International Review of Management and Marketing

ISSN: 2146-4405

available at http: www.econjournals.com

International Review of Management and Marketing, 2020, 10(5), 92-102.

Extracting Innovative Buyers by Scoring Using Innovator Theory

Ryo Iwata[1]*, Kaoru Kuramoto[2], Kenyuu Matsumoto[2], Satoshi Kumagai[2]

[1]DO HOUSE Inc., Sales Promotion Division, Japan, [2]Aoyama Gakuin University, Department of Industrial and Systems Engineering, Japan. *Email: r.iwata@dohouse.co.jp

Received: 22 April 2020 **Accepted:** 25 July 2020 **DOI:** https://doi.org/10.32479/irmm.9815

ABSTRACT

For companies that want to sell a high volume of products, it is important to identify innovative buyers to help with product marketing efforts. The purpose of this paper is to construct a model extracting whether users are innovative buyers or not from their purchase histories at physical stores and from access logs from an online-to-offline (O2O) site. Innovative buyers are users who influence other users' product purchases, also known in innovator theory as innovators and early adopters. They purchase products quickly, visiting physical stores such as supermarkets and convenience stores. In other words, innovative buyers are known to have high cosmopolite natures. In extracting innovative buyers, we estimated the speed of user product purchases and their cosmopolite natures. This estimation index can also be referred to as innovator scores. We went on to verify this method with socioeconomic status points, personality points and communication points (SPC points), using consciousness data and profile data collected from a panel on an O2O site. Thus, we showed that innovative buyers could be extracted using this new method, and the accuracy was higher than that of traditional methods measuring only the speed from product sale start to user purchase.

Keywords: Innovator scores, Innovative buyers, SPC points, O2O, Cosmopolite natures, Extraction model
JEL Classifications: M31, O39, C38

1. RESEARCH BACKGROUND

In marketing consumer products with a short life cycle, such as foods, it is important for products to effectively penetrate into a market. Saito (1994) conducted a survey of the market penetration of Rao, released in the early 1990s, and found that the market penetration with innovators and early adopters steadily increased through the first 24 weeks. These demographics served as market leaders, so that the penetration of late markets, such as late majority and laggards, increased from the 17th week. Note that Kirin Ichiban Shibori, Kirin Fine Pilsner, and Suntory Jias were all released at the same time; Kirin Ichiban Shibori penetrated to innovators and early adopters and is still a standard product, but the others have withdrawn from the market. From this, we conclude that products that penetrate the market with innovators and early adopters are more likely to penetrate other market segments. In the future, it is expected that data regarding these demographics will be analyzed and used effectively by companies that want to leverage the data in their marketing.

In addition, CyberAgent and Digital InFact (2019) reported that the market size of stores utilizing digital advertising (O2O advertising) will reach 258.6 billion yen in 2024, approximately 6.4 times that of 2019. Because of this, it is important for companies to develop a method to capture innovators and early adopters from buyers (O2O-type users) who go between online and offline shopping, such as web sites and physical stores.

In innovator theory, Rogers (2003) found a way to specify innovators and early adopters. His innovator theory showed that users can be classified into five groups, according to their level of early adoption of innovation: innovators, early adopters, early majority, late majority, and laggards. There are differences in lifestyle and personality in each group.

The purpose of this study is to use the innovator theory advanced by Rogers (2003) to extract innovative buyers using purchase histories and access logs from an O2O site where online and

This Journal is licensed under a Creative Commons Attribution 4.0 International License

offline users gather. By utilizing this information about innovative buyers who are interested in a product, marketing activities for a new product can be more effectively developed and products can quickly penetrate into the market.

The objective of this paper is to propose a method of extracting innovative buyers, focusing on differences in behavior data and consciousness data for users who are active shoppers in physical stores and O2O sites. In this paper, Chapter 2 explains traditional innovative buyers extraction methods and the purpose of this research. Chapter 3 explains our proposed new extraction method. Chapters 4 and 5 explore the estimate of a proposed method using actual data, results, and considerations. Finally, Chapter 6 offers a conclusion.

2. TRADITIONAL METHODS AND PURPOSE OF THIS RESEARCH

Previous research demonstrated a method of analyzing buyers' consciousness data and using a questionnaire as a method of extracting innovators and early adopters. For example, Saito (1994) prepared 700 variables for the purpose of constructing a psychological scale, categorized user value criteria from questionnaires, and produced groups that corresponded to the classifications of innovator theory.

On the other hand, more recent extraction methods take into account the propagation rate and order of product adoption between users and items for the purpose of utilizing it for marketing (Bass, 1969; Ichikawa et al., 2012; Ishikawa et al., 2007; Kawamae et al., 2009; Kawamae et al., 2007; Mahajan and Muller, 1998; Mahajan et al., 1995; Mahajan et al., 1990; Menjo and Yoshikawa, 2008; Muller et al., 2009; Peres et al., 2010; Rusmevichientong et al., 2004; Song et al., 2007; Song et al., 2006). Many studies have attempted to extract users who adopt products earlier as innovative buyers. For example, Song et al. (2006) carried out an experiment on information network flow created from past purchase histories and discovered a method for predicting future purchases. Song et al. (2007) also found a method for identifying opinion leaders in innovator theory using the InfluenceRank algorithm. Rusmevichientong et al. (2004) extracted users' innovation level by focusing on the relative order of users who adopted the same products on Amazon.com. As a real-world example, Mahajan et al. (1990) conducted a study using data on eleven kinds of durable consumer products, and found a way to extract users who adopted quickly as innovative buyers. Mahajan and Muller (1998) also proposed a diffusion modeling framework for companies to identify conditions that target the majority of innovator theory.

These studies showed that there is no need for user-intensive input such as consciousness analysis using a questionnaire. Although there are studies that analyze the behavior data of users who make purchases only online or only at physical stores, there is no study examining the O2O market that is expected to grow rapidly in the future. The purpose of this research is to construct an extraction method that can be applied to O2O-type users without requiring user-intensive input.

3. MODEL FOR EXTRACTING INNOVATIVE BUYERS

3.1. Overview of the Model

Rogers (2003) showed that socioeconomic status variables, personality variables, and communication variables differ depending on users' innovation adoption behavior. This suggests that O2O-type users have different online and offline behavior properties, and may also have different socioeconomic status variables, personality variables, and communication variables. This study considers whether it is possible to extract whether users are innovative buyers or not by using differences in behavior data collected from purchase histories at physical stores and access logs from an O2O site. Therefore, to identify innovators, characteristic hypotheses were set for socioeconomic status variables, personality variables, and communication variables (Figure 1). We hypothesize that there are no differences in socioeconomic status variables between innovative and non-innovative buyers (Hypothesis 1). We also hypothesize that there are differences between personality variables and communication variables (Hypothesis 2 and Hypothesis 3).

From Hypothesis 3, it can be determined whether a certain user is an innovative buyer by comprehensively estimating the speed of the user's product purchase and their cosmopolite nature. By definition, cosmopolite nature is the degree to which a person is oriented toward an external social system. Generally speaking, cosmopolitan innovators travel extensively and are involved in matters that exceed regional system boundaries. For example, Ryan and Gross (1950) showed that hybrid corn innovators in Iowa traveled to the state capital of Des Moines more than the average farmer. Furthermore, Rogers (2003) showed that innovative physicians who adopted new drugs participated more frequently in specialist meetings in different regions than those who did not. The same is true of innovators among O2O users. If we assume that they go to physical stores such as supermarkets and convenience stores more than the average user, then we can add cosmopolite nature to the formula when extracting innovative buyers.

Thus, we developed a model for extracting and verifying the innovativeness of O2O users based on a method of extracting innovative buyers, by estimating the speed of user product purchases and their cosmopolite natures, and by applying the hypotheses of socioeconomic status, personality, and communication. An extraction model makes it possible to extract the innovation level of a user by using an innovator score that can be calculated from purchase histories (users who purchased a certain product at a physical store) and access logs from an O2O site.

Figure 2 summarizes a conceptual diagram of an innovative buyers' extraction model.

To test our hypotheses, we use SPC points calculated with socioeconomic status variables, personality variables, and communication variables as points. SPC points are used to verify whether there is a difference between innovative buyers and non-innovative buyers. To compare the extraction accuracy between a new method and a traditional method, SPC points when estimated by a traditional method is also calculated.

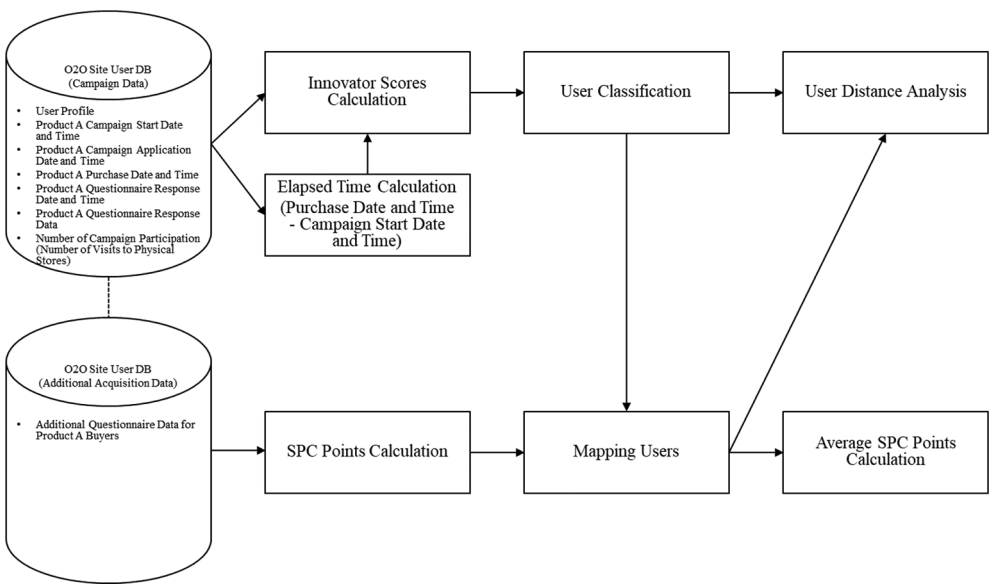
Figure 1: List of hypotheses regarding socioeconomic status, personality, and communication

Hypothesis 1: There is no correlation between socioeconomic status variables and innovation.
- Hypothesis 1-1: Innovators are older than Laggards.
- Hypothesis 1-2: When it comes to schooling, there is no crucial difference between Innovators and Laggards.
- Hypothesis 1-3: Innovators have higher literacy than Laggards.
- Hypothesis 1-4: When it comes to social status, there is no crucial difference between Innovators and Laggards.
- Hypothesis 1-5: When it comes to social mobility, there is no crucial difference between Innovators and Laggards.
- Hypothesis 1-6: In units of organization (large schools, large companies, etc.), there is no crucial difference between Innovators and Laggards.

Hypothesis 2: There is a positive correlation between personality variables and innovation.
- Hypothesis 2-1: Innovators are more empathetic than Laggards.
- Hypothesis 2-2: Innovators are less self-righteous than Laggards.
- Hypothesis 2-3: Innovators are more capable of dealing with abstract notions than Laggards.
- Hypothesis 2-4: Innovators are more rational than Laggards.
- Hypothesis 2-5: Innovators are more intelligent than Laggards.
- Hypothesis 2-6: Innovators are more willing to accept variation than Laggards.
- Hypothesis 2-7: When it comes to dealing with uncertainties and risks, there is no crucial difference between Innovators and Laggards.
- Hypothesis 2-8: Innovators have a more favorable attitude toward science than Laggards.
- Hypothesis 2-9: Innovators are not as fatalist as Laggards.
- Hypothesis 2-10: Innovators are more ambitious (in terms of schooling, status, occupation, etc.) than Laggards.

Hypothesis 3: There is a positive correlation between communication variables and innovation.
- Hypothesis 3-1: When it comes to social participation, there is no crucial difference between Innovators and Laggards.
- Hypothesis 3-2: Innovators are more highly connected via interpersonal networks in social systems than Laggards.
- Hypothesis 3-3: Innovators have a more cosmopolite nature than Laggards.
- Hypothesis 3-4: Innovators have more contact with change agents than Laggards.
- Hypothesis 3-5: When it comes to contact with mass media communication, there is no crucial difference between Innovators and Laggards.
- Hypothesis 3-6: Innovators have more contact with interpersonal communication than Laggards.
- Hypothesis 3-7: Innovators seek more information about innovation than Laggards.
- Hypothesis 3-8: Innovators have more information about innovation than Laggards.
- Hypothesis 3-9: Innovators have a higher degree of opinion leadership than Laggards.

Figure 2: A model for extracting innovative buyers



When extracting users' innovation level, we estimate the cosmopolite natures of users in addition to speed of user product purchases. Therefore, the time elapsed until the purchase of an object product is used to estimate the speed of product purchase, and the number of a user's participation in the campaign conducted on an O2O site (defined in this paper as "number of CP participation")[1] is used to estimate cosmopolite natures. These innovator scores are calculated for each user.

The date and time when the user purchases an object product is defined as PD, and an CP start date and time is defined as SD. In addition, the number of CP participation, which is the number of visits to physical stores by users, is defined as CP. The difference between PD and SD is defined as elapsed time, and users are clustered based on the magnitude relationship between elapsed time and CP.

For user set $J=\{1,2,\cdots, n\}$, if $PD_j–SD$ is arranged in ascending order and is $s(j)$, then Eq. (1) is obtained.

$$PD_{s(1)}–SD<_{PDs(2)}–SD<$$

$$<PDs(j–1)\text{-}SD<PDs(j)–SD$$

$$=PD_{s(j+1)}–SD<\cdots<PD_{s(n)})–SD \quad (1)$$

1 This number measures a user's campaign participation concerned with products other than the object product (in other words, the number of visits to physical stores).

The score *PSS* of the elapsed time of a user is defined as Eq. (2)[2].

$$PSS_{s(j)} = n-j+1,\ (1,2,\cdots n) \tag{2}$$

If $c(j)$ is the order in which CP_j is arranged in ascending order, then Eq. (3) is obtained.

$$CP_{c(1)} < CP_c(2) < \cdots < CP_c(j-1)$$
$$= CP_{c(j)} < CP_{c(j+1)} < \cdots < CP_{c(n)} \tag{3}$$

The score *CPS* of the user's number of CP participation is defined as Eq. (4)[3].

$$CPS_{c(j)} = j,\ (j=1,2,\cdots,n) \tag{4}$$

Innovator scores for user set $J = \{1,2,\cdots,n\}$ is defined as Eq. (5).

$$TS_j = PSS_{s(j)} + CPS_{c(j)},\ (j=1,2,\cdots,n) \tag{5}$$

Using the average μ_{TS} and the standard deviation σ_{TS} of innovator scores, we can classify user set J into five clusters in Eq. (6) through Eq. (10) (TS_I: Innovators, TS_{EA}: Early Adopters, TS_{EM}: Early Majority, TS_{LM}: Late Majority, and TS_L: Laggards).

$$TS_I = \{j \in J \mid TS_j \geq \mu_{TS} + 2\sigma_{TS}\} \tag{6}$$

$$TS_{EA} = \begin{Bmatrix} j \in J \mid \mu_{TS} + 2\sigma_{TS} \\ > TS_j \geq \mu_{TS} + \sigma_{TS} \end{Bmatrix} \tag{7}$$

$$TS_{EM} = \begin{Bmatrix} j \in J \mid \mu_{TS} + \sigma_{TS} \\ > TS_j \geq \mu_{TS} \end{Bmatrix} \tag{8}$$

$$TS_{LM} = \begin{Bmatrix} j \in J \mid \mu_{TS} > TS_j \\ \geq \mu_{TS} - \sigma_{TS} \end{Bmatrix} \tag{9}$$

$$TS_L = \left\{ j \in J \mid \mu_{TS} - \sigma_{TS} > TS_j \right\} \tag{10}$$

3.1. Hypotheses Verification with SPC Points

We hypothesize that a proposed method (hereinafter referred to as a "new method") will have higher personality and communication variables if a user is highly innovative. However, we hypothesize that higher innovation does not increase socioeconomic status variables, and that there is no difference between innovative and non-innovative buyers. The reason for this is that the object products are consumer products sold in supermarkets and convenience stores, and there is a high possibility that they can be purchased without regard to a user's socioeconomic status. Innovator theory provides that socioeconomic status and innovation are positively correlated with the need for high economic status to address uncertainties and risks of adopting expensive innovations. However, many consumer products, such as foods, are inexpensive (costing about several hundred yen). Furthermore, as Rogers (2003) points out, innovation cannot be measured solely by economic factors, and there are many innovators who do not have high economic status.

The details of SPC points calculation and hypotheses verification are given below.

1. A questionnaire on socioeconomic status, personality, and communication is conducted for users who have purchased an object product. For examples of questionnaire items, see below
   - Socioeconomic status: Final educational background (select from elementary school, junior high school, high school, etc.), occupation (select from company employee, company management/officer, public servant, etc.), and so on
   - Personality: Knowledge acquisition frequency (select from almost every day, 4 or 5 times a week, 2 or 3 times a week, etc.), future thinking (select from your own future is determined by your own will, your future is determined by your fate, etc.), and so on
   - Communication: New product cognition information (select what you know about the new product), word-of-mouth about new product (select from 1 people, 2 people, 3 people, etc.), and so on.
2. Based on data collected in step 1 and other user profile data, points x_s of socioeconomic status, points x_p of personality, and points x_C of communication are assigned to each user according to the response result. For example, let *SA*, *SB*, *SC*,⋯ be *p* estimation items for socioeconomic status, and let *SI*, *SJ*, *SK*,⋯ be the answer of user *j* for each item. Specific examples of *p* estimation items are single-response-type question items and multiple-response-type question items. In the following, for simplicity of formula representation, $p = 3$ without loss of generality. User *j*'s points x_{jS} of socioeconomic status, points x_{jP} of personality, and points x_{jC} of communication are defined as Eq. (11) through Eq. (13), respectively

$$x_{jS} = x_{jSI}^{SA} + x_{jSJ}^{SB} + x_{jSK}^{SC} \tag{11}$$

$$x_{jP} = x_{jPI}^{PA} + x_{jPJ}^{PB} + x_{jPK}^{PC} \tag{12}$$

$$x_{jC} = x_{jCI}^{CA} + x_{jCJ}^{CB} + x_{jCK}^{CC} \tag{13}$$

In this study, the minimum point of each estimation item was set to 0 and the maximum point was set to 4, and point allocation processing of options was performed according to the number of options and contents of options. Then, Table 1 was constructed to calculate SPC points for each user. For example, if *SA* is a single-response-type question item and corresponds to the estimation item of age, based on Hypothesis 1-1 in Figure 1 we would use 0 in their 20s, 1 in their 30s, 2 in their 40s, 3 in their 50s, and 4 in their 60s or older, and the point width was distributed uniformly. Additionally, for example, if *CB* is a multiple-response-type question item and corresponds to the estimation item of new product cognition information, the points of each option are assigned the same value based on Hypothesis 3-8, and the total points of all the options are set to 4. The "Other" option was set to 0, because points may vary greatly depending on the contents of a user's response.

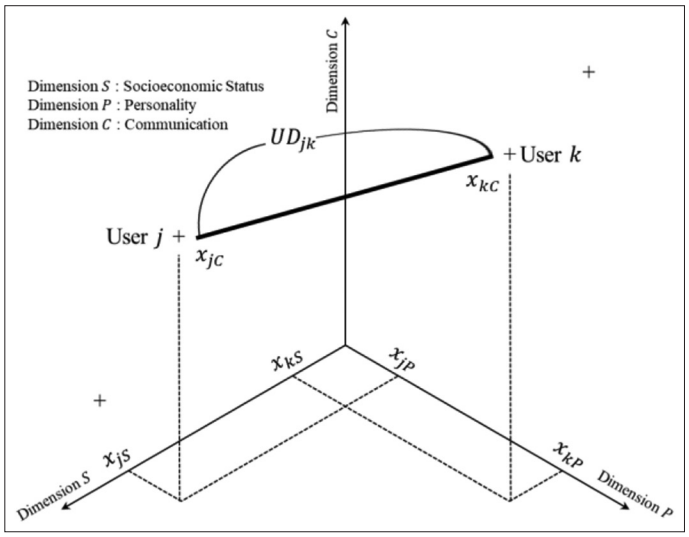
2 However, when $PD_{s(j)} - SD = PD_{s(j+1)} - SD$ as in Eq. (1), $PSS_{s(j)} = PSS_{s(j+1)} = n-j$.

3 However, when $CP_{c(j-1)} = CP_{c(j)}$ as in Eq. (3), $CPS_{c(j-1)} = CPS_{c(j)} = j-1$.

3. The features of users are mapped using socioeconomic status points, personality points, and communication points calculated in step 2. As Figure 3 shows, the similarity relationship is expressed by analyzing the distance between users (Okada et al., 2001).

The distance UD_{jk} between the point representing user j and the point representing user k is defined as Eq. (14)

Figure 3: Analysis of distance between users in a map



$$UD_{jk} = \sqrt{\sum_{g=1}^{r} \left(x_{jt} - x_{kt}\right)^2} \tag{14}$$

where x_{jt} is the coordinates of user j in dimension t, and r is the number of dimensions. Then, as Table 2 shows, the distances between n users are calculated.

Table 2 shows the number of combinations that can be made from users, that is, $n\ C_2 = n \times (n-1)/2$ distances. (The table shows only the lower triangular part of distance matrix of $n \times n$, excluding the main diagonal elements.) Because Table 2 shows $UD_{jk} = UD_{kj}$.

4. The distance between users calculated in step 3 is aggregated for each cluster, and the similarity relationship between clusters is expressed. The average distance $\overline{CD}_A$ between point sets representing users in cluster A is defined as Eq. (15)

$$\overline{CD}_A = \frac{\sum_{h=1}^{{}_{n_A}C_2} UD_{lm}}{{}_{n_A}C_2}$$

$$= \frac{\sum_{h=1}^{n_A \times (n_A - 1)/2} \sqrt{\sum_{g=1}^{r} \left(x_{lt} - x_{mt}\right)^2}}{n_A \times \left(n_A - 1\right)/2} \tag{15}$$

where n_A is the number of users in cluster A. The average distance $\overline{CD}_{AB}$ between point sets representing users in cluster A and point sets representing users in cluster B is defined as Eq. (16).

Table 1: SPC points calculation method

Estimation items	Answers and points				
SA (age)	20s 0	30s 1	40s 2	50s 3	60s or older 4
SB (final educational background)	Elementary School 0	Junior High School 2/3	…	Graduate School 4	Other 0
⋮	⋮	⋮	⋮	⋮	⋮
PA (knowledge acquisition frequency)	Almost every Day 4	4 or 5 times A week 18/5	…	Fewer than Once a year 2/5	Do not Acquire Knowledge 0
⋮	⋮	⋮	⋮	⋮	⋮
CA (new product cognition information)	Not using A 4/11	Not using B 4/11	…	Don't know About new Product 0	Other 0
CB (word-of-mouth about new product)	0 People 0	1 People 2/5	…	9 People 18/5	10 People or More 4

Table 2: Distance matrix between users in a map

	User 1	User 2	…	User j–1		…		User n–1	User n
User 1									
User 2	UD_{21}								
User 3	UD_{31}	UD_{32}							
⋮	⋮	⋮							
User j	UD_{j1}	UD_{j2}	…	$UD_{j,j-1}$					
⋮	⋮	⋮		⋮					
User n–1	$UD_{n-1,1}$	$UD_{n-1,2}$	…	$UD_{n-1,j-1}$	…		$UD_{n-1,n-2}$		
User n	$UD_{n,1}$	$UD_{n,2}$	…	$UD_{n,j-1}$		…		$UD_{n,n-1}$	

$$\overline{CD}_{AB} = \frac{\sum_{i=1}^{n_A \times n_B} UD_{oq}}{n_A \times n_B}$$

$$= \frac{\sum_{i=1}^{n_A \times n_B} \sqrt{\sum_{g=1}^{r} \left(x_{ot} - x_{qt}\right)^2}}{n_A \times n_B} \quad (16)$$

Then, as Table 3 shows, the average distances between five clusters are calculated.

Table 3 shows the number of combinations that can be made from 5 clusters (that is, 15 average distances). (The table shows only the lower triangular part of the distance matrix of 5 × 5, including the main diagonal elements.) Because Table 3 shows $\overline{CD}_{EAI} = \overline{CD}_{IEA}$.

5. Hypotheses are verified using the map diagram created in step 3. For example, comparing users classified as innovators with users classified as laggards, in accordance with hypotheses, the figures for personality and communication in the map are higher for innovators. In addition, average distances within a cluster and between clusters are compared using average distances calculated in step 4. For example, comparing users classified as innovators with users classified as laggards, in accordance with hypotheses, the average user distance within innovators is smaller than the average user distance between innovators and laggards.

3.2. Comparison with Traditional Method

We verify whether innovative buyers can be extracted by the speed of user product purchases used in a traditional method, and compare it with a new method.

1. Clustering is performed based on the magnitude relationship of elapsed time, representing the difference between a user's object product purchase date *PD* and a *CP* start date *SD*. For user set $J=\{1,2,\cdots,n\}$, if PD_j–*SD* is arranged in ascending order and is $s(j)$, then Eq. (17) is obtained.

$$PD_{s(1)} - SD < PD_{s(2)} - SD$$

$$< \cdots < PD_{s(n)} - SD \quad (17)$$

Using the average μ_{PD-SD}, the standard deviation σ_{PD-SD} of elapsed time data and coefficient $w(w \geq 0)$[4] for adjustment during innovator extraction, classify user set into five clusters in Eq. (18) through Eq. (22) (PD_I–*SD*: Innovators, PD_{EA-SD}: Early Adopters, PD_{EM-SD}: Early Majority, PD_{LM-SD}: Late Majority, and PD_{L-SD}: Laggards).

PD_I–*SD*

$$= \left\{ \begin{matrix} j \in J \mid PD_j - SD \\ \geq \mu_{PD-SD} - 2\omega\sigma_{PD-SD} \end{matrix} \right\} \quad (18)$$

PD_{EA}–*SD*

$$= \left\{ \begin{matrix} j \in J \mid \mu_{PD-SD} - 2\omega\sigma_{PD-SD} \\ > PD_j - SD \geq \mu_{PD-SD} - \omega\sigma_{PD-SD} \end{matrix} \right\} \quad (19)$$

PD_{EM}–*SD*

$$= \left\{ \begin{matrix} j \in J \mid \mu_{PD-SD} - \omega\sigma_{PD-SD} \\ > PD_j - SD \geq \mu_{PD-SD} \end{matrix} \right\} \quad (20)$$

PD_{LM}–*SD*

$$= \left\{ \begin{matrix} j \in J \mid \mu_{PD-SD} > PD_j - SD \\ \geq \mu_{PD-SD} + \omega\sigma_{PD-SD} \end{matrix} \right\} \quad (21)$$

PD_L–*SD*

$$= \left\{ \begin{matrix} j \in J \mid \mu_{PD-SD} + \omega\sigma_{PD-SD} \\ > PD_j - SD \end{matrix} \right\} \quad (22)$$

2. As in Table 1, points x_s of socioeconomic status, points of x_P personality, and points *xc* of communication are assigned to each user
3. Map users based on the calculated socioeconomic status points, personality points, and communication points. In addition, as Figure 4 shows, users of cluster *A* in a traditional method, users of cluster *A* in a new method, and users of cluster *A* that coincide with a traditional method and a new method are mapped

4 The reason for using coefficient ω is that the standard deviation is too large for the average of the elapsed time data, the value of $\mu_{PD-SD}-2\omega\sigma_{PD-SD}$ becomes minus, and there is a high possibility that innovators cannot be extracted successfully.

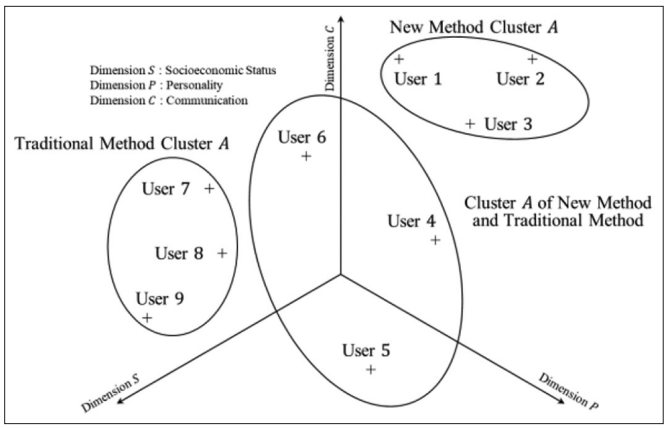
Table 3: Average distance matrix between clusters

Cluster items	Innovators (*I*)	Early adopters (*EA*)	Early majority (*EM*)	Late majority (*LM*)	Laggards (*L*)
Innovators (*I*)	$\overline{CD}_I$				
Early Adopters (*EA*)	$\overline{CD}_{EAI}$	$\overline{CD}_{EA}$			
Early Majority (*EM*)	$\overline{CD}_{EMI}$	$\overline{CD}_{EMEA}$	$\overline{CD}_{EM}$		
Late Majority (*LM*)	$\overline{CD}_{LMI}$	$\overline{CD}_{LMEA}$	$\overline{CD}_{LMEM}$	$\overline{CD}_{LM}$	
Laggards (*L*)	$\overline{CD}_{LI}$	$\overline{CD}_{LEA}$	$\overline{CD}_{LEM}$	$\overline{CD}_{LLM}$	$\overline{CD}_L$

Table 4: Calculation of average SPC points for each method

Cluster items	Socioeconomic status (*S*)		Personality (*P*)		Communication (*C*)	
	New method (*NM*)	Traditional method (*TM*)	New method (*NM*)	Traditional method (*TM*)	New method (*NM*)	Traditional method (*TM*)
Innovators (*I*)	$\bar{x}_{IS}^{NM}$	$\bar{x}_{IS}^{TM}$	$\bar{x}_{IP}^{NM}$	$\bar{x}_{IP}^{TM}$	$\bar{x}_{IC}^{NM}$	$\bar{x}_{IC}^{TM}$
Early adopters (*EA*)	$\bar{x}_{EAS}^{NM}$	$\bar{x}_{EAS}^{TM}$	$\bar{x}_{EAP}^{NM}$	$\bar{x}_{EAP}^{TM}$	$\bar{x}_{EAC}^{NM}$	$\bar{x}_{EAC}^{TM}$
Early majority (*EM*)	$\bar{x}_{EMS}^{NM}$	$\bar{x}_{EMS}^{TM}$	$\bar{x}_{EMP}^{NM}$	$\bar{x}_{EMP}^{TM}$	$\bar{x}_{EMC}^{NM}$	$\bar{x}_{EMC}^{TM}$
Late majority (*LM*)	$\bar{x}_{LMS}^{NM}$	$\bar{x}_{LMS}^{PM}$	$\bar{x}_{LMP}^{NM}$	$\bar{x}_{LMP}^{TM}$	$\bar{x}_{LMC}^{NM}$	$\bar{x}_{LMC}^{TM}$
Laggards (*L*)	$\bar{x}_{LS}^{NM}$	$\bar{x}_{LS}^{TM}$	$\bar{x}_{LP}^{NM}$	$\bar{x}_{LP}^{TM}$	$\bar{x}_{LC}^{NM}$	$\bar{x}_{LC}^{TM}$

Figure 4: Comparison of classification results for each method



4. As Table 4 shows, the average SPC points are calculated for each cluster, classified by a traditional method and a new method
5. Comparing a traditional method with a new method using the map diagrams created in steps 3 and 4 and the average SPC points shows a more effective innovative buyers extraction method.

4. ESTIMATION OF INNOVATIVE BUYERS EXTRACTION MODEL

4.1. Summary of Estimation

We verified whether innovative buyers in Figure 1 can be extracted using a new method. In addition, we verified whether innovative buyers can be similarly extracted by speed of user product purchases used in a traditional method, and compared it with the new method.

4.2. Estimation Data

We used behavior data and questionnaire data collected by DO HOUSE Inc. To be concrete, behavior data is recorded data such as product name, user ID, date and time of purchase, and number of CP participation of an object product purchased by users visiting an O2O site and physical stores. The data collection period was November 2018, and the number of users analyzed was 17,450. Furthermore, in March 2019 we extracted some users corresponding to innovators, early adopters, early majority, late majority, and laggards, and obtained questionnaire data for 2,161 users. The application object was beverage brands that included new products among consumer products and had a comparably large number of purchasers. Table 5 summarizes the data on innovator scores and SPC points.

4.3. Classification of Object Users by Innovator Scores

Innovator scores were calculated and innovative buyers were extracted based on the magnitude relationship. The distribution of innovator scores is shown in Figure 5.

Using a new method, the top users with high innovator scores could be extracted as innovative buyers. As Figure 5 shows, the distribution has a shape close to the normal distribution shown in innovator theory. In addition, Figure 6 shows that the cumulative number of purchasers also has a shape close to the S-shaped curve shown in innovator theory.

4.4. Hypotheses Verification (SPC Points)

Hypotheses are verified for each cluster for the object users.

1. Verification Method 1: Mapping object users With SPC points

Users were mapped by calculating SPC points. The results are shown in Figure 7 (only innovators and laggards are plotted). Maps are divided into three patterns: socioeconomic status and personality, socioeconomic status and communication, and personality and communication. The thick line is the cluster of innovators, and the thin line is the cluster of laggards.

Figure 7 shows that innovators have higher personality points and communication points than laggards. It is also clear that the difference in socioeconomic status between innovators and laggards is small.

2. Verification Method 2: Average distance analysis between object clusters using SPC points

Average distances between clusters were analyzed. The results are shown in Table 6. Average distance tables are divided into four patterns: socioeconomic status, personality, and communication; socioeconomic status and personality; socioeconomic status and communication; and personality and communication.

Table 6 shows that the average distances between clusters are closer for users within innovators, and the average distances tend to be greater for users between innovators and laggards.

Table 5: Data summary for innovator scores and SPC points

Data items	Innovator scores calculation data		SPC points		
	Elapsed time to purchase	Number of CP participation (number of visits to physical stores)	Socioeconomic status (*S*)	Personality (*P*)	Communication (*C*)
Number of samples	17,450 users	17,450 users	2,161 users	2,161 users	2,161 users
Estimation items	-	-	6 items	10 items	9 items
Average	79 hours 25 minutes 47 seconds	40 visits	12.312 points	26.270 points	8.784 points
Median	79 hours 26 minutes 30 seconds	17 visits	12.067 points	24.271 points	8.384 points
Minimum	0 hours 13 minutes 00 seconds	0 visits	2.333 points	8.091 points	0.983 points
Maximum	157 hours 50 minutes 00 seconds	478 visits	23.333 points	39.583 points	31.959 points
Standard deviation	45 hours 32 minutes 41 seconds	55 visits	4.346 points	4.567 points	3.433 points

Table 6: Average distances between clusters in socioeconomic status/personality/communication map

Average distances between clusters in socioeconomic status/personality/communication map					
Cluster items	**Innovators (*I*)**	**Early Adopters (*EA*)**	**Early Majority (*EM*)**	**Late Majority (*LM*)**	**Laggards (*L*)**
Innovators (*I*)	14.007				
Early Adopters (*EA*)	13.775	13.085			
Early Majority (*EM*)	14.477	13.401	13.313		
Late Majority (*LM*)	15.458	14.059	13.627	13.589	
Laggards (*L*)	16.086	14.382	13.708	13.411	12.913
Average distances between clusters in socioeconomic status/personality map					
Cluster items	**Innovators (*I*)**	**Early Adopters (*EA*)**	**Early Majority (*EM*)**	**Late Majority (*LM*)**	**Laggards (*L*)**
Innovators (*I*)	10.134				
Early Adopters (*EA*)	9.936	9.621			
Early Majority (*EM*)	10.488	9.945	9.927		
Late Majority (*LM*)	11.036	10.399	10.157	10.239	
Laggards (*L*)	11.073	10.348	9.980	10.025	9.716
Average distances between clusters in socioeconomic status/communication map					
Cluster items	**Innovators (*I*)**	**Early Adopters (*EA*)**	**Early Majority (*EM*)**	**Late Majority (*LM*)**	**Laggards (*L*)**
Innovators (*I*)	8.825				
Early Adopters (*EA*)	8.721	8.298			
Early Majority (*EM*)	8.982	8.396	8.443		
Late Majority (*LM*)	9.532	8.725	8.634	8.593	
Laggards (*L*)	9.918	8.880	8.688	8.470	8.069
Average distances between clusters in personality/communication map					
Cluster items	**Innovators (*I*)**	**Early Adopters (*EA*)**	**Early Majority (*EM*)**	**Late Majority (*LM*)**	**Laggards (*L*)**
Innovators (*I*)	9.055				
Early Adopters (*EA*)	8.892	8.251			
Early Majority (*EM*)	9.483	8.460	8.255		
Late Majority (*LM*)	10.348	8.993	8.464	8.346	
Laggards (*L*)	11.184	9.537	8.748	8.326	8.040

Figure 5: Innovator scores distribution

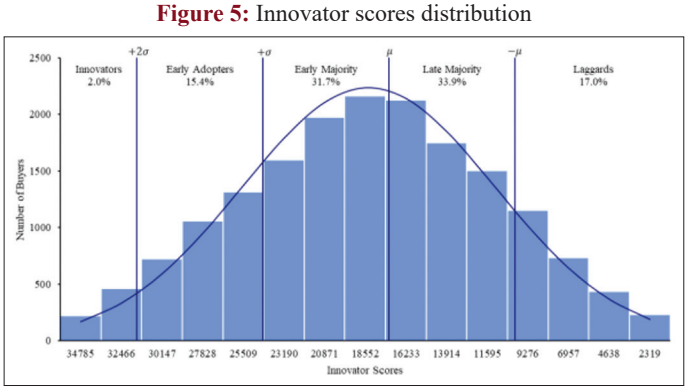
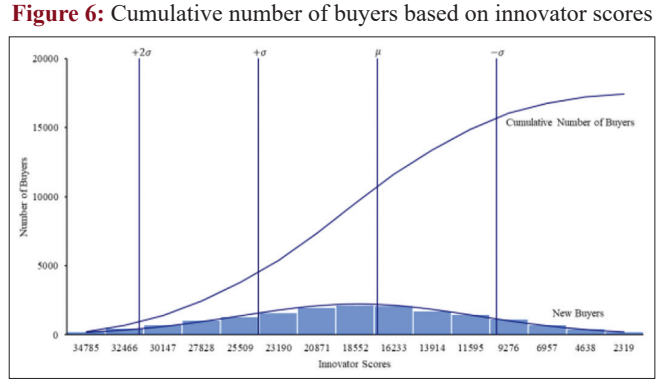


Figure 6: Cumulative number of buyers based on innovator scores



4.5. Comparison with Traditional Method

Using the map diagram derived in step 3 of Chapter 4 and the average SPC points calculated in step 4, a new method and a traditional method were compared and estimated. Figure 8 shows the distribution of innovative buyers extracted by a traditional method before these comparative estimations.

As Figure 8 shows, the distribution is not the normal distribution shown in innovator theory, but has a shape close to a uniform

Figure 7: Socioeconomic status/personality/communication maps

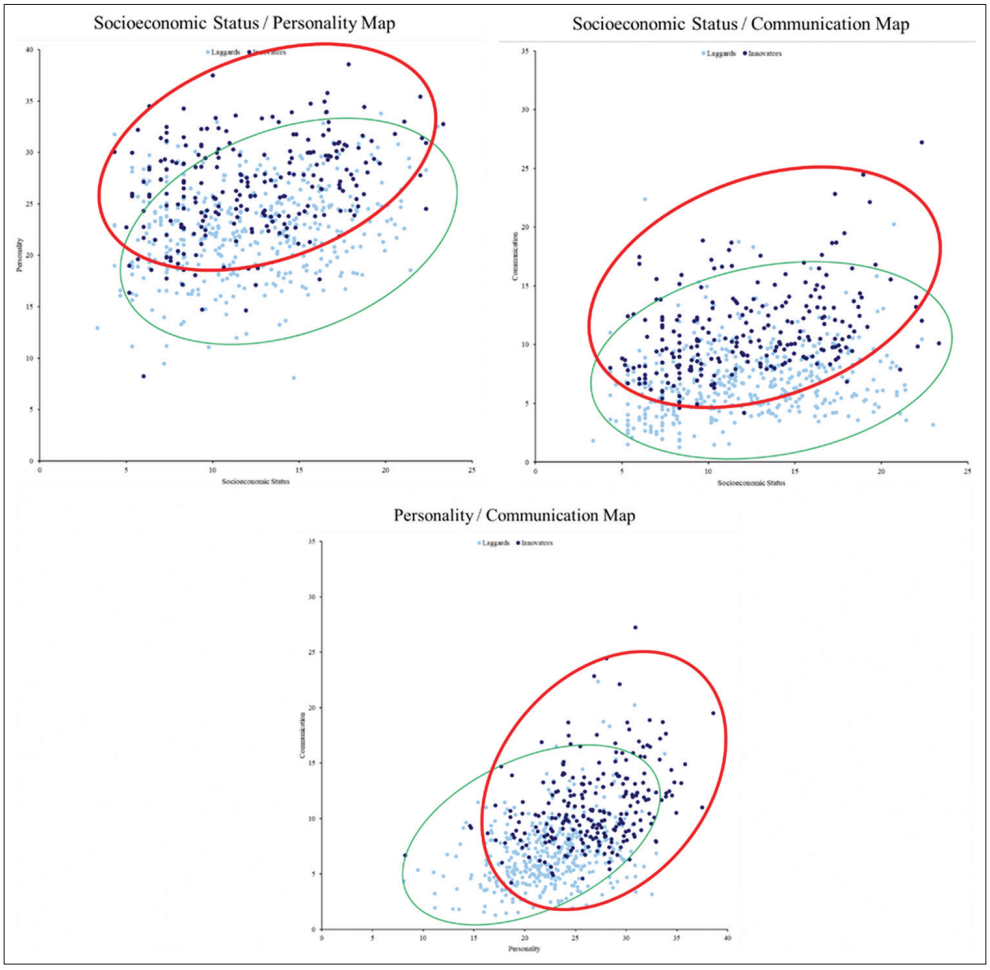


Figure 8: Elapsed time distribution

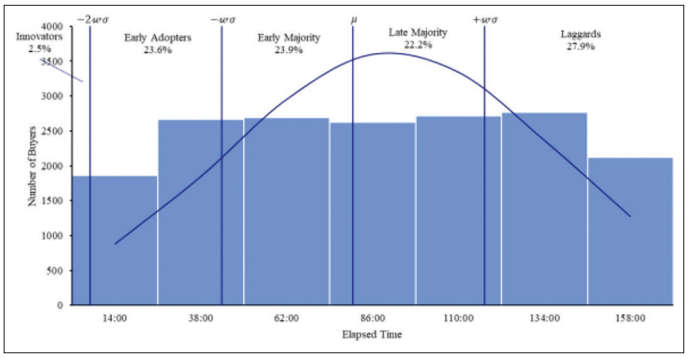
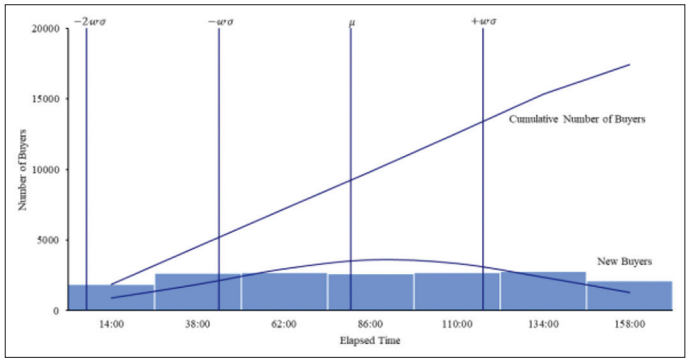


Figure 9: Cumulative number of buyers by elapsed time



distribution. In addition, Figure 9 shows that the cumulative number of buyers also has a shape close to a straight line instead of an S-shaped curve.

The comparison result between a new method and a traditional method is shown in Figure 10 and Table 7. The results of the map comparison when extracting innovators for each method are divided into three patterns: socioeconomic status and personality, socioeconomic status and communication, and personality and communication (Figure 10). The solid line is the cluster of innovators extracted by a new method, and the dotted line is the cluster of innovators extracted by a traditional method.

Figure 10 and Table 7 show that a new method of classifying users by innovator scores (on most map diagrams) and average SPC points satisfies the conditions of each cluster with higher accuracy than a traditional method of classifying users based only on the speed of product purchase. We can therefore conclude that this new innovative buyers extraction method is effective.

5. CONSIDERATION

A new method of classifying users by innovator scores is overall more accurate than a traditional method of classifying only by speed of user product purchases. In fact, looking at each cluster

Figure 10: Socioeconomic status/personality/communication maps of innovators for each method

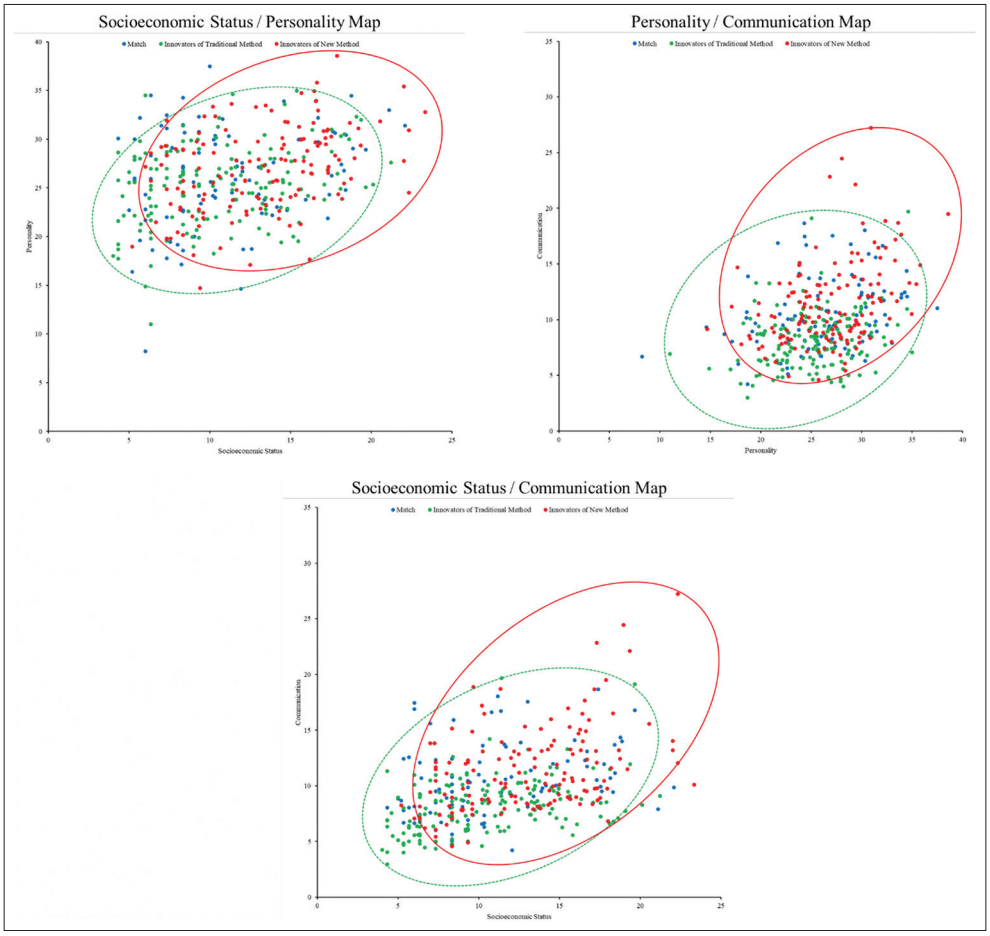


Table 7: Average SPC points – results for each method

Cluster items	Socioeconomic status (*S*)		Personality (*P*)		Communication (*C*)	
	New method (*NM*)	Traditional method (*TM*)	New method (*NM*)	Traditional method (*TM*)	New method (*NM*)	Traditional method (*TM*)
Innovators (*I*)	12.291	10.713	26.654	25.579	11.045	9.019
Early adopters (*EA*)	12.160	12.766	25.656	26.801	9.578	9.980
Early majority (*EM*)	12.277	12.160	24.055	24.923	8.944	9.007
Late majority (*LM*)	12.802	12.471	23.045	23.543	7.949	8.065
Laggards (*L*)	12.151	12.602	22.395	22.477	6.860	8.155

classified by a traditional method, there are many clusters that differ from hypotheses set in advance and are not suitable for extracting innovative buyers. A new method is more accurate than a traditional method in most clusters. However, at average distances between clusters, discussed in Chapter 4.4, users within innovators are farther apart than the distance between users who are innovators and early adopters. This suggests that the small number of samples of innovators themselves have a large effect on average distances, and to reduce average distances, it is necessary to consider an effective method.

Classification by this new method estimates users based on only two items: speed of product purchases and cosmopolite natures. Since it does not consider other user behaviors in terms of communication and personality, it may not be suitable as an index value indicating innovation. A classification method using data suitable for measuring innovation is required. For example, Rogers (2003) pointed out that there is a positive correlation between user communication variables and innovation. For this reason, the accuracy may be improved by using a user's word-of-mouth as an estimation item. We want to consider this for future tasks.

6. CONCLUSIONS

This paper proposes a method for extracting innovative buyers using innovator scores that can be calculated from users' purchase histories at physical stores and from access logs from an O2O site.

As a result of verifying the method using beverage brands as an application object, it was confirmed that a new method is more accurate than a traditional method using only the speed of product purchase. This new method is characterized by the ability to extract

innovative buyers in a short period of time, using behavior data without requiring user-intensive input. For this reason, it is easy to analyze the behavior and consciousness of large-scale innovative buyers online and offline using this new method.

In the future, it is necessary to improve the accuracy by searching for more effective user estimation items for extraction. Furthermore, we will continue to analyze the behavior and consciousness of innovative buyers and consider ways to use them more effectively in marketing measures.

REFERENCES

Bass, F.M. (1969), A new product growth for model consumer durables. Management Science, 15(5), 215-227.

CyberAgent and Digital InFact. (2019), Estimating and Forecasting the Market Size of Store-bought Digital Advertising. CyberAgent. Available from: https://www.cyberagent.co.jp/news/detail/id=23249.

Ichikawa, Y., Nakamura, M., Kishimoto, Y., Kobayashi, T. (2012), A Proposal of Extracting Innovative Users with Web Access Log of an E-commerce Site. Japan: IPSJ SIG Technical Report, Information Processing Society of Japan, 2012-GN-83. p1-7.

Ishikawa, M., Morita, T., Izumi, N., Yamaguchi, T. (2007), Content portal recommendation engine based on access log analysis. The 21st Annual Conference of the Japanese Society for Artificial Intelligence, 3G8-2, 1-3.

Kawamae, N., Sakano, H., Yamada, T. (2009), Personalized recommendation based on the personal innovator degree. In: Proceedings of the 3rd ACM Conference on Recommender Systems, RecSys 2009. New York, USA: ACM. p329-332.

Kawamae, N., Yamada, T., Ueda, N. (2007), Personalized ranking by identifying relative innovators. FIT, 2007, 99-102.

Kawamae, N., Yamada, T., Ueda, N. (2007), Personalized recommendation by identifying innovator. The Institute of Electronics Information and Communication Engineers, IEICE Technical Report, AI 2007-4 (2007-5), 107(78), 19-24.

Mahajan, V., Muller, E. (1998), When is it worthwhile targeting the majority instead of the innovators in a new product launch? Journal of Marketing Research, 35(4), 488-495.

Mahajan, V., Muller, E., Bass, F.M. (1995), Diffusion of new products: Empirical generalizations and managerial uses. Marketing Science, 14(3), G79-G88.

Mahajan, V., Muller, E., Srivastava, R.K. (1990), Determination of adopter categories by using innovation diffusion models. Journal of Marketing Research, 27(1), 37-50.

Menjo, T., Yoshikawa, M. (2008), Trend prediction in social bookmark service using time series of bookmarks. Data Engineering Workshop, 2008, B9-5.

Muller, E., Peres, R., Mahajan, V. (2009), Innovation Diffusion and New Product Growth. United States: Marketing Science Institute.

Okada, A., Kijima, M., Moriguchi, T. (2001), New Frontier of Management Science 6-mathematical Model of Marketing. Japan: Asakura Publishing Co., Ltd.

Peres, R., Muller, E., Mahajan, V. (2010), Innovation diffusion and new product growth models: A critical review and research directions. International Journal of Research in Marketing, 27(2), 91-106.

Rogers, E.M. (2003), Diffusion of Innovations. 5th ed. New York: Free Press.

Rusmevichientong, P., Zhu, S., Selinger, D. (2004), Identifying early buyers from purchase data. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004. New York, USA: ACM. p671-677.

Ryan, B., Gross, N. (1950), Acceptance and diffusion of hybrid corn seed in two Iowa communities. Iowa Agriculture and Home Economics Experiment Station, 29(372), 1.

Saito, T. (1994), Creating a Food Market-proposal of Hit Product Development Equipment Japan-VALS. Tokyo, Japan: Tokyu Agency Inc.

Song, X., Chi, Y., Hino, K., Tseng, B.L. (2007), Identifying opinion leaders in the blogosphere. In: Proceedings of the 16th ACM Conference on Information and Knowledge Management, CIKM 2007. New York, USA: ACM. p971-974.

Song, X., Tseng, B.L., Lin, C.Y., Sun, M.T. (2006), Personalized recommendation driven by information flow. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2006. New York, USA: ACM. p509-516.