Empirical Study of Indecision Behavior in Customer's Purchase Process using EC Data

Takashi Namatame†
Faculty of Science and Engineering
Chuo University, Tokyo, Japan
Tel: (+81) 3-3817-1932, Email: nama@indsys.chuo-u.ac.jp

Tomoaki Tabata
School of Information and Telecommunication Engineering
Tokai University, Tokyo, Japan
Tel: (+81) 3-3441-1171, Email: tabata@tokai-u.jp

Abstract. In marketing study, consumer behavior is one of the main topics. Consumer behavior have some processes, i.e. accessing to and organizing information, evaluating alternatives, deciding purchase item and post-purchase evaluating. In this study, we focus on the customer's state before purchase. Frequently, many customers cannot decide purchase item soon, so they are in indecision state for several times. In this state, they have some candidate alternatives, and after then are specifying. To focus on these phenomenon is important for consumer's purchasing behavior study and marketing strategy, because if we can approach to these customers effectively, marketing effect are rising and we may get high customer loyalty. In our previous study, we propose some indices to represent indecision state for each customer at each time. In this study, using real marketing data of an EC site, we elucidate customer's potential purchase behavior before purchase. For our analysis, we use purchase record, clickstream on this EC site and customers' demographic data. We analyze the relationship the intensity of our indecision index and the other data, then we find some important knowledge about specifying process of purchase. Using our result, store or manufacturer can do efficient approach to their customers.

Keywords: consumer behavior, indecision, ID-POS data, web access log data

1. INTRODUCTION

Consumer behavior is one of the central issues in marketing study. Because marketing aims to get the customer satisfaction through marketing activities. In recent marketing study, analysis are done with some data or record. Mainly, we often have used POS data of ID-POS data that are past purchase records of customers. However, POS or ID-POS data is the result of purchase decision, so we can grasp the process of decision from them directly. In recent year, electronic commerce (EC) are growing, so the purchasing from EC site is become ordinary. On EC site, server records the clickstream of all visitors as web access log. Through the web access log analyzing, we can grasp the purchase process of each customer.

When consumer purchase some items, many of them cannot decide what item should be purchased, soon. In these cases, they are in "indecision." In case of real retail shop, sales person can approach them and induce to purchase through listening their needs or problem about purchase decision. However, on EC site, it is very difficult to grasp customers' indecision because we cannot see the customers' faces. So the customers are left from EC shop.

If we can find the indecision situation, and grasp how the customer cannot decide, then we may resolve their indecision and induce to purchase. To do this, a conceptual and realistic model is needed.

So far, in consumer behavior no compensative discrete choice model were utilized to indecision study. For example, Montgomery (1983) assume that choice behavior is to explore an alternative which is supreme characteristics. Then he explains indecision situation to make loop the consumer behavior in no compensatory discrete choice model's algorithm.

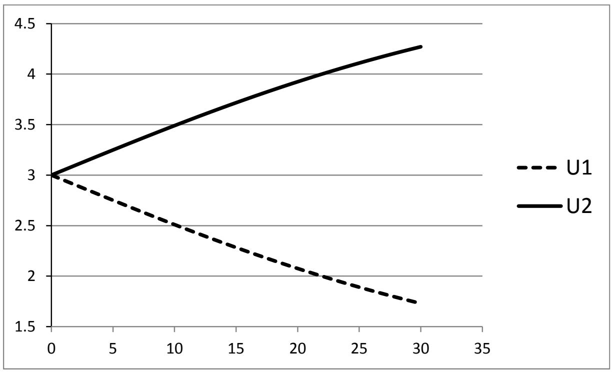

Figure 1: no indecision case

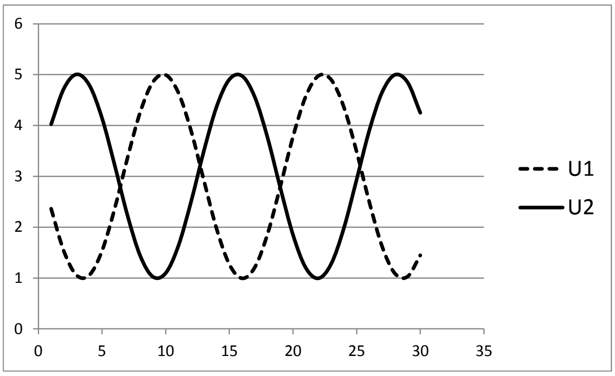

Figure 2: indecision case

On the other hand, the other method has been tried. Consumer decide the purchase item through two steps, that is, in the first step consumer narrow down some candidate items, then in the second step, they do the final decision using a compensatory model (e.g., Lussier and Olshavsky, 1979, Gensch, 1987).

To capture an indecision phenomenon in real market, we need to expand some compensatory model to explain consumers' indecision process. For this matter, we have some expanded compensatory discrete choice model (Tabata et al. 2015, 2015). Moreover, we have analyzed some real cases. However, we remain some problems for them. For example, we have not revealed the process of indecision or trigger event or timing from indecision to decision.

So, in this study, we analyze real purchase records and web access log data on an EC site using our models, explain the indecision phenomena in detail. By this study, we can build a concierge system to reduce the indecision of a consumer on EC site. First, we define a probabilistic choice model with dynamic utility. Next, we propose some indices of indecision status. Then we demonstrate an example of our model using real EC site purchase and access data, and consider our proposed method. Finally, we summarize our study and point out our future work.

2. DYNAMIC UTILITY MODEL

Compensatory models with random utility like multinomial logit model are main stream to explain consumer choice. These multiple attribute attitudes models are based on "expectation-value" theory by Fishbein (1963).

Let U_i is the attitude (utility) of a consumer for item (or category) i, x_{ij} is the evaluation level for item i, attribute j, a_j is the weight (importance) for attribute j. Then, utility U_i is expressed as the next equation

$$U_i = a_1 x_{i1} + a_2 x_{i2} + \cdots + a_m x_{im} \tag{1}$$

Tabata el al. (2015) have expanded equation (1) to be dynamic. Assume x_i is not changed in short time, we can express the instantaneous utility $u_i(t)$ and the cumulative utility $U_i(t)$ as below.

$$u_i(t) = \sum_{j=1}^{m} a_j(t) x_{ij} \tag{2}$$

$$U_i(t) = \int_{t-s}^{t} u_i(s) \mathrm{d}s \tag{3}$$

Where s is a parameter of reviewing time. When the utility functions are given above, a probability of choosing item i at decision-making time T is defined the following.

$$P_i(T) = \frac{U_i(T)}{\sum_\iota U_\iota(T)} \tag{4}$$

In equation (2), we can consider some situations with respect to the shape of $a_j(t)$. When $a_j(t)$ is increasing or decreasing monotonically as Figure 1, then utility value is converging. However, $a_j(t)$ have some periodic fluctuation or not converged, then indecision situation may be occur.

In this model, we obtain the value of $a_j(t)$ from data sequence. $a_j(t)$ is a function with respect to t, so we need to assume a function and estimate the parameters. Moreover, we assume that this model have dynamic change by time, so we need to check that $a_j(t)$ is changed with respect to t. From next section, we estimate $a_j(t)$ from real data, then we observe the time change of $u_i(t)$ or $U_i(t)$.

3. ANALYSIS

In this section, we demonstrate an analysis using real purchase and web access log data. Through our analysis, we consider the characteristics of indecision phenomena.

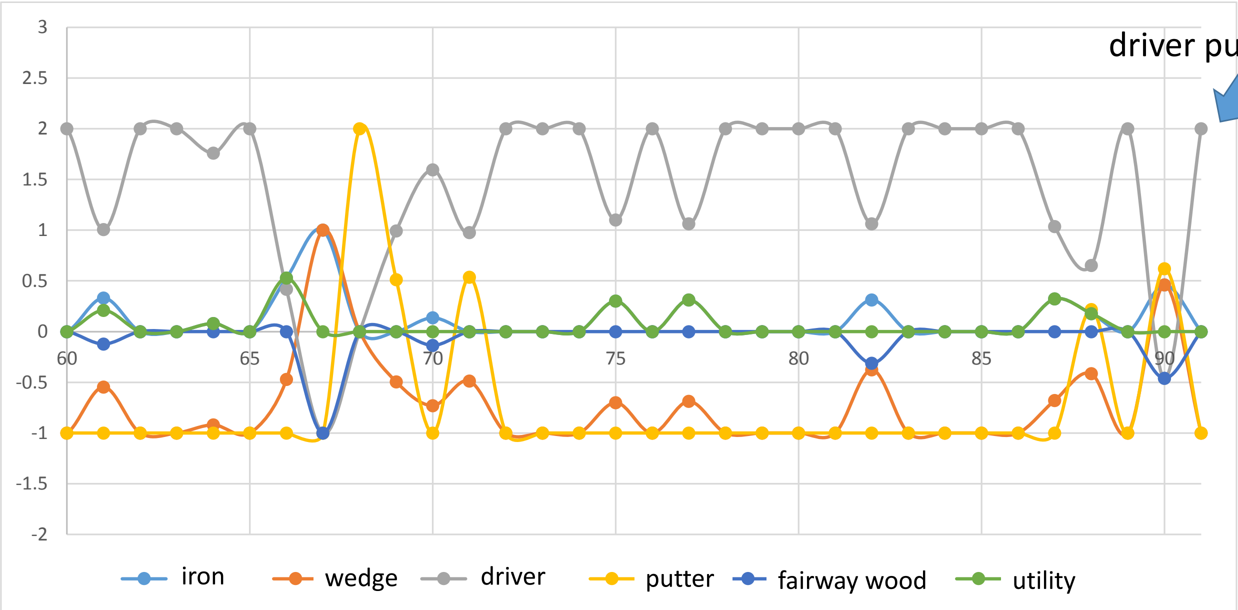

Figure 3: an example of dynamic utility path

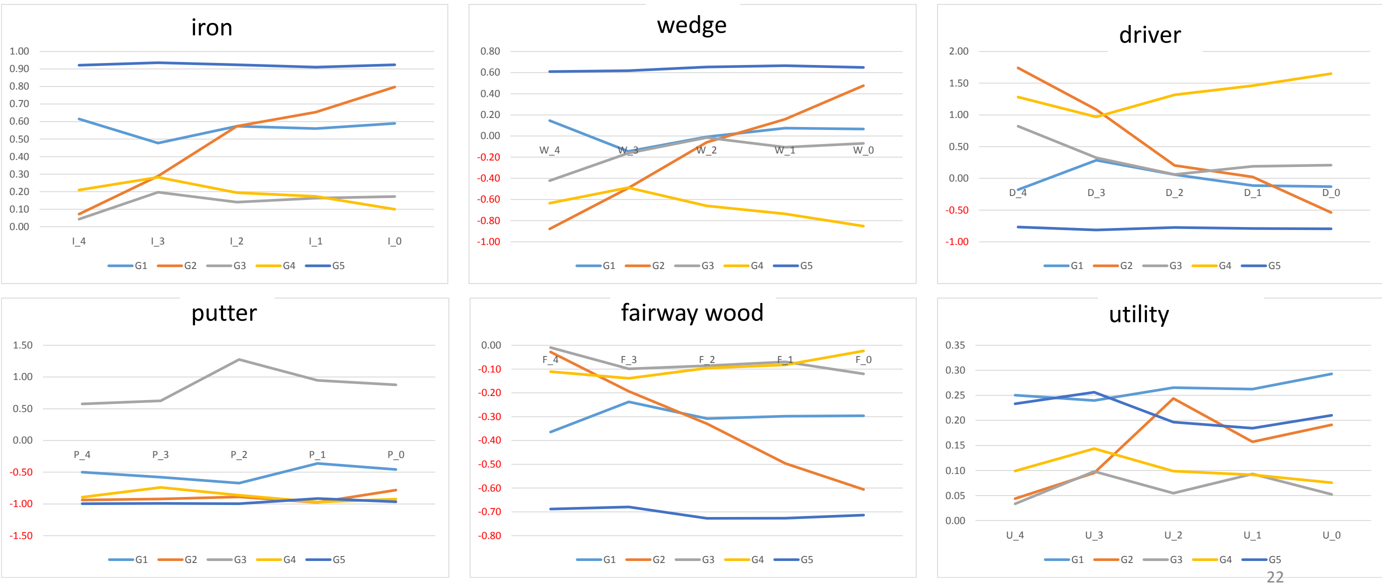

Figure 4: clusters of dynamic utility

3.1 DATA DESCRIPTION

In this study, we use a golf item selling EC site of Japan, and we use item purchase records and web browsing log in the EC site. The summary of data is below.

Term: 1/1/2012 – 12/31/2012
No. of customer: 300 (purchased golf club)
Categories: Iron, Wedge, Driver, Putter, Fairway-wood, Utility
Web access log: browsing category page

3.2 WEIGHTS $a_j(t)$ AND UTILITIES $u_i(t)$

We analyze the data to grasp the change of $a_j(t)$ and $u_i(t)$ through below steps.

We call *session* that is the sequence data from landing on a site page to leaving off the other site for each customer at a time. In each session, we consider that each item browsing is as an item choice behavior, then we estimate $a_j(\tau)$ (τ is each item browsing times in a session). Then we obtain the sequence of $a_j(t)$ and $u_i(t)$. Concretely, for example, there are some sessions, that is $SS_1, \cdots, SS_s$ for a consumer. In each session, item browsing behavior is as a choice behavior, then we estimate the optimal $a_j(SS_1), \cdots, a_j(SS_s)$ using multinomial logit model. In this study, the maximum session time is 120 minutes. Then, we obtain $u(SS_1), \cdots, u_i(SS_s)$ when $a_j(SS_1), \cdots, a_j(SS_n)$ are substituted in equation (2).

We set the problem that one item is chosen from 6 kind of golf clubs. We assume that "effect of tee shot (A)", "effect of second shot (B)", "effect of green approach (C)" and "effect on green (D)" as the attributes of golf club. The value of attributes are shown in Table 1. As shown this figure, this customer purchased a driver after quickly utility change, so the customer might be under indecision just before purchase.

Table 1: items and attributes

Item	A	B	C	D
Iron (IR)	2	3	3	1
Wedge (WD)	1	2	3	1
Driver (DR)	3	1	1	1
Putter (PT)	1	1	1	3
Fairway wood (FW)	2	2	1	1
Utility (UT)	2	3	2	1

4. RESURTS OF UTILITY CHANGE

If we assume that utility is no time-dependent structure, then the rank of utility is not change. So, we conclude that utility is depended on time, so utility is the function of time. Figure 3 is an example of utility transaction about a customer.

Table 2 shows that accuracy in case of matching truly purchasing item and the maximum utility item at each time. As shown in Table 2, when we consider longer term, the accuracy rate is decreasing. It shows that the longer term customer has, the more indecision is caused.

It is costless to get information about items on EC site, so customer is easy to click many items and it may contain some useless clicks. However, when the time is closer to purchase, the accuracy rate are rising. So many customers click the purchase item at close of purchase time.

Figure 4 shows the change of $u_i(t)$ of a customer who purchased a driver at session 91. From this figure, we grasp that the customer has high utility about driver generally. However, from 65 to 70 sessions, the utility values of driver were not always highest, so it is seemed the customer was indecision situation.

As shown above example, using dynamic utility model, we can distinguish whether each purchase was whether systematically or not. Through these observations, we can consider a new marketing approach.

Table 2: terms and average accuracy

Terms	Ave. Accuracy
Only just before session	68.0%
Before 3 sessions	53.9%
Before 5 sessions	46.8%
Before 7 sessions	42.7%
Before 10 sessions	40.9%
Before 20 sessions	37.6%

To observe the utility change structure, we consider the change features of utility. We use 5 sessions' utilities of just before purchase. Then we do hierarchical cluster analysis using these data. Figure 4 is the 5 clusters cases for each item. All clusters have increase case and decrease case. When we change the number of clusters, we found increase and decrease cluster as same as 5 cluster case. This result shows the various types of customer exist and it seems realistic. Moreover, we found some cases when the utility of an item was increasing, one of the other item were decreasing. This phenomenon shows there are some confliction between the items. To use our indecision model, we can consider these realistic change of customer's psychological situation.

5. INDICES OF UTILITY CHANGE

As shown in the previous section, we have clarified some characteristics of dynamic utility change of purchase process. However, those changes are various, and thus we suggest some indices to categorize them.

Moreover, in this section, we propose some indices for utility change phenomena to compare difference among customers.

5.1 INDEX FOR DEPTH OF INDECISION

First, we define the *degree of indecision*. This index means whether customer touch various brands or not. To explain the depth of degree of indication, we refer entropy. Let P_i be the ratio of choice of item i, then the degree of indecision is the next equation.

$$D = -\sum_i P_i \log P_i \tag{5}$$

5.2 INDICES FOR PHYSICAL FACTORS

Second, we define two indices of heterogeneity. These indices are based on the correlation between brands, then we defined a customer's heterogeneity.

One is the *brand heterogeneity*, which means the difference of change of brands to an objective brand. Let $\xi_{i,\iota}^{(band)}$ is the correlation coefficient of time series utility of brand i and ι. Then the brand heterogeneity is defined the average of one minus correlation coefficients as the next equation.

$$RB_i = \frac{1}{2(n-1)} \sum_{\iota \in N \setminus i} \left(1 - \xi_{i,\iota}^{(Brand)}\right) \tag{6}$$

The range of RB_i is $0 \leq RB_i \leq 1$, if RB_i is approach to 1 then brand heterogeneity is high, and vice versa.

Another is the *attribute heterogeneity*. This index is the same sense of the previous brand heterogeneity. Let $\xi_{i\iota}^{(Attribute)}$ is correlation coefficient of attribute brand i and ι, then the attribute heterogeneity is defined as the next equation.

$$RA_j = \frac{1}{2(m-1)} \sum_{\iota \in N \setminus i} \left(1 - \xi_{i\iota}^{(Attribute)}\right) \tag{7}$$

The range of RA_i is $0 \leq RA_i \leq 1$, and we can interpret same as RB_i.

5.3 INDICES FOR PSYCHOLOGICAL FACTOR

Third, we consider the psychological aspect of utility change. When a customer cannot decide the attribute of purchase, the customer may look at various items. In this period, the customer's utility for each item is change largely. So, the rank of utilities of items may change, then the transition graph of utilities has some cross points. In this study, we focus on the cross points of utility transition, then we defined two transition indices of psychological factors

One is the *brand transition* index defined as the next equation.

$$CB_i = \frac{2 \sum_{i \in N \setminus i} \zeta_{i,\iota}^{(Brand)}}{\sum_{i \in A_{kc}} \sum_{\iota \in N \setminus i} \zeta_{i,\iota}^{(Brand)}} \tag{8}$$

where $\zeta_{i,\iota}$ is the number of cross points of utilities transition of brand i and ι, N is the number of items and A_{kc} is an evoked set. The range of CB_i is from 0 to 1, and when the value is higher then the brand is seemed indecision brands, and vice versa.

Another is the *attribute transition* defined as the next equation.

$$CA_j = \frac{2 \sum_{\nu \in J \setminus j} \zeta_{j,\nu}^{(Attribute)}}{\sum_{\omega \in X(A_{kc})} \sum_{\nu \in J \setminus j} \zeta_{\omega,\nu}^{(Attribute)}} \tag{9}$$

where $\zeta_{j,\nu}^{(Attribute)}$ is the number of the cross points of attribute j and ν, $X(A_{kc})$ is the attribute set for the evoked set A_{kc}. Also the range of this index is in 0 to 1, and when the value is high then the customer is under indecision about attributes.

6 CASE STUDY

In this section, we focus on some typical customers' result to explain various utility change structures.

Customer ID-7852 purchased a driver on May 1 and reset his brand examination. After that he resumed his examination from May 4 and purchased a putter on May 29. This customer increases the utility of the putter just before his purchase time. But he did not necessarily choose the putter a while ago. On the other hand, the degree (depth) of "indecision" becomes lower as approaching purchase time. In heterogeneity of attributes, "Green" is high. It is bisection situation which choose "Green" or not. In transition of attributes, "Green" is low and seldom has switching points with other attributes. Thus, it is thought that "Green" is not able to be excluded.

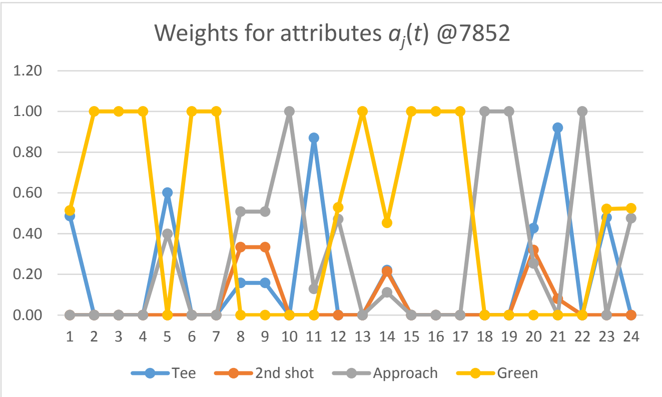

Figure 5: change of weights for attributes $a_j(t)$ (ID-7852)

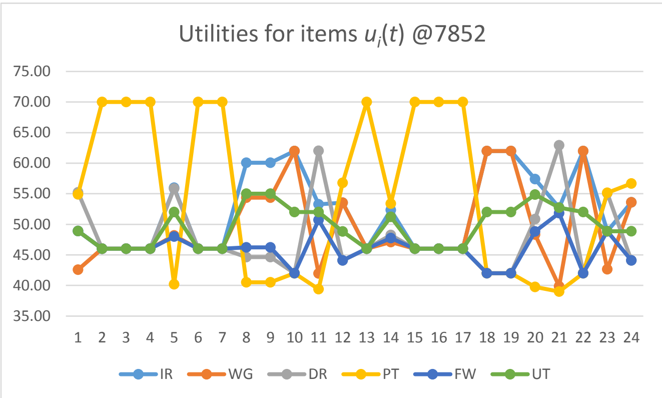

Figure 6: change of utilities for items $u_j(t)$ (ID-7852)

Table 3: changes of choice probabilities (ID-7852)

From purchasing time	IR	WG	DR	PT	FW	UT	D	Time
before 1 session	17.8%	17.8%	14.7%	18.8%	14.7%	16.2%	2.5782	0:00:00
before 3 sessions	18.2%	17.5%	15.6%	17.0%	15.0%	16.6%	2.5818	41:35:52
before 5 sessions	18.3%	16.4%	17.0%	15.5%	15.7%	17.1%	2.5827	65:25:57
before 15 sessions	17.8%	16.8%	16.0%	17.5%	15.3%	16.6%	2.5830	427:15:54
All sessions	17.5%	16.5%	15.9%	18.2%	15.4%	16.5%	2.5827	616:16:22

Table 4: changes of choice probabilities (ID-10940)

From purchasing time	IR	WG	DR	PT	FW	UT	D	Time
before 1 session	15.3%	15.3%	15.3%	23.3%	15.3%	15.3%	2.5639	0:00:00
before 3 sessions	17.0%	16.2%	14.7%	20.0%	15.5%	16.6%	2.5779	22:28:46
before 5 sessions	16.8%	15.4%	16.2%	19.3%	15.8%	16.5%	2.5810	70:29:18
before 15 sessions	16.8%	15.7%	16.3%	19.4%	15.6%	16.3%	2.5809	405:48:42
All sessions	18.1%	16.3%	16.9%	16.0%	15.7%	17.1%	2.5834	3099:51:01

Table 5: heterogeneity and transition of attributes (ID-7852)

Attribute	Tee	2nd shot	Approach	Green	Ave.
heterogeneity	0.592	0.530	0.643	0.769	0.633
transition	0.586	0.563	0.506	0.345	0.500

Table 6: heterogeneity and transition of items (ID-7852)

Category	IR	WG	DR	PT	FW	UT	Ave.
heterogeneity	0.479	0.578	0.508	0.763	0.532	0.439	0.550
transition	0.345	0.376	0.398	0.181	0.358	0.341	0.333

In heterogeneity of brands, “Putter” is exceptionally high. It is considered that the putter is a choice candidate.

In transition of attributes, “Putter” is low and seldom has switching points with other attributes. “Putter” is a stable choice candidate.

Next, we focus on another case. ID-10940 purchased putter at the 53rd session. The transition of $a_j(t)$and $u_i(t)$ are shown in Figure 7 and 8, respectively. These figures shows that this customer did not have much interest in putter at early times. However, from middle term, the customer turned the eyes to putter, and after some comparing, the customer purchased a putter. This customer did not clarify own criteria about importance of attributes. As shown Table 8, the index values of all attributes are similar, and the transition of attributes are changed severely. However, the putter and putter related attribute, i.e. “Green”, are faced in latter half. The value of putter heterogeneity in Table 9 is high, so the customer seems to be in indecision which category should be purchased.

Then, we show another example. Customer ID-10540 purchased “utility” at the last session. This customer is distinctive. The customer purchase “utility” however the value of utility is not high whole the term. As a whole, there is no high utility value shown as Table 8. Utility values of many categories are entwined as shown in Figure 10 and the rank of attribute values are changing (Figure 9). So, the customer might cannot decide purchase category or in the first place the customer might be not interested in purchase. However, the customer might feel to purchase at the last term.

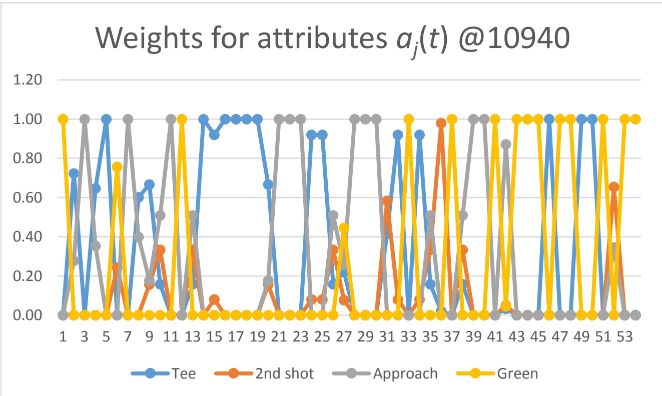

Figure 7: change of weights for attributes $a_j(t)$ (ID-10940)

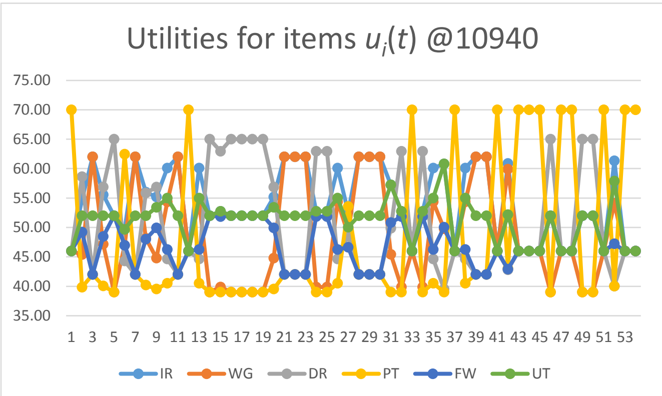

Figure 8: change of utilities for items $U_j(t)$ (ID-10940)

Table 7: changes of choice probabilities (ID-10540)

From purchasing time	IR	WG	DR	PT	FW	UT	D	Time
before 1 session	20.3%	16.6%	13.2%	13.0%	16.7%	20.3%	2.5625	0:00:00
before 3 sessions	19.2%	15.3%	16.4%	13.0%	16.9%	19.2%	2.5725	46:40:13
before 5 sessions	19.0%	15.6%	15.5%	14.3%	16.7%	19.0%	2.5769	79:48:34
before 15 sessions	18.7%	15.1%	17.5%	13.5%	16.8%	18.4%	2.5759	386:02:42
All sessions	18.0%	15.0%	17.7%	15.1%	16.5%	17.7%	2.5809	3101:31:16

Table 8: heterogeneity and transition of attributes (ID-10940)

Attribute	Tee	2nd shot	Approach	Green	Ave.
heterogeneity	0.680	0.575	0.664	0.697	0.654
transition	0.474	0.545	0.442	0.539	0.500

Table 9: heterogeneity and transition of items (ID-10940)

Category	IR	WG	DR	PT	FW	UT	Ave.
heterogeneity	0.496	0.593	0.552	0.740	0.547	0.459	0.564
transition	0.317	0.334	0.373	0.310	0.366	0.300	0.333

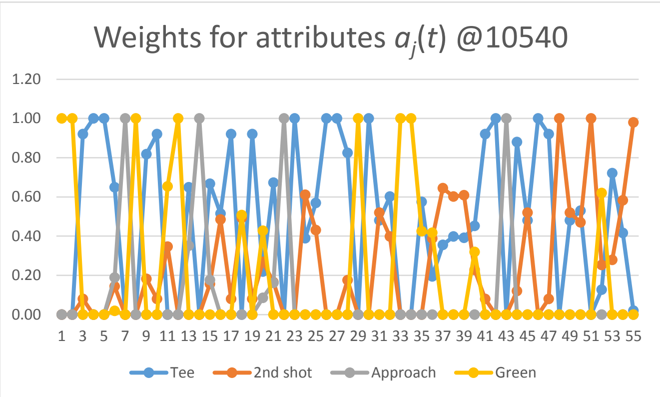

Figure 9: change of weights for attributes $a_j(t)$ (ID-10540)

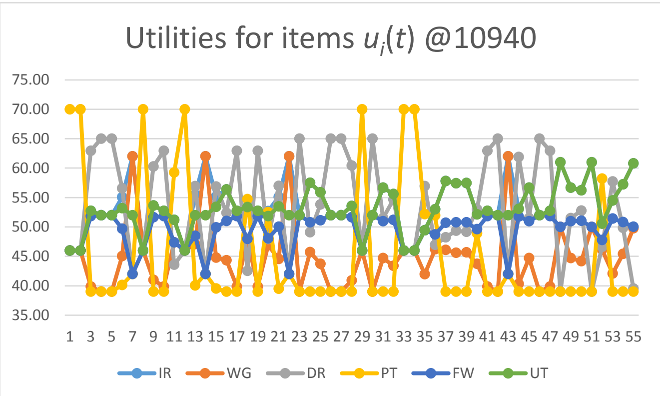

Figure 10: change of utilities for items (ID-10540)

Table 10: heterogeneity and transition of attributes (ID-10940)

Attribute	Tee	2nd shot	Approach	Green	Ave.
heterogeneity	0.705	0.640	0.629	0.675	0.662
transition	0.452	0.527	0.493	0.527	0.500

Table 11: heterogeneity and transition of items (ID-10540)

Brands	IR	WG	DR	PT	FW	UT	Ave.
heterogeneity	0.469	0.593	0.588	0.737	0.529	0.453	0.562
transition	0.353	0.298	0.385	0.272	0.333	0.359	0.333

Through these analyses, we know there are various types of indecision processes. It is difficult to categorize these results, so some more studies are needed to catch up customers' essentially psychological purchase factors.

In addition, we obtain that we can confirm a state of indecision from indices even if we do not extract it from figures.

7. CONCLUSION AND FUTURE WORKS

In this study, we focus on indecision situation on consumer behavior. To describe the situation, we showed a probabilistic choice model with dynamic utility change and some indices to explain the indecision phenomena. Then we demonstrated an example of real purchase and web access log data of EC site. Our result shows some typical process from indecision to decision, and capture the prediction of purchase.

Some future works are remained. First, in this study, the parameters of each specification are given, however, the expectation of item may be varied for each customer, so we need to give the value of parameter carefully. Second, we demonstrated only one example, so we need to inspect using the other cases, e.g., daily use category or food.

ACKOWLEDEGEMENTS

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research(C) 16K03944.

REFERENCES

Coombs, C.H.. and Avrunin, G.S. (1977) Single-peaked functions and the theory of preference, *Psychological Review*, **84**, 216-230.

Fishbein, M. (1963) An investigation of relationships between beliefs about an object and the attitude toward that object, *Human Relations*, **16**, 233-240.

Gensch, D.H. (1987) Empirical evidence supporting the use of multiple choice models in analyzing a population,' *Journal of Marketing Research*, **24**, 197-207.

Lussier, D.A. and Olshavsky, R.W. (1979) Task complexity and contingent processing in brand choice, *Journal of Consumer Research*, **6**, 154-165.

Montgomery, H. (1983) Decision rules and the search for a dominance structure, in Humphreys, P.C., Svenson, O. and Vári (eds.), *Analysing and Aiding Decision Process*, 343-369.

Tabata, T, Namatame, T. and Ohno, T. (2015a) Purchase support system in online-shop," *Journal of Japan Association for Management Systems*, **31(3)**, 229-236. (in Japanese)

Tabata, T, Namatame, T. and Ohno, T. (2015b) A scheme of dynamic analysis for compensatory discrete choice model, *Journal of Japan Association for Management Systems*, **31(3)**, 271-280. (in Japanese)