	0	4.3	0	0	0	0	0
桂離宮周辺	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
東福寺周辺	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.5	0	0	0	0	0	0	0	0	0	0	0	0	0
東寺周辺	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16.2	0	0	0	0	0	0	0	0
伏見稲荷大社周辺	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
醍醐寺周辺	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
城南宮周辺	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
伏見周辺	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
京北方面	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

銀閣寺周辺->嵐山方面

京都駅->清水寺周辺

三十三間堂->祇園方面

Figure 7-3 OD pairs with most complaints on road congestion

7.2 Simulation scenarios

In the simulation, we try scenarios with single and combinations of strategy and evaluate the effectiveness of such strategies by looking at the revealed trip statistics and see if it reduces or increases the traffic to the most crowded attraction area, and whether the trips are more concentrated or dispersed. Successful strategies include those that encourage to disperse the tourists' travel demand and make trips to different areas more balanced and distributed.

For the set of strategies, we consider possible changes in the public transport network and attraction areas that might alleviate the problem. We divide the strategies into the following three categories and design simulation scenarios respectively. For each type of strategy, we identify the targets for improvement by referring to the "complaints" data as well as intuitive and empirical experience.

7.2.1 Reducing the travel impedance

The overcrowded site, packed buses, and road congestion have been considered as key issues affecting tourist satisfaction in Kyoto[43]. As identified in 7.1, many tourists were upset when traveling between certain OD pairs, with complaints on traffic jams and congestion in the cabin.

For those OD pairs that have frequent complaints and discontent on congestion and operation interval, we assume a reduction in travel time to simulate the strategies such as increasing the operating frequency or introducing alternative lines. The above strategies help to increase the level of service of transit and will result in a shortening of waiting time. More supplies of a same bus line and alternative lines help to disperse the tourist's demand as well.

Besides, due to the limited road space, most bus stations in Kyoto city can only serve one bus a time. Buses cannot overtake in the queue, and at the same time, bus bunching is frequently happening. To overcome this inherent shortcoming, potential strategies also include expanding the bus platform so that multiple buses can be served at the same time to increase the capacity of bus stops.

We ran simulation scenarios with strategies listed in table 7.1 respectively. For each OD pair targeted we apply a strategy of shortening the edge travel time by 20%. In strategy 4, due to the clustering nature of the attraction areas, we presume a reduction in the travel time of the bus line that operates between them. To evaluate the effects of such strategies, we first looked at the shift of trip patterns as illustrated in table 7.2.

Table 7.1 The edge strategy set

Strategy #	**Target OD(s)**	**Target(s) OD in Japanese**	**Strategy Effect**
1	28-24	三十三間堂周辺-祇園方面	A decrease in the edge travel time by 20%
2	15-23	銀閣寺周辺-嵐山方面	Same as above
3	29-27	京都駅周辺-清水寺周辺	Same as above
4	ODs between 15, 16, 17, 24	銀閣寺周辺，哲学の道周辺，平安神宮周辺，祇園方面	Same as above
5	29-28	京都駅周辺-三十三間堂周辺	Same as above
6	14-25	嵯峨野方面-河原町・新京極方面	Same as above

Table 7. 2 enumerates significant changes in the travel pattern frequencies after applying different TDM strategies on edges. Note that ODs are excluded and only effective strategies are illustrated. We define effective strategies as those causing >= 5 numbers of change in attraction visit frequency to avoid random error, e.g. alternative paths with very similar utility in tour prediction. Moreover, with the limited space available we cannot list all of them but some significant changes that worth mentioning. Finally, we aligned routes that are either similar or close to each other clustered together for a better review.

The strategy 1 shortens the travel time between 24: 祇園 and 28: 三十三間堂周辺 increases the number of travels to the Sanjusangen-do area (三十三間堂周辺) by inserting an extra visit based on the most frequent ones. Specifically, it reduces the travel patterns like going directly from 24-29 by encouraging tourists to have another visit to the Sanjusangen-do area in between. However, it tends to increase the travel pattern 24-25: 祇園—河原町・新京極方面 as well due to the two areas are close to each other and are complementary in terms of intrinsic utilities.

Third, strategies 2~6 are quite successful in increasing the travel patterns in which the Path of Philosophy area is included. More people are encouraged to include 16: 哲学の道周辺 and 17: 平安神宮周辺 as extra visits in their travel patterns. Among the four strategies, the strategy 4 which alleviates the travel impedance between the four areas that locate around the Path of Philosophy area (哲学の道) had a significant effect in introducing new travel patterns that go round that area, at the meantime suppressing a great number of tours going directly from 24-25. As a result, this strategy is successful in introducing travel to the expected area and it helped to disperse the most congested trips. Strategy 6, which alleviates the travel cost from 14: 嵯峨野方面 to 25: 河原町・新京極方面, contributed to the increase in such patterns as well.

Strategy 5 decreases the travel impedance from 29: 京都駅周辺 to 28: 三十三間堂周辺 had a similar effect as strategy 1 in encouraging an intermediate visit to area 28 before ending the tour. It reduced the most frequent travel patterns and dispersing the travel demand after alleviating the travel impedance to near areas, e.g. the Path of Philosophy area (哲学の道).

For other travel patterns, tours which travel directly from 24 to 12: 北野天満宮周辺 or 11: 下鴨神社周辺 are suppressed by most strategies. Another two travel patterns that include attraction located at the south side of the city tend to increase after strategy 6 is applied. However, strategy 6 failed our expectation that it did not attract more visits to 14: 嵯峨野方面 or 23: 嵐山方面 as expected. This might be because the shortening in travel time was too trivial, or improving a single OD pair is not effective in attracting more visits or introducing another beneficial travel pattern.

We also looked at the change in attraction visit frequency after introducing such edge strategies. As illustrated in Figure 7.5, although strategies 1~3 reduced the traffic to the very crowed area 29: 京都駅周辺, strategy 1 and 2 raised the number of visits to 24: 祇園 and 25: 河原町・新京極方面 by a little bit. They also tend to increase the number of visits to the Arashiyama area (14 and 23) but not as much as we expected. Strategy 1 helps introduce more travels to 28: 三十三間堂周辺. All three strategies increased the travel demand to areas like 15: 銀閣寺周辺, 27: 清水寺周辺 and 33: 伏見稲荷大社周辺, of which area the intrinsic utilities tend to be under-estimated.

On the other hand, strategy 4 is quite successful is introducing traffic to 16: 哲学の道周辺 and 17: 平安神宮周辺 areas. Strategy 5 appeared to help disperse the travel demand to area 29: 京都駅周辺 and introducing more visits to areas like 17: 平安神宮周辺, 28: 三十三間堂周辺, which matched our expectations. Strategy 6 suppressed the traffic to abovementioned areas, increase the demand to 14: 嵯峨野方面 but reduced the number of visits to the Arashiyama area (23), which failed our expectation.

Finally, although transit fares are considered as another key factor in representing the negative utility of traveling on each edge, we did not enumerate fare strategies separately as the effect of such change in the amount of fare can be converted to time with the coefficient VOT. Thus, we enumerated the time strategies only to represent the change in travel impedance during the simulation.

Table 7.2 The edge strategy effects on trip pattern frequencies

Trip Patte	Predicted	STRAT.1	STRAT.2	STRAT.4	STRAT.5	STRAT.6
Most frequent						
24-25-29	467				-12	
24-25	171	+15		-30	+5	+5
24-29	99	-16		-6	-9	
24-29-25	10	-9			-8	
Sanjusangen-do area (三十三間堂周辺)						
24-25-28-2	28	+7	-5		+11	-8
24-28-29	1	+17			+9	
The Path of Philosophy area (哲学の道, 平安神宮周辺)						
24-16-17-2	10		-5	+14	-5	-4
24-17-25-2	1		+5			
24-16-17-2	2		+5	+18	+5	+5
24-16-17	0			+14	+5	
25-24-16-1	0			+6		
24-15-16-1	1			+4		
24-25-17	4				+6	
24-17-29	2					+4
Others						
24-25-12	0	+5				
24-25-18	7	-5	-4			
24-18	1	+5				
24-12	7	-4	-6		-4	
24-11	6	-6	-6			-4
24-4	0		+5			
24-25-29-3	0					+6
24-34	2					+4

*		
	4 上賀茂神社周辺	24 祇園方面
	11 下鴨神社周辺	25 河原町・新京極方面
	12 北野天満宮周辺	28 三十三間堂周辺
	16 哲学の道周辺	29 京都駅周辺
	17 平安神宮周辺	33 伏見稲荷大社周辺
	18 御所周辺	34 醍醐寺周辺

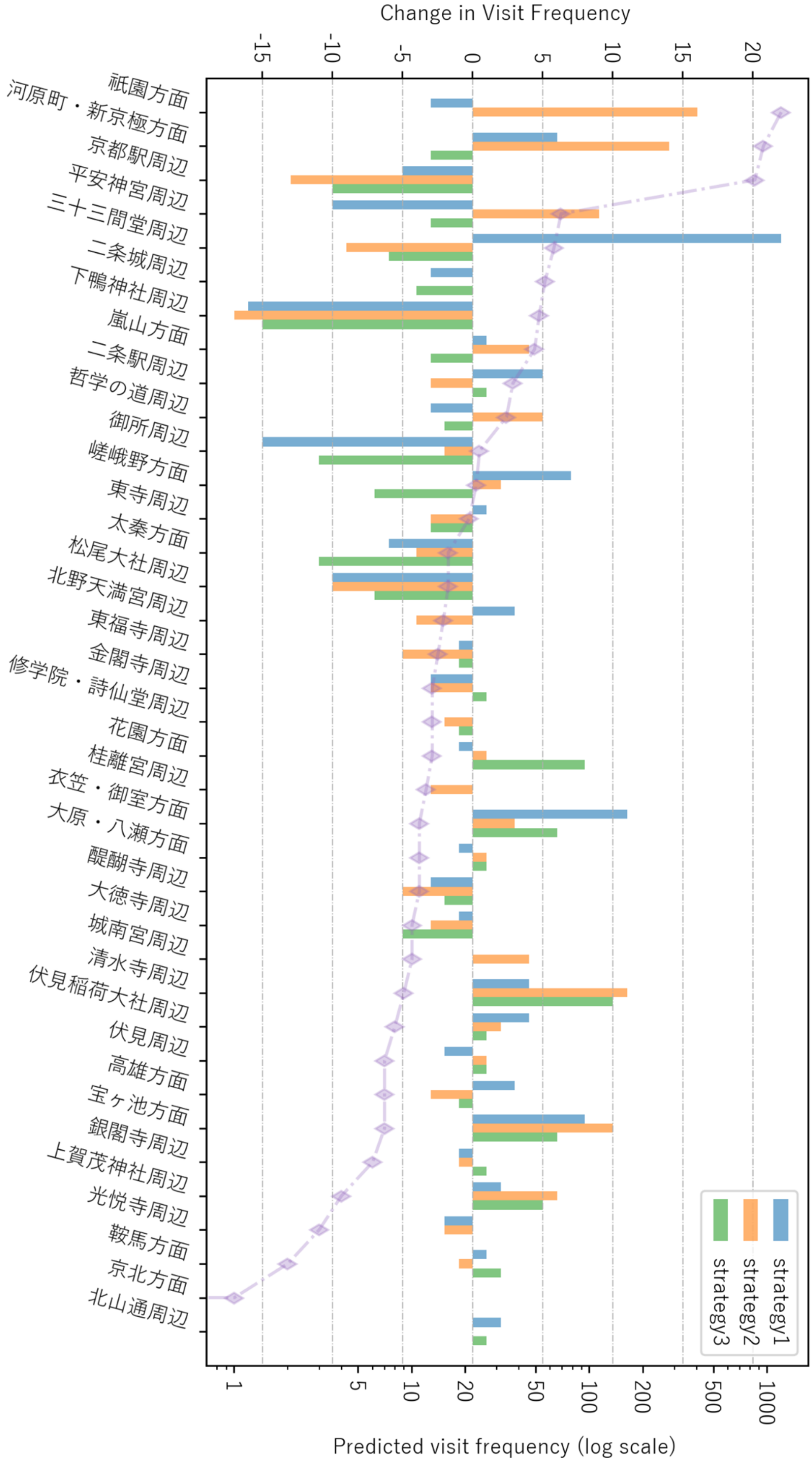


Figure 7-4 Effects to the attraction visit frequencies by strategy (1-3)

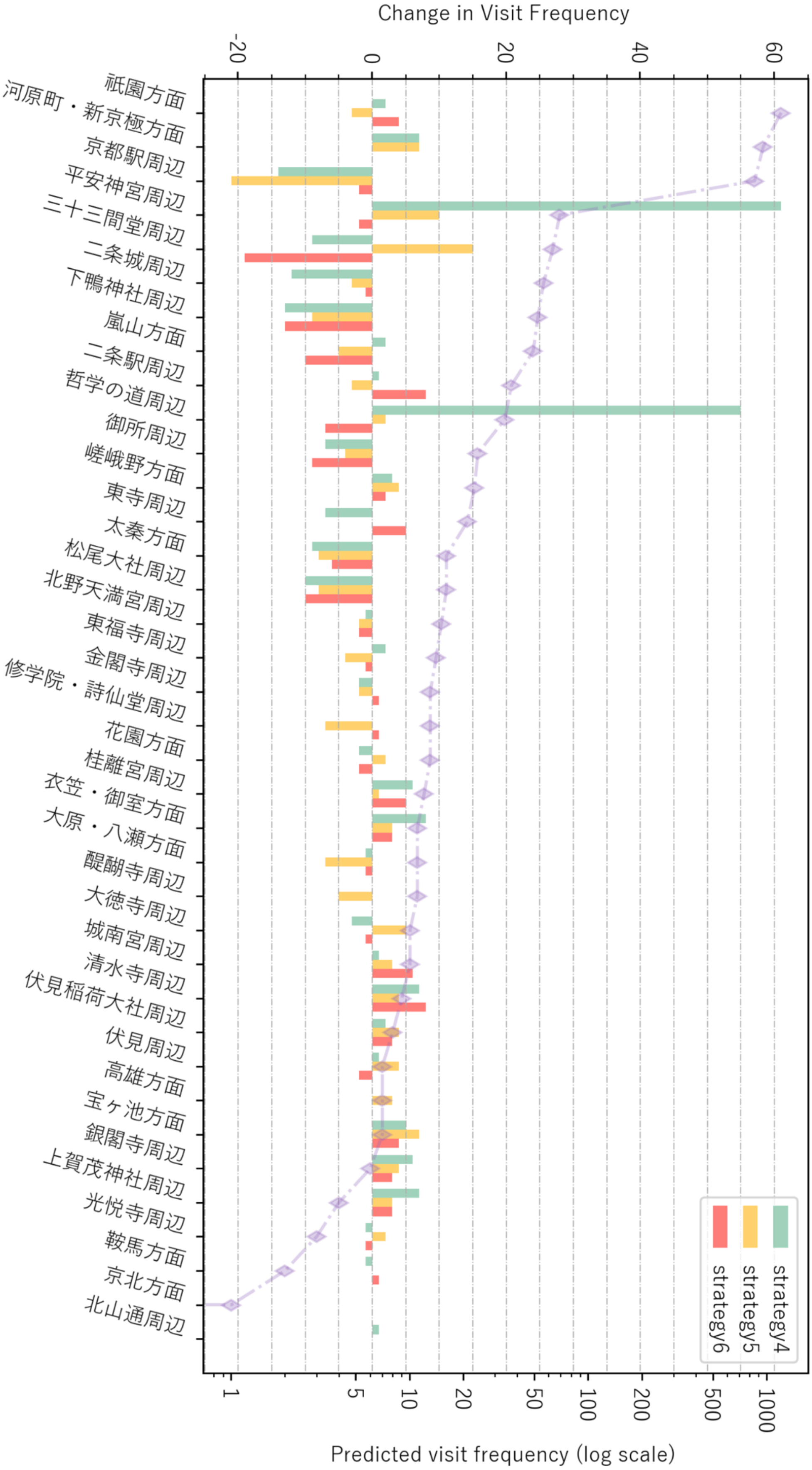


Figure 7-5 Effects to the attraction visit frequencies by strategy (4-6)

7.2.2 Enhancing the attractiveness of spots

Finally, we adjust the attractiveness of certain sightseeing spots to see if this reduces or increases the traffic to the most crowded one if nearby attractions become more interesting.

The strategy set is illustrated in table 7.3. For the effects of each strategy, we presume an increase in the utilities of the two dimensions: "gourmet" and "leisure activities" by introducing more facilities like pubs, restaurants, museums and art galleries. For each node in the strategies and for each dimension to be applied, we simulate an improvement in the intrinsic utilities by one third. At last, we reconsider the intrinsic utilities of several under-estimated areas that should have higher utilities based on empirical our knowledge and impose the same increase in the utilities intuitively.

Table 7.3 The node strategy set

Strategy #	Target Node(s)	Target Node(s) in Japanese	Strategy Effect	Comment
1	1, 2, 35, 36, 37	大原・八瀬方面, 鞍馬方面, 城南宮周辺, 伏見周辺, 京北方面	A 1/3 increase of utility in "gourmet" and "leisure activities"	Outskirt areas
2	20, 21	二条城周辺, 二条駅周辺	Same as above	The Nijo area
3	19, 22	花園方面, 太秦方面	Same as above	The Hanazono area
4	26, 30 14, 23	松尾大社周辺, 桂離宮周辺 嵯峨野方面, 嵐山方面	Same as above	The Katsura and Arashiyama area
5	31, 32, 33	東福寺周辺, 東寺周辺, 伏見稲荷大社周辺	A 1/3 increase of utility in all dimensions	The Fushimi area (伏見)
6	10, 14, 15, 23, 33	金閣寺周辺, 嵯峨野方面, 銀閣寺周辺, 嵐山方面, 伏見稲荷大社周辺	A 1/3 increase of utility in "red leaves and temple shrines"	Under-estimated areas

The effects of node strategies on the shift of travel patterns are illustrated in table 7.4. Compared to the effects of edge strategies that modify the network travel impedance, node strategies have a more intuitive and direct impact on the shift of travel patterns.

Firstly, almost all strategies tend to decrease the frequency of the most popular travel patterns which include 24, 25 or 29 in the tour. Specifically,

Even though we expected an increase of the travel patterns where distant attraction areas are included by applying strategy 1, such effect turned out to be not significant probably due to the travel cost for getting there outweighed the limited improvement in their attractiveness. Nevertheless, it dispersed the number of the most frequent pattern: 24-25-29 and we also observed a little increase in the tours that go around the Path of Philosophy area (哲学の道).

We improved the attractiveness of some potential areas near those popular sites by strategy 2 and 3 and see if such strategies help disperse the traffic and introduce new patterns. Specifically, by increasing the intrinsic utilities of attractions 20, 21 in the Nijo area, we observed a drop in the frequency of the pattern 24-25-20-29 but an increase in 24-25-20. Such a shift is quite interesting in that, compared to the former in which the tourist inserts area 20: 二条城周辺 as an extra visit before going to 29: 京都駅周辺, he now stopped after visiting area 20 because he obtained enough satisfaction. As a result, the accumulated utility or in other words the fatigue led to a decreasing utility obtained for going to next places. Such travelers now chose to end their tours instead of detouring to the Kyoto station area for a last visit. On the other hand, strategy 3 helped to disperse the travel demand of the most frequent pattern, increased the frequency of tours going around the Path of Philosophy area which is not expected, while was not effective in introducing new trip patterns that include the target areas we tend to promote.

Strategies 4~6 all have significant effect in reducing the most frequent tour patterns and tours inside the Nakakyo-ku area. In strategy 4, we improved the attractiveness of two less-known sightseeing spots 26: 松尾大社周辺 and 30: 桂離宮周辺. To account for the error in the under-estimated attractiveness of the Arashiyama area 14 and 23, we presume an increase in the intrinsic utilities of the two areas as well, and see if people will be encouraged to visit such areas as well given that more of them tend to include Arashiyama area in the journey. As a result, such strategy introduced new patterns in which people tend to choose 23: 嵐山方面 instead of 24: 祇園方面 as their first visit. Several patterns with 23 as the leading visit are observed, with a diminish of pattern 24-23. However, we did not observe an significant increase in the patterns that go to 26 or 30 as we expected. On the contrary, there is an increase in the number of people that visit 32: 東寺周辺.

We improved the attractiveness of areas that are on the south side of the Kyoto Station in strategy 5. The result met our expectation as it introduced new travel patterns going to that area and suppressed the most frequent patterns. In strategy 6 we tested the effect for improving the underestimated areas. We observed more patterns going to the Arashiyama area similar to the result in scenario 4, but no significant improve in travel patterns that include 10: 金閣寺周辺, 15: 銀閣寺周辺 or 33: 伏見稲荷大社周辺 if we improve those areas along, especially if we compare it to the effect by strategy 5 which improved the south area altogether.

Finally, we examined the impact of above strategies in the change of attraction visit frequencies. Strategy 1~3 all reduced the traffic to the crowded Kyoto station area, but increase the trips to 25: 河原町・新京極方面. Each of the strategies tend to increase the visits to the improved areas targeted, e.g. strategy 1 increased the visits to 35: 城南宮周辺 and 36: 伏見周辺 whereas strategy 2 increased the trips to the Nijo area. In contrast, strategy 3 was not effective in introducing more visits to its target areas. Addtionaly, despite an increase of utilities for area 37: 京北方面, we hardly observed an increase to the area in any of the scenarios.

On the other hand, strategy 4~6 have more direct effects in changing the attraction visit frequency. As illustrated in Figure 7.7, Strategy 4 and 6 have similar effects in reducing the traffic to the most crowded area 24: 祇園方面 and bringing more visits to the Arashiyama area 14 and 23. Compared to strategy 6, strategy 4 is more significant to increase the visits to such areas while the latter tends to disperse the shifted travel demand to other improved areas as well. Strategy 5 has the most direct and intuitive impact as it increased the number of visits to its targeted areas as expected.

Table 7.4 The node strategy effects on trip pattern frequencies

Trip Pattern	Predicted	STRAT.1	STRAT.2	STRAT.3	STRAT.4	STRAT.5	STRAT.6
Most frequent							
24-25-29	467	-21	+13	-17	-40	-38	-46
24-25	171	+8	-14	+8	-12	-7	-13
24-29	99		-5		-7	-16	
24	58				-5	-8	
24-25-28-29	28		-5		-11	-7	-11
29	21					-9	
Nakakyo(中京)							
24-25-20	5		+5				
24-27-25-29	0			+6	+6		
24-12	7	-6	-6	-5	-6	-5	-5
24-25-20-29	11		-8		-8		-8
24-11	6		-5			-5	
24-25-18	7			-5			-6
24-25-20-21-29	8				-7		-6
The Path of Philosophy area (哲学の道, 平安神宮周辺)							
24-25-16-17	3	+4		+4			
24-16-17-25-29	10		-7				
24-16-17	0			+5			
24-16-17-25	2				+8	+5	
24-17-25-29	0					+5	
Arashiyama area (嵐山周辺)							
24-23	5				-5		-5
23-14-25-29	0				+39		+34
23-29	0				+17		+19
23-25-29	0				+11		+12
23	0				+7		+8
23-14-24-25-29	0				+7		+7
23-14-25	0				+7		+6
23-14-24-25	0				+5		+7
23-24-25-29	0						+5
Fushimi Area (伏見方面)							
33-31-28-29	0					+23	
32-29	0					+14	
33	0					+10	
31-28-29	0					+8	
32	2				+8	+5	
31-29	1					+5	

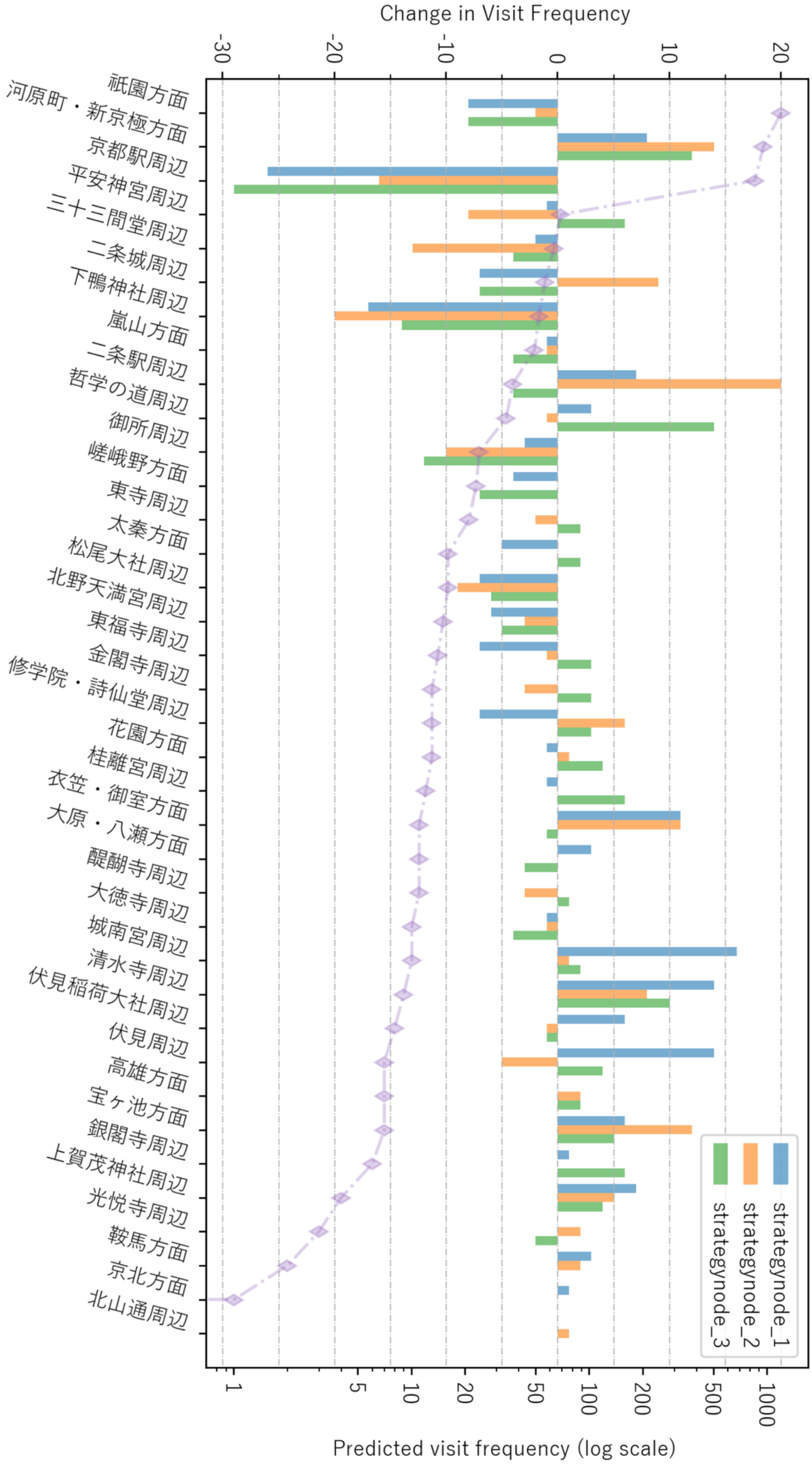


Figure 7-6 Effects to the attraction visit frequencies by node strategy (1-3)

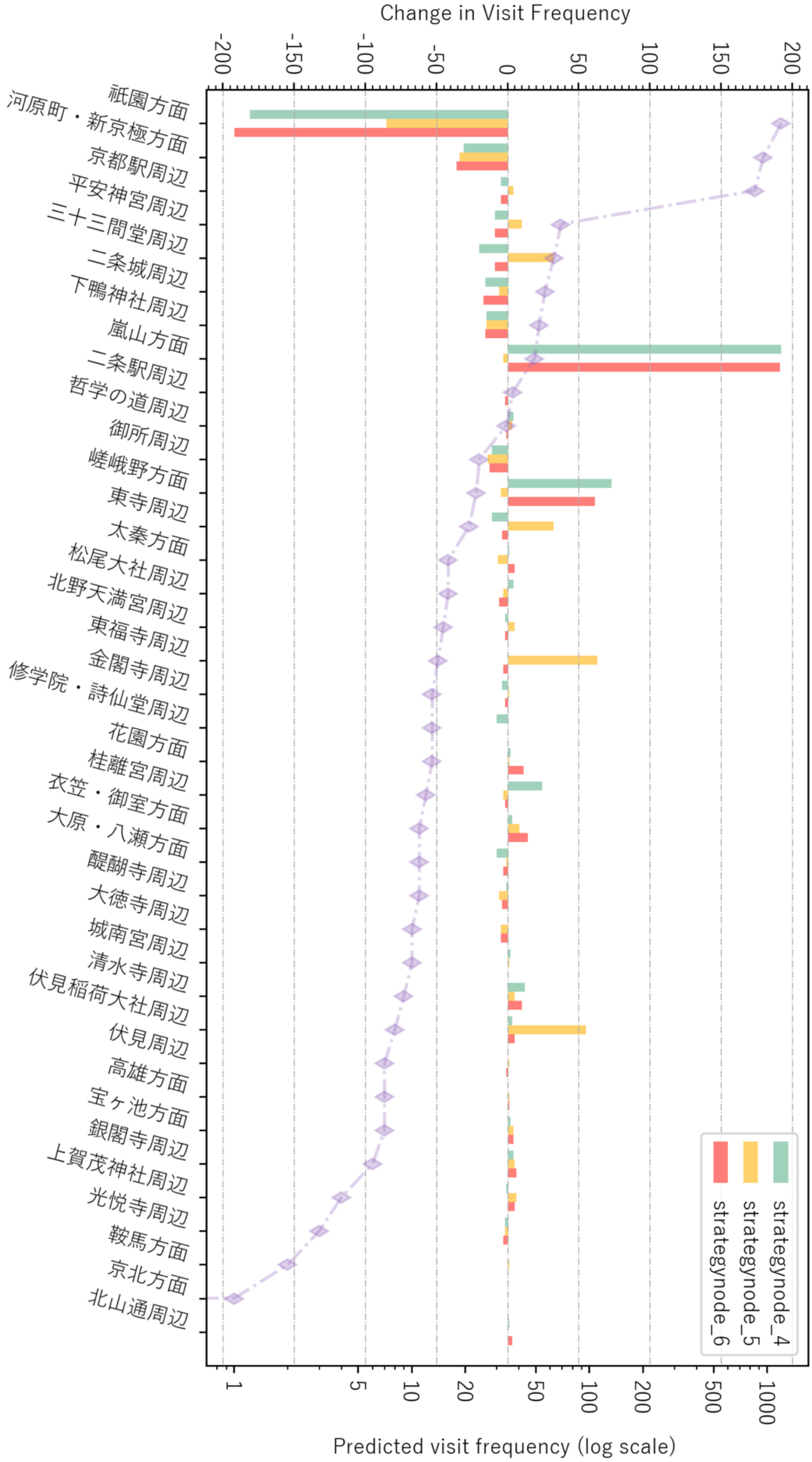


Figure 7-7 Effects to the attraction visit frequencies by node strategy (4-6)

Chapter 8 Conclusions and discussions

Travel is a derived demand from activity participation. Although trip-based and activity-based models have been popular measures for estimating travel demand of ordinary travelers, there is limited literature discussing travel demand forecasting on tourists. On the other hand, although recent studies that take advantage of big data have succeeded in rebuilding tourists' trajectories by which tourists' travel patterns can be found, these statistical approaches cannot predict something that has not happened before and are of limited help for travel demand management.

8.1 Contributions

To supplement the studies in tourist travel demand estimation, we contribute to this field by proposing a tour-based model for modeling the travel behavior of each decision-maker. Specifically, we extend the existing literature in the following two aspects:

First, we introduce a multi-dimensional formulation to evaluate the preference of tourists and the utility of attractions in different categories. In the prediction of tourist preference, we discovered that tourists indeed have different purposes towards the goal of achieving satisfaction for city tourism. They have various tastes and preferences on tourism attractions and reflect significantly different behaviors in choosing destinations and making tours. Specifically, travelers have different patterns of travel purposes. From the survey data collected in 2006 by the Kyoto city government, we observed three patterns that can divide the tourists into 3 clusters with different combinations of travel purposes. 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

We also explored the relations between which clusters they belong to with their socio-demographics and other trip-related attributes. Specifically, a 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 indicate 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 center during the tourist season, especially during the red leaves and cherry blossoms seasons which Japan is famous for.

Besides, 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.

Table 8.1 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	

Our multi-dimensional criteria replace the common "single-score" method for measuring the attractiveness of tourism sites and can better explain the variance in tourists' tastes and the intrinsic utilities of tourist destinations.

Second, we suggest that the choice of destinations will be "history-dependent" in that there is diminishing marginal utility gained by visiting additional POIs throughout 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.

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:

$$u_{ij}^{T} = -(t_{ij} + 0.013c_{ij}) \tag{8.1}$$

$$u_{n,\,k}^{P} = 393.72 \times \boldsymbol{P}_n^T \cdot (\boldsymbol{U}_{i_k} \circ (1 - F(\boldsymbol{A}_{n,k}; 0.859,\ 0.391))$$

where $\circ$: entrywise product; $\boldsymbol{P}_n$:preference of tourist n; $\boldsymbol{U}_{i_k}$: intrinsic utility of k'th POI;

$\boldsymbol{A}_{n,k}$: accumulated utility; T_{i_k}: activity time

We then calculated the model fitness in contrast to the null case illustrated by (8.2) and evaluated the significance of the solution by a numerical approach.

$$\text{McFadden's } R^2 = 1 - \frac{\ln \hat{L}\,(M_{Full})}{\ln \hat{L}\left(M_{Intercept}\right)} \tag{8.2}$$

$$\text{Model's } R^2 = 1 - \frac{Q(\boldsymbol{\theta}mod)}{Q(\boldsymbol{\theta}null)} = 1 - \frac{1736.47}{6759.52} = 0.743$$

It is revealed that compared with the ordinary modeling approach where tourists perceive an identical interesting level of attractions, our model can better describe tourists' behavior in choosing destinations and deciding the visiting order. As a tour-based travel demand forecasting system for tourists, its non-aggregate nature allows us to simulate tourism management strategies for a variety of scenarios.

Specifically, we ran simulation scenarios with different kinds of strategies as elaborated in Chapter 7, and evaluated their performance by looking at the shift in tourists' travel patterns as well as the change in the visit frequency for each attraction area. The results show that the two strategies of adjusting the travel impedance on edges and improving the attraction node utilities can both change people's travel patterns, while the latter a direct and intuitive effect on introducing tours and visits to the improved area(s).

Our non-aggregate model of tourist travel behavior overcomes some inherent shortcomings of conventional travel demand models using statistical approaches. Ordinary travel demand models often use statistical approaches to predict trips on an aggregate level, e.g. using Markov Chain's theory to estimate the transition probability matrix. Such statistical approaches are of limited help for travel demand management due to the absence of an analytical model in representing people's travel behavior, i.e. they cannot predict the change in people's travel patterns if any change is introduced to the network. Moreover, trips in such modeling context are considered independent from previous ones, nor is travel history accounted for.

This research complemented above studies by taking a tour-based non-aggregate approach to model tourists' behaviors in making tours. In contrast to trip-based models, we see the change in trip distribution as an aggregated result from the shift of each individual's travel pattern. Therefore, our simulation system does not only estimate the change in the visit frequency for each attraction area, but also look at how people tend to change their travel patterns. As such, we can predict the change in the proportion of different origins that generate trips to current area, as well as the shift in the destinations among all outgoing trips. For example, the attraction might attract tourists that are supposed to visit nearby attractions, or there might be new trip chains generated based on the current destination.

8.2 Limitations

Below are some limitations in this study we would like to claim and several suggestions for further research.

First, the data used in our research is obsolete as we obtained it from a survey conducted in 2006. A lack of newer data causes problems like the data has a low resolution in describing the spatial-temporal trajectories of each tourist. Destinations are defined as attraction areas in the survey, which encloses multiple POIs at a time. The intrinsic utilities of the areas are aggregated from enumerating the POIs that belong to each category, accounting for both the quality of that POI as well as the popularity, indicated by the number of reviews attached to that place.

The preference prediction has some defects in nature as well. When performing the clustering of tourists by their travel purpose, we found the survey data limited the number of choices to a maximum of three, instead of asking respondents to answer "yes or no" at each question. Besides, the sample size is limited, nor there are enough features in the socio-demographic data which can be used to predict tourists' preference. As a result, the limited features tend to underfit the observed pattern which resulted in a relatively low R-square in the multinomial logistic regression.

Second, since we model the tourists' behaviors as a non-discrete choice problem, there does not appear to be a closed-form formulation between objective values and parameters. Because there are no analytical gradients for the objective function, we had to use heuristics to approach to the global optimum. Moreover, comparing the differences between predicted and observed paths is tricky, because there is no fixed metric to distinguish alternative paths except for paths that are the same as the observed one. Therefore, compared to a "right and wrong" problem, we must adjust the penalty function based on the type of travel patterns we want to predict and the type of error we emphasize.

Finally, the transport network is static with fixed edge properties regarding travel. The time dimension is not considered either to simplify the model formulation. Ideally, the transport network should be dynamic, which includes road traffic and time-dependent travel time. Besides, we acknowledge that people change their behaviors and the intrinsic utility of the attraction areas may also change throughout the day. For example, the utility of "gourmet" will increase during lunch or dinner. However, as our interest is mainly the behavioral model, we leave a time-dependent graph network as part of future work.

8.3 Future work

Data used for our model calibration are from a tourism survey conducted by the 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.

For part of the future work, we hope to include the time dimension in the tour prediction and consider a time-dependent network that takes into consideration the road traffic and dynamic travel time. Ideally, we hope to develop a combined destination and duration choice model such that the activity time at attractions are also predicted.

Although we model the decision making a deterministic process where tourists discriminate against the possible paths and pick up the one with the highest utility, the probabilistic choice theory suggests that people make choices under uncertainty. To confront this problem, we can adopt the concept of weighted models or mixing distribution, which means the weighted average of several functions in the statistics literature. Specifically, we can estimate the parameters stochastically, e.g. the density of parameters can be specified to be normal with mean β and covariance W, or their distribution can be discrete, with parameters taking a finite set of distinct values. Then, similar to the concept of mixed logit which is a mixture of the logit function evaluated at different β's with $f(\beta)$ as the mixing distribution, we can predict the trip table as a weighted average of the predicted trips frequency evaluated at different values of estimated parameters, with the weights given by the density $f(\beta)$. If so, our deterministic approach for predicting the trips will be a special case where the mixing distribution $f(\beta)$ is degenerate at fixed parameters $\beta *$, which is the best set of parameters estimated so far.

Finally, although we evaluated the quantitative effects of various TDM strategies in the simulation, we did not optimize trip distribution through a combination of those strategies. Our next step is to develop a problem formulation, define the objective to optimize, and solve the optimization problem by introducing several strategies together to optimize the trip distribution statistics.

Reference

[1] Koens, K., Postma, A., & Papp, B. (2018). Is overtourism overused? Understanding the impact of tourism in a city context. Sustainability, 10(12), 4384.

[2] Duignan, M. (2019). 'Overtourism'? Understanding and Managing Urban Tourism Growth beyond Perceptions: Cambridge Case Study: Strategies and Tactics to Tackle Overtourism.

[3] Kyoto City (2017). Total survey on tourism in Kyoto. https://www.city.kyoto.lg.jp/sankan/cmsfiles/contents/0000222/222031/28chousa.pdf.

[4] Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management, 29(3), 469-486.

[5] Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place?. Tourism Management, 59, 267-280.

[6] 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.

[7] Bhat, C. R., & Koppelman, F. (1999). Activity-Based Modeling for Travel Demand. Handbook of Transportation Science. edited by RW Hall, Kluwer Academic Publishers, Norwell, Massachusetts, 35-61.

[8] Chu, Z., Cheng, L., & Chen, H. (2012). A review of activity-based travel demand modeling. In CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient (pp. 48-59).

[9] Castiglione, J., Bradley, M., & Gliebe, J. (2015). Activity-based travel demand models: A primer (No. SHRP 2 Report S2-C46-RR-1).

[10] Ilägcrstrand, T. (1970). What about people in regional science?. In Papers of the Regional Science Association (Vol. 24).

[11] Kitamura, R., & Fujii, S. (1998). Two computational process models of activity-travel behavior. Theoretical foundations of travel choice modeling, 251-279.

[12] Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: promises, progress and prospects. International Journal of Urban Sciences, 18(1), 31-60.

[13] Bowman, J. L., & Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. Transportation research part a: policy and practice, 35(1), 1-28.

[14] Ben-Akiva, M. E., & Bowman, J. L. (1998). Activity based travel demand model systems. In Equilibrium and advanced transportation modelling (pp. 27-46). Springer, Boston, MA.

[15] Miller, H. J. (2016). Time geography and space–time prism. *International encyclopedia of geography: People, the earth, environment and technology*, 1-19.

[16] Auld, J. (n.d.). Introduction to Activity-Based Modeling Joshua Auld Transportation Research and Analysis Computing Center Argonne National Laboratory November 28, 2011. - [PPTX Powerpoint]. Retrieved January 10, 2020, from https://fdocuments.in/document/introduction-to-activity-based-modeling-joshua-auld-transportation-research.html

[17] Ben-Akivai, M., Bowman, J. L., & Gopinath, D. (1996). Travel demand model system for the information era. Transportation, 23(3), 241-266.

[18] Bowman, J. L., Bradley, M., Shiftan, Y., Lawton, T. K., & Ben-Akiva, M. (1999). Demonstration of an activity-based model for Portland. In World Transport Research: Selected Proceedings of the 8th World Conference on Transport ResearchWorld Conference on Transport Research Society(No. Volume 3).

[19] Kitamura, R., Pas, E. I., Lula, C. V., Lawton, T. K., & Benson, P. E. (1996). The sequenced activity mobility simulator (SAMS): an integrated approach to modeling transportation, land use and air quality. Transportation, 23(3), 267-291.

[20] Cullen, I., & Godson, V. (1975). Urban networks: the structure of activity patterns. Progress in planning, 4, 1-96.

[21] 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.

[22] Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. Annals of tourism research, 33(2), 403-423.ISO 690

[23] Becken, S., Simmons, D., & Frampton, C. (2003). Seg-menting Tourists by their Travel Pattern for Insights into Achieving Energy Efficiency. Journal of Travel Research,42(1), 48-56. Doi:10.1177/0047287503253938

[24] Garcia, A., Arbelaitz, O., Linaza, M. T., Vansteenwegen, P., & Souffriau, W. (2010, July). Personalized tourist route generation. In International Conference on Web Engineering(pp. 486-497). Springer, Berlin, Heidelberg.

[25] Gavalas, D., Konstantopoulos, C., Mastakas, K., & Pantziou, G. (2014). Mobile recommender systems in tourism. Journal of network and computer applications, 39, 319-333.

[26] Umanets, A., Ferreira, A., & Leite, N. (2014). GuideMe–A tourist guide with a recommender system and social interaction. Procedia Technology, 17, 407-414.

[27] 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.

[28] 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.

[29] 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.

[30] Abbaspour, R. A., & Samadzadegan, F. (2011). Time-dependent personal tour planning and scheduling in metropolises. Expert Systems with Applications, 38(10), 12439-12452.

[31] Butt, S. E., & Cavalier, T. M. (1994). A heuristic for the multiple tour maximum collection problem. Computers & Operations Research, 21(1), 101-111.

[32] 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.

[33] Sasaki, K. & Nishii, K. (2010). A process model of destination set choice in tourist excursion with combination utility of the destination set.

[34] Kinoshita, Y., & Yokokishizawa, H. (2008, October). 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.

[35] Hamming distance. (2019). Retrieved from https://en.wikipedia.org/wiki/Hamming_distance.

[36] 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.

[37] 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.

[38] Algorithmic techniques for the Tourist Trip Design Problem (Master's thesis, Πανεπιστήμιο Πειραιώς).

[39] 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.

[40] 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.

[41] Levenshtein distance. (2020, February 8). Retrieved from https://en.wikipedia.org/wiki/Levenshtein_distance

[42] Road Bureau, Ministry of Land, Infrastructure Transport and Tourism. (2010, November 25). 道路事業の評価手法に関する検討委員会. Retrieved February 3, 2020, from https://www.mlit.go.jp/road/ir/ir-council/hyouka-syuhou/4s.html

[43] Johnston, E. (n.d.). Overcrowded site and packed buses main complaints of Kyoto tourists as visits hit record levels. Retrieved from https://www.japantimes.co.jp/news/2018/07/05/national/overcrowded-sites-packed-buses-main-complaints-kyoto-tourists-record-number-visitors-flock-city/#.Xi_vQWj7SUk