Tailor-made Selection of Policy Measures for Households Based on the Detailed Attributes by Segmentation Approach with Decision Tree Analysis

Noriko Otani, Yusuke Fukuoka, Nao Sugiki and Kazuaki Miyamoto

Abstract

In the situation of an ageing and population-decreasing society like in Japan, it is necessary for a local government to manage policy measures more efficiently and effectively in order to realize its policy objectives. Since the diversity of households in such a city is becoming wider than ever, it is necessary to consider their attributes in more detail in implementing a policy measure. The purpose of the present study is to develop a method to select a set of policy measures for each household based on its detailed attributes like tailor-made medicine for a patient. In order to find target segments of households to which a policy measure is effective, the decision tree approach is employed for data mining. Additional analysis for selecting policy measures to each household is conducted based on the information obtained by a questionnaire survey. The results of application to households in Toyama City, Japan show the usefulness of this approach.

Noriko Otani: Faculty of Informatics, Tokyo City University (otani@tcu.ac.jp)
Yusuke Fukuoka: Osaka Branch Engineering Division, Katahira & Engineers Inc. (fukouka@katahira.co.jp)
Nao Sugiki: Transportation Department, Docon Co., Ltd. (ns1491@docon.jp)
Kazuaki Miyamoto: Faculty of Urban Life Studies, Tokyo City University (miyamoto@tcu.ac.jp)

1. Introduction

In the situation of an aging and population-decreasing society such as in Japan, it is not affordable for a local government to keep providing the households in the city with the same level of public services by means of a conventional approach. The local government faces financial difficulties caused by less income from decreasing tax payers and more expenditure for increasing elders. Therefore, it is necessary for a local government to manage policy measures more efficiently and effectively in order to realize its specific policy or alleviate facing problems in the city.

In accordance with aging and population-decreasing, the diversity of households in the city is becoming wider than ever in such a society. Services demanded by a household depend on attributes in terms of both type and quantity. In addition, in the case of a local government implementing a policy measure for a specific policy purpose, some households react desirably from the viewpoint of the policy, whereas others may neglect it or react even undesirably. This suggests that it is necessary to consider household attributes in more detail in implementing a policy measure. In this study, a set of detailed attributes of a household is referred to as household micro-data.

On the other hand, a market segmentation approach to promote sales of a specific product is popular in commercial marketing. It employs a data mining approach to find target segments of customers based on their attributes. If the product is replaced by a policy measure, it may be interpreted that the local government finds the target segment that may react desirably to the policy measure. In addition, in the field of medical care, tailor-made approach programs based on genetic characteristics are being developed in many places. According to the "Tailor-made Medical Treatment Program (2015)", one of the leading programs, the objective is to develop new medications and treatments through elucidating the specific causes of diseases and to realize personalized medicine where selections of drugs and treatments can be appropriate and optimized based on genetic characteristics to avoid their side effects. If genetic characteristics of a patient are replaced by household micro-data, it is interpreted as that tailor-made selection of policy measures for the household may be effective for urban planning and implementation like tailor-made medical treatments for a patient. Thus, the purpose of the present study is to develop a method to select policy measures for each of the households based on the micro-data.

In this study, Toyama City, a city in Japan with a population of around a half million, was selected as a case study area, and a questionnaire survey was conducted. Toyama City is famous for promoting a so-called "compact

city" policy in Japan. Therefore, the policy considered in this study is the "compact city" or "transit oriented development" initiative.

The present paper is structured as follows. First, publications regarding social marketing, which is an extension of a marketing approach to social problems, were reviewed to confirm the state of the arts related to the basic idea of this study. Then, in order to find a target segment of households for the compact city policy, the decision tree approach is employed among data mining methods, and is applied to households in Toyama City. In the following portion, an additional analysis for selecting policy measures each of the households based on the information obtained by the questionnaire survey. Finally, after confirming the usefulness of this approach, results and further research are summarized.

2. Market Segmentation

Market segmentation is a kind of marketing strategy that aims to approach a particular consumer group intensively. In the strategy, the consumer market for a specific product or service is divided into some groups of consumers who show the same reaction to the product or service and have common properties. The divided group is called "market segment," and the group that is targeted to promote the product is called "target segment." Table 1 shows the bases that are used to segment the consumer market.

Commercial marketing strategies are getting adopted in governments gradually as social marketing. The society in which citizens are living is considered as a market, while citizens are considered consumers. Based on the identification of the citizens' various requirements, appropriate and effective services can be provided.

In the following sections, some examples of adopting market segmentation to public services, or a social marketing approach, are illustrated.

Table 1. Segmentation Bases in Consumer Markets

Base	Examples
Geographic	region, scale of metropolitan area, urban/suburb/rural area, climate
Demographic	gender, age, family size, lifecycle
Socioeconomic	income, occupation, religion, education
Psychographic	activities, lifestyle, interests, opinions
Behavioral	purchase records, royalties, benefits, occasions

2.1 Marketing Water Services

Njiru and Sansom attempted to provide the rationale for using a strategic marketing approach in the urban water sector in Mombasa, Kenya (Njiru & Sansom 2001). In this study, target segments were set with appropriate service options at sustainable water charges that consumers are willing to pay for. Options for equipment are house connections, yard taps, and water kiosks, with or without storage tanks, at appropriate price levels, whereas other options are payment options. This paper concludes that strategic marketing offers a systematic and flexible approach to providing improved and sustainable services to all consumer groups or segments, including the poor.

2.2 Increasing Transit Ridership

The Metropolitan Transit Development Board and the San Diego Association of Governments used market segmentation to identify potential transit customers. After identifying transit corridors with low ridership, MTDB used this approach to target potential customers based on their demographic profile. Surveys of riders and non-riders in the corridor further provided the transit agency with an opportunity to ad-dress concerns – for example, safety and security – in specially tailored promotional materials. Besides this case, many public transport operators have conducted a segment marketing approach. However, the best chance for public transport to regain market share is in the adaptation of land-use policies favorable for public transport (Laconte 2002), as the segment marketing approach naturally has its limitations.

2.3 Family Planning

There have been case studies of market segmentation approaches to family planning in many developing countries, such as Rumania, Bangladesh, Bolivia, Egypt, and Jordan, to name a few. In an application in Nicaragua (Winfrey & Lacayo 2013) public- and private-sector users are segmented into different sub-groups, each with its own pro-file, and make recommendations about how best to meet the needs of each group. The document highlights not only demographic and economic variations but also differences in values and attitudes, all of which are key drivers of family planning demands.

3. Household Segmentation

We propose the selection method of policy measure alternatives in ur-ban planning. In our proposed method, the households in a city are considered as consumers in a market. The policy considered in this study as an example is the so-called "compact city" or Transit Oriented Development initiative. Desirable actions of households from the policy viewpoint in response to a policy measure are considered as a target product. Like the market segmentation strategy, households are divided into some segments, and appropriate policy measures for the target segment are selected. In order to make segments whose house-holds show the same reaction to the policy and have common proper-ties, detailed household attributes are used for segmentation. We call the procedure "household segmentation." This method makes it possible to provide each segment of households with a tailor-made selection of policy measures and public services.

3.1 Household Micro-data

Data used in household segmentation are called "household micro-data." Household micro-data are different from macro-data or mezzo-data such as the population in the census data or statistical data published by local governments. The detailed data for each household, such as the number of members, location, car ownership, housing type, and members' properties (age, gender, relationship to house-holder, and so forth) are contained in household micro-data. Although it is more difficult to acquire micro-data than macro-data or mezzo-data, more accurate and detailed analyses will be possible by using micro-data. In addition, the estimation method of household micro-data in a city has been developed by Sugiki et al. (2012).

3.2 Decision Tree

Not only statistical methods but also machine learning techniques, such as the decision tree, association rule, neural network, and support vector machine are used for data mining. In order to obtain more effective rules in data mining, it is important to use an appropriate technique according to data form, data structure, objective, and so on.

Household segmentation needs to make household segments which have common properties, show the same reaction to a policy measure, show the same level of satisfaction for a situation, or show the same intension for future behavior. Even if some households show the same reaction, level of

satisfaction, or intension, households with different properties are segmented to different segments. In order to know what kind of households react to a specific policy measure desirably or undesirably, characteristics of each segment should be revealed. Therefore, the classifier that represents the criteria of classification definitely, such as the decision tree, is more suitable for this study than the classifier whose process is unknown, namely the black box. In addition, it is possible to make a decision tree contain multiple rules. Then, the decision tree is adopted in household segmentation.

Decision tree is a technique to represent classification rules in a tree structure. A decision tree is generated using some examples that consist of some attribute values and the class that the example belongs to. An attribute is assigned to each non-terminal node, and a condition of the attribute is assigned to each label of edge. Each terminal node indicates a class, namely a classification result. One pair of a non-terminal node and an edge that connects the non-terminal node means a one-branch condition in tracing.

After tracing the decision tree from the root node to the terminal node according to attribute values of an example, the reached terminal node indicates the class that the example belongs to. The path between a root node and a terminal node represents a classification rule. The antecedent is the conjunction of all branch conditions that is represented by a pair of a non-terminal node and an edge, and the consequence is the class of terminal node.

In generating a decision tree, values of nodes and labels of edges are decided to classify several examples as accurately as possible. CART (Breiman et al. 1984), ID3 (Quinlan 1986), and C4.5 (Quinlan 1993) are representative systems for generating a decision tree. These are able to classify unknown examples that were not used for generating trees by pruning, boosting, and so forth.

Otani (Otani et al. 2004) proposed a system SESAT for generating simple and accurate decision trees based on symbiotic evolution that is one of the evolutionary computation algorithms. The simplicity of a tree structure is as important as accuracy, especially in consideration of the comprehensibility to a human. In addition, SESAT is strong in processing numerical attribute values. As these characteristics are suit-able for this study, SESAT is used for household segmentation in this study.

3.3 Segmentation Procedure

It is possible to make a decision tree contain multiple classification rules. Fig. 1 shows an example of a decision tree that indicates rules for classifying

households into two classes, for example, to stay at or move from their present house. The numerical value under the terminal node means the number of households that reached the terminal node after tracing from the root node. This tree contains the following two rules about households that agree to the policy.

1. They own one car or less, and there are two children or more.
2. They own two cars or more, there are four members or more, and their housing is 2 km or further from the nearest station.

However small the households are that satisfy the second rule, the first rule is more general than the second one. Implementing policy measures to more residents are more effective. Therefore, the path that the most households trace from the root node to the terminal node should be focused on, and extracted as a rule of the target segment.

As SESAT searches the optimal solution using random numbers, sub-optimal solutions can be obtained. Considering the diversity of households and noise data, multiple rules may cover various attributes. Therefore, generating a decision tree is repeated, and characteristics of the segment are generated by combining some rules obtained.

The reactions to a policy, the levels of satisfaction for a situation, or intentions for future behavior are set to classes. The segmentation procedure is as follows.

1. Generate a decision tree
2. Extract a rule that is satisfied by the most households in the rules of the target class
3. Repeat 1 - 2 several times
4. Combine rules and generate characteristics of the target segment

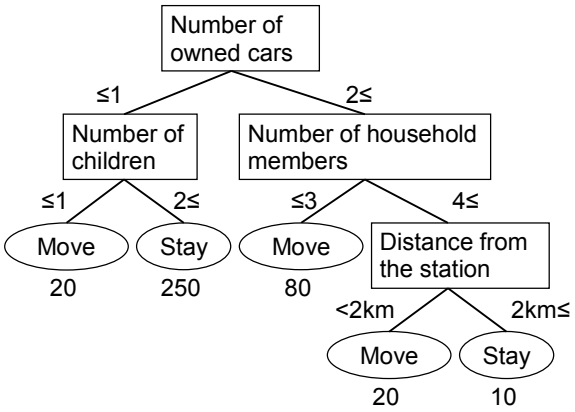

Fig. 1. Example of Decision Tree

4. Application for Selecting Policy Measures in Toyama City

In this section, the household segmentation is applied for selecting policy measures in Toyama City, a local city that is working hard on the so-called compact city policy in Japan. The policy measure that makes the population concentrate in the city center and the vicinity of transport stations is promoted. For annotation, this application was conducted only for confirming execution possibility of our proposed method. The discussed policy measures are never relevant to Toyama City's decision.

4.1 Target Policy and the Measures

We targeted the policy measure "the project to promote living in the center of the town" (Toyama 2005) based on the compact city policy. This policy aims to concentrate the population on the public transport corridor. Further, the subsidy system for living in the center of the town was also implemented. Subsidies are granted to households that buy their housing in the designated area or move to the designated area. The application in this section aims to realize this concentration of the population.

The area that is 2 km far from the nearest station is defined as a walkable area from the station. The four types of move action are labeled to either desirable or undesirable as follows.

- Stay inside or Move from *inside* to *inside* of the walkable area from station (desirable)
- Move from *inside* to *outside* of the walkable area from station (undesirable)
- Move from *outside* to *inside* of the walkable area from station (desirable)
- Stay outside or Move from *outside* to *outside* of the walkable area from station (undesirable)

4.2 Data

Questionnaires collected by the government of Toyama City and our research group were used as the household micro-data. The outline of questionnaires is as follows.

- Period: from Dec 2011 to Jan 2012
- Area: Toyama City in Japan

- Target households: 14,073 households sampled randomly from 140,743 households
- Sent questionnaires: 14,073
- Collected questionnaires: 5,089 (36.2%)
- How to send and collect: mail (reply envelopes were enclosed)

Questionnaires with serious missing values were discarded, and 1,326 questionnaires were used. Part of the questionnaire items is as follows.

- Members' properties (age, gender, occupation, etc.)
- Present housing information (selection reason of the housing, complaint, worry, etc.)
- Wish to move in the future
- Transportation in daily life
- Wish for transportation
- Recognition of policies in Toyama City
- Wish to use the subsidy system for living in the center of the town

4.3 Classes and Attributes

In order to know what kind of households live or move on the public transport corridor, the indicators "Wish to move in the future" and "Wish to use the subsidy system for living in the center of the town" were focused on. The responses to these two prompts were combined to classes as follows. The number of households concerned is given in parentheses.

- Class 1: "Wish to move," "Wish to use the subsidy system" (272 households)
- Class 2: "Wish to move," "Do not wish to use the subsidy system" (95 households)
- Class 3: "Do not wish to move" or "Undecided" (959 households)

If the obtained rules are easy to understand, it becomes easy to plan, select, and explain policy measures. The following items were selected from questionnaire items, and were used as attributes.

- Household properties (average age of members, number of members)
- Household members
- Distance from the housing to Toyama station and the nearest station
- Past and present residential information
- Number of owned cars
- Experience of move

4.4 Characteristics of Target Segment

Characteristics of households that belong to Class 1 were obtained by the proposed method as follows.

- They live in an apartment for rent
- They live in the area that is 4 - 5 km away from Toyama station
- They live inside of a walkable area from the station
- They are young couples in their 20s - 30s, or families with one child who is a preschooler or an elementary school or junior high school student
- Their average age of household members is 30.0 years old
- They have an experience of moving
- They own one car
- The selection reason of the present housing is “housing expenses” or “distance from the workplace” or “convenience for shopping”
- They are satisfied that “shopping facilities” and “facilities” are located near their housing
- They are worried about “housing expenses” and “snow shoveling” in the future
- They plan to move to “their own detached house” or “condominium for rent” because of “buying a house” or “parents’ care” or “transfer or change of job” in the near future
- If a station or a bus stop is established near their housing, they change “frequency for children to go to preparatory school,” “frequency of eating out with friends,” “workplace,” “school,” and “place for eating out with friends”

The above result shows that these households are generally satisfied with their housing and wish to move to their own detached house. It is unclear whether they hope to move to the inside of a walkable area from the station or not. However, they live inside of the walkable area from the station, and they wish to use the subsidy system. Policy measures should be selected to control moving to the outside of walkable areas from the station.

4.5 Additional Analysis

To identify the times and reasons of moves, characteristics of households that have experiences of moving were analyzed. Classes of the decision tree are the following five classes. The target class is Class B.

- Class A: Desirable move from *inside* to *inside* (842 households)
- Class B: Undesirable move from *inside* to *outside* (204 households)

- Class C: Desirable move from *outside* to *inside* (119 households)
- Class D: Undesirable move from *outside* to *outside* (107 households)
- Class E: No experience of move (440 households)

Although decision trees were generated using the same attributes as Class 1 – 3, the accuracy was at about 60 percent. Therefore, characteristics of Class B households were extracted by comparing data of Class 1 and Class B. The situations when Class B households have moved were as follows.

- When the householder earns enough money to support the household, they have no need to use public transport because of their cars, and they have enough time
- When they have complaints about their present housing, and they are attracted to the nature that is not in the center of the town
- When they need to live with their parent(s)
- When they want to live in their own detached house, not an apartment house

In addition, the average age of household at the time of move was calculated by subtracting "years of living at the present housing" from "present average age of household." Average, median, and standard deviations of calculation results (without minus values) and the average age of household in Class 1 are shown in Table 2.

Considering the largeness of standard deviation and wideness of the target age group, the average age of household at the time of move is estimated at 36.5 years old using the median value. Therefore, it is roughly estimated that Class 1 households will move after 6.5 years, or in 5 - 10 years.

Table 2. Average Age of Household

	Class B (at the time of move)	Class 1 (present)
Average	38.6	30.1
Median	36.5	30.0
SD	14.6	7.87

4.6 Planning and Selection of Policy Measures

In the previous decision tree analysis, target segments for promoting a compact city policy are identified for Classes 1 and B. Households belonging to the target segment of Class 1 wish to move, and wish to use the subsidy system. In addition, it is expected that they will move in 5 to 10 years from

now. For households belonging to Class 1, the questionnaire survey revealed that most of them did not know of the existence of the subsidy to live in the center area from the city before the survey.

As for households belonging to the Class B segment, they moved undesirably from the viewpoint of a compact city. This means that households who currently live inside of a walkable area from the station and belonging to the Class B segment are potential movers to the outside. In order to promote the compact city policy, policy measures should seek to make them stay inside.

Then, by setting these two segments as target segments, adequate policy measures should be selected from the viewpoint of a compact city policy. Candidate measures for promoting a compact city policy are listed below.

- Information or Enlightment Measures: To give them real information of lords they would have in the case of living outside of the area
- Economic Measures: To give them economic merits to stay inside the area
- Investment Measures: To provide them with more convenient public and private services at the present location; to develop better housing inside the area

As for the households belonging to the segment of Class 1, it can be said that intensive public relations of a subsidy system for them is an adequate and effective measure. This is a set of policy measures including information and economic measures. In order to conduct the public relations effectively, it is necessary to know where they are living now. Since the estimation method of household micro-data in a city is available as previously mentioned, it is possible to provide the city with the information.

As for households belonging to the segment of Class B, from the viewpoint of a compact city policy, measures to discourage them from moving to outside of the walkable area from the station should be selected. Among these candidate measures, the information or enlightment measure may be the most feasible.

In this study, the type and level of loads that households currently living outside of the area have are identified based on the questionnaire survey. Table 3 shows the type and level of loads of household members in their sixties living outside of the walkable area from the station, as well as their age when they moved to their present housing. Since this table represents factual information, it is expected that presenting the information to the households which belong to the Class B segment to some extent discourages them from moving to outside of the area.

Table 3. Load in living outside of the walkable area from station

		Age when moved to present housing (number of sample households)				
		20s (54)	30s (36)	40s (35)	50s (22)	60s (12)
Public transport	Access	○	○	○	○	●
	Frequency	●	○		○	
	Time	○	●	○	○	○
Daily life	Shopping			○	○	●
	Hospital	○		○	●	
	Snow shoveling	○	●	○	●	○
	Education	●			●	
Entertainment	Theatergoing	●	○	○	○	○
	Eating out	○	●	○	○	○
Economy	Housing expenses	●	○	○	○	○

●: Extremely large load ○: Large load

5. Concluding Remarks

This study developed a method to select policy measures to households based on the micro-data. The decision tree analysis successfully identified households which belong to target segments from the policy viewpoint. Although examples presented in this paper are limited, the approach is applicable to other kind of target segmentation as applied in commercial marketing.

As for policy measure selection, we admit that it requires further considerations, for the alternatives discussed in this paper are still very limited. However, the usefulness of a tailor-made approach in selecting policy measures is presented with two target segments of households.

As previously mentioned, an estimation method of household micro-data for a whole city has been developed and applied to Toyama City to generate all the households in the city. This means that the distribution of households belonging to the target segments is known. Therefore, it is possible to effectively implement the policy measures in the city.

Since the questionnaire survey included data used for this study that was originally conducted for building a household micro-simulation model, the survey is not very adequate for this study. It is requested to conduct an original questionnaire survey that is designed for this research purpose in the future.

References

Breiman L, Friedman JH, Olshen RA, & Stone CJ (1984) Classification and regression trees. Wadsworth & Brooks

Laconte P (2002) Smart segments for urban public transportation: an international survey of practices. Japan Railway & Transport Review 32: 4–11

Njiru C & Sanson K (2001) Strategic marketing of urban water services. Proceedings of the 27th WEDC Conference, pp. 256–259

Otani N & Shimura M (2004) Generating the simple decision tree with symbiotic evolution. Journal of Japanese Society for Artificial Intelligence 19(5): 399–404

Quinlan JR (1986) Induction of decision trees. Machine Learning 1(1): 139–159

Quinlan JR (1993) C4.5: Programs for machine learning, Morgan Kaufmann

Sugiki N, Vichiensan V, Otani N, Miyamoto K (2012) Agent-based household micro-datasets: an estimation method composed of generalized attributes with probabilistic distributions from sample data and available control totals by attribute. Asian Transport Studies 2(1): 3–18

Tailor-made Medical Treatment Program. (2015) Program objective. http://www.biobankjp.org/english/plan/purpose.html. Accessed 26 Feb 2015

Toyama City (2005) Project for promoting to live in the center of the town. http://www.city.toyama.toyama.jp/toshiseibibu/toshisaiseiseibika/jyutakuseisaku/kyojusuishinjigyo.html. Accessed 22 Feb 2015

Winfrey W & Lacayo Y (2013) Nicaragua family planning market segmentation analysis, prepared for the enhancing equity and sustainability of public-sector family planning project. Futures Institute, PATH. http://www.path.org/publications/detail.php?i=2259. Accessed 26 Feb 2015