Study on the Relationship Between Loyalty Program and Consumer Behavior on EC Site

Yusuke Nakasatomi[1(✉)], Takashi Namatame[2], and Kohei Otake[1]

[1] School of Information and Telecommunication Engineering, Tokai University,
2-3-23, Takanawa, Minato-ku, Tokyo 108-8619, Japan
6BJMl206@cc.u-tokai.ac.jp, otake@tsc.u-tokai.ac.jp
[2] Faculty of Science and Engineering, Chuo University, 1-13-27, Kasuga,
Bunkyo-ku, Tokyo 112-8551, Japan
nama@indsys.chuo-u.ac.jp

Abstract. Point programs are offered as part of the loyalty program. In this study, we analyze customer's purchase history and usage trend of points at golf portal site. We first classified loyal customers by RFM analysis. We also categorized into two unique types according to customer's point use behavior. Next, we performed logistic regression analysis for discrimination of loyal customers. Through these analyzes, we clarified the characteristics of point use tendencies for loyal customers.

Keywords: Point use behavior · RFM analysis · Logistic regression analysis

1 Introduction

Currently, many retailers provide point programs as part of the loyalty program. Many of the loyalty programs conducted by retailers are called FSP (Frequent Shopper Program) and provide benefits according to customers' purchase price [1]. The method of giving benefits in FSP can be classified using two types such as continuous or non-continuous and linear or nonlinear. In the linear and continuous loyalty program, registered members can acquire points depending on a specific amount of money, and they can use available point in 1 point unit. In a previous research [2], it is reported that many consumers tend to use according to the balance at the time of purchase based on mental accounting theory. On the other hand, there are very few studies that clarify the use tendency of customer's point program at EC site based on actual use of loyalty program.

In this study, we clarify the characteristics of point usage trends based on purchase history, customer attributes, and point usage history of customers who are using linear and continuous point programs set at a certain golf portal site.

2 Dataset

In this study, we use a dataset provided by the golf portal site. Specifically, we use ID-POS (Identification-Point of Sales) data, point balance data, point history data, customer attribute information data. All data are linked by customer ID.

The data period is one year from 2017/08/01 to 2018/07/31. The data summary is following.

- ID-POS data: purchase date, product code, product classification code, product unit price, sales volume, etc.
- Point balance data: balance of customers points.
- Point history data: processing date, use or acquisition flag, number of points, etc.
- Customer attribute information data: sex, age, place of residence, date of registration, etc.

In this study, we focused on the customers who registered at a certain golf portal site and we targeted customers with purchase history between 2017/08/01 and 2018/07/31. In addition, we excluded customers who have historically returned products, never gotten points or never used any of them.

As a result of extraction under the above conditions, the number of target customers was 78,193.

In the targeted EC site, five categories (golf accessories, golf gear, men's wear, lady's wear and other) were set as major categories. Number of purchases and the average of purchase price for each product category in dataset is shown Table 1.

Table 1. Number of purchases and average of purchase price

	Golf accessories	Golf gear	Men's wear	Lady's wear	Other
Number of purchases	506045	199454	385673	107434	4876
Average purchase price	4801.105	22115.080	5474.102	6173.782	3877.261

3 Analysis Aimed at Feature Extraction of Points Usage Tendency

In this section, we describe the analysis aimed at feature extraction of point use tendency. First, we evaluate the loyalty of customer, then we classify the customers into two types, loyal customers and general customers. Next, we calculate "point retention value" using point history data.

3.1 Divide Method of Loyal Customers

We performed RFM analysis [3] as a divide method of customers' loyalties. RFM analysis is a method of representing and discriminating customers' loyalty using three indicators, i.e., Recency, Frequency and Monetary. Recency is determined by how recently customer purchased products. Frequency is determined by how often customer purchase products. Monetary is determined by how much money customer used.

In this study, we evaluated RFM using the following data.

- Recency: Time after last purchased
- Frequency: Total number of purchases during the data period
- Monetary: Total usage amount during the data period

Customer ranks are given in 10 levels, higher value means higher loyalty for each of the three viewpoints, and the total value of viewpoints is set as a comprehensive index. Based on the obtained comprehensive indices, we decided the top 30%, 28,393 customers as loyal customers. The outline of the customer after divide is shown Table 2.

Table 2. Outline of the customers' RFM value after discrimination

	Loyal customers	General customers
No. of customers	28,939	49,800
Recency: average days	41.06	158.36
Frequently: average value	12.32	2.49
Monetary: average value	111790.03	15877.05

3.2 Usage Trend of Customer's Point

Next, we categorize the usage type of the point from the point history data. As an index for classification, in this study, we calculated the maximum value (number of days) from point acquisition to usage as "point retention value". An image of the calculation "point retention value" is shown in Fig. 1.
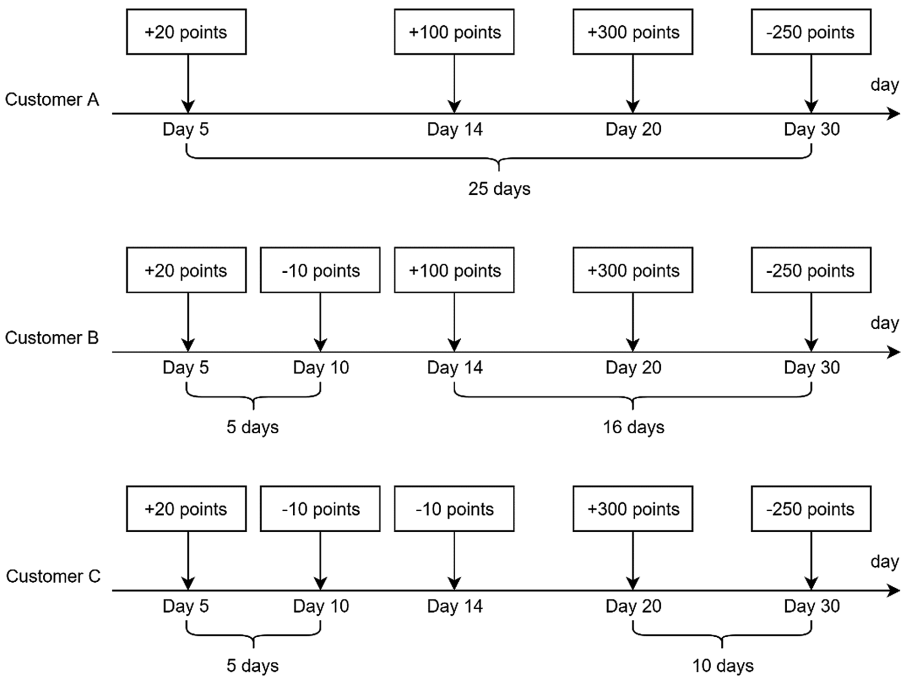

Fig. 1. An image of the calculation "point retention value" for each customer

From the Fig. 1, customer A's point retention value is 25 and customer A is a point saving-type which is holding without using point for a long period of time from point acquisition. On the other hand, customer C's point retention value is 10 and customer C is an immediate-type that uses points quickly after earned points.

We named the saving-type is within the top 10% of the points holding value, the immediate-type is the lowest point holding value is less than 10%.

3.3   Result of the Analysis of Point Usage Tendency

We use customers' loyalty and the usage type of points, we clarify the point usage tendencies.

First, the number of loyal customers belonging to the two types. As a result, it was found that about 59% of the saving-types are composed of loyal customers. On the other hand, in the immediate-type, it was found that about 88% of customers are loyal customers. Table 3 shows the comprehensive index calculated by RFM analysis for each usage type of point. From the Table 3, the saving-type has a larger variation in customer loyalty than the immediate-type.

Table 3. Percentage of customers and their avg. point balance by point usage type

	General customers	Loyal customers	Average point balance
Immediate-type	12%	88%	390.51
Saving-type	41%	59%	1351.94

Next, we show the product categories for each usage type are shown in Figs. 2 and 3. In the targeted EC site, five categories (golf accessories, golf gear, men's wear, lady's wear and other) were set as major categories. We totaled the number of purchasing based on major categories.
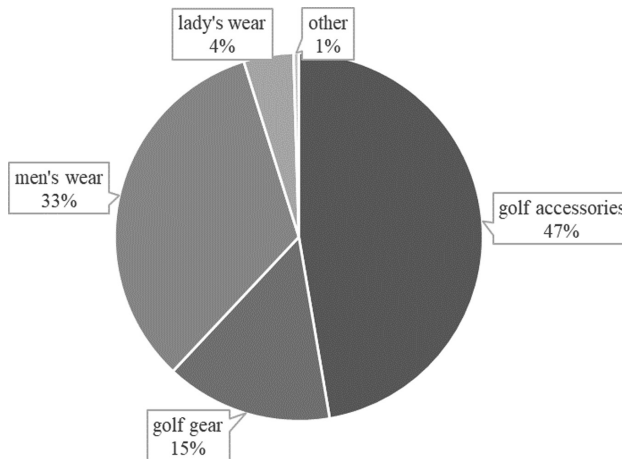

Fig. 2. Pie chart of purchases by category in the immediate-type
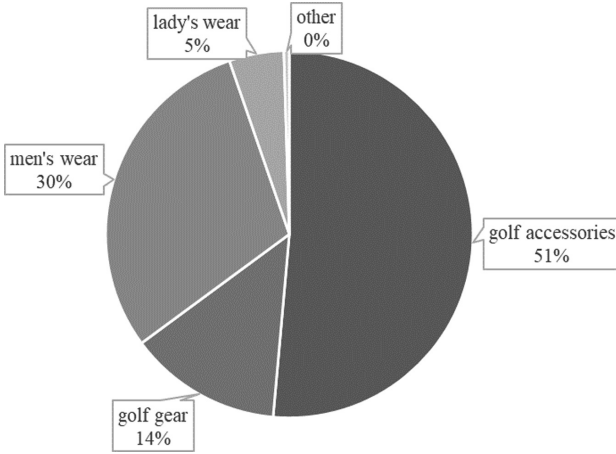
Fig. 3. Pie chart of purchases by category in the saving-type

From the Figs. 2 and 3, purchase of golf accessories is somewhat higher (about 3%) for the immediate-type. However, it cannot be said that there is a characteristic tendency. We create discriminant models of loyal customers and discover characteristic trends in the next step.

4   Analysis Using Logistic Regression Model

In this section, we describe analysis aimed at discriminating loyal customers using logistic regression. We created a model to distinguish the loyal customers calculated in 3.1 by binomial logistic regression and try to grasp the features.

The binomial logistic regression model is a type of classifier that performs class discrimination. By interpreting significant explanatory variables in the constructed model, it is possible to clarify the characteristics that affect the presence or absence of loyalties. In the binomial logistic regression analysis, the loyal customer probability p_i is expressed by the following equation.

$$p_i = \frac{\exp\{\sum_{j=0}^{m} \beta_j X_{ij}\}}{1 + \exp\{\sum_{j=0}^{m} \beta_j X_{ij}\}} \tag{1}$$

Here, X_{ij} is factors affecting loyalties and β_j is Parameters for each explanatory variable (β_0 is Intercept).

4.1   Variable Selection

In this analysis, we use the following variables.

- Explanatory variables: point retention value, point balance at 2018/07/31, Purchasing availability by product categories (golf accessories, club, men's wear, lady's wear, others)
- Objective variable: loyal customer classification

The objective variable is the classification result of the customer based on RFM analysis. 1 for loyal customers, 0 for general customers. Moreover, we set point retention value calculated in Sect. 3.2 as explanatory variables. In addition, we set explanatory variables for purchase history by five purchase categories, set 1 for purchase and 0 for no purchase. As shown in Sect. 3.3, when classifying the tendency of point use into the two types of the immediate-type and the saving-type, a difference was seen in purchasing behavior by product category. Therefore, we considered that customer loyalty would also be affected by product category.

4.2   Result

First, we performed under-sampling to estimate the model because two level of the objective variable is not balanced. The total number of cases is 78,193, the number of cases classified as loyal customers is 28,393, and the number of cases classified as general customers is 49,800. This is randomly extracted from majority data, here classified as a general customer, to match the number of minority data, here classified as a loyal customer.

Next, we performed normalization on the explanatory variables. Specifically, point retention value and point balance were normalized. We estimate the model 10 times and show the best results of the analysis in Table 4.

Table 4. Result of logistic regression analysis

Variables	Regression coefficient	P > |z|	Odds ratio
Point retention value	–0.0694	0.000	0.9329
Point balance	0.0840	0.000	1.0876
Sex	–0.0432	0.011	0.9576
Supplies/accessories	0.2331	0.000	1.2624
Club	0.9009	0.000	2.4618
Men's wear	0.1805	0.000	1.1977
Lady's wear	1.0557	0.000	2.8738
Others	0.5314	0.000	1.7012

Next, in order to verify the generalization performance of the model, we performed cross-validation. Cross-validation is a method that divides data used for model estimation and data used for model evaluation and applying it for validation and confirmation of

the validity of the analysis. In this study, we divided into 10 pieces of the dataset. The averages of the correct answer rates were obtained, and a value of 80% was obtained.

Next, we create confusion matrix using prediction results. Confusion matrix is to check the number of each of the samples judged correctly for y = 0 and y = 1 of the sample and the number judged erroneously in the crosstable. The results are shown Table 5.

Table 5. The confusion matrix of logistic regression model

		Predict class	
		Positive	Negative
Actual class	True	TP: 10,281	FN: 4,844
	False	FP: 730	TN: 7,603

We calculated prediction accuracy of the constructed model by using the following Eqs. (2) to (5).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{4}$$

$$\text{F-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

These indexes are better when the values are closed to 1.

Evaluation was conducted ten times, and the average value was obtained. The results are shown Table 6.

Table 6. The evaluation of the model

Accuracy	Precision	Recall	F-measure
0.7711	0.6268	0.8794	0.7319

From Table 6, we considered that the model has a certain discrimination accuracy. The ROC curve is shown Fig. 4.
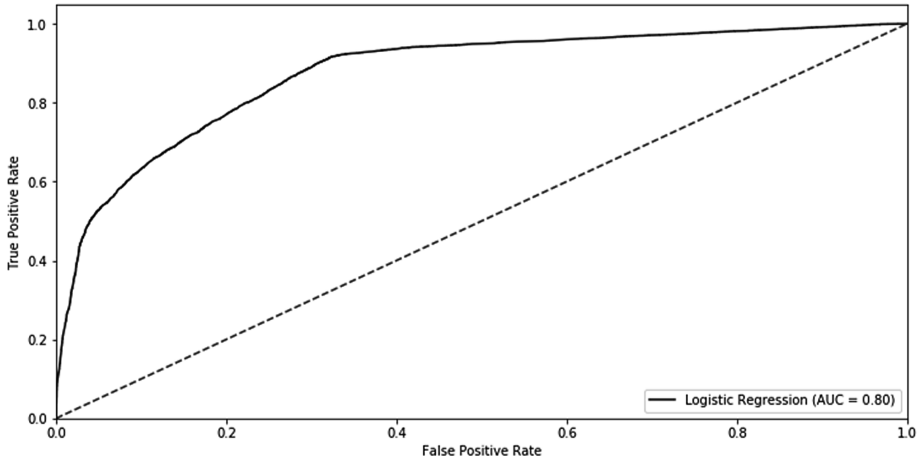
Fig. 4. A graph of the ROC curve

As shown in this figure, the ROC curve was exceeded to the diagonal line which is the expected value when it was predicated randomly. Hence also two figures, own result has enough predictability.

5 Discussion

In Sect. 3, we classified customer's point usage type into two types, the immediate-type and the saving-type, using point retention value. From Figs. 2 and 3 there was no big difference in the proportion of purchased genres between the saving-type and the immediate-type. In addition, the average purchase price by genre purchased from Table 1 greatly differs depending on genre. From these results, the use period of customer's point does not fluctuate depending on the amount of purchase. This seems to be the same result as the previous study [2] that the commodity price does not have a big influence on consumer's point use behavior.

In addition, since points are used immediately from Table 3, it is shown that the proportion of loyal customers increased to 88% in the immediate-type where point balance tends to be low. On the other hand, the saving-type, which saves points and tends to increase the point balance, shows that the percentage of loyal customers is lower than the immediate-type at 59%. The previous research [4] point out that customers who use points in linear and continuous point programs are more priced and purchasing behavior than customers who do not use them. In this study, we use RFM analysis that uses cumulative purchase price, purchase frequency, purchase period for discrimination of loyal customers, suggesting that customers using points as well as previous studies can be better customers.

Next, in the model shown in Sect. 4, as Table 4 shows, partial regression coefficients worked positively for all purchases by category. The influence of lady's wear and club category was higher loyalty because the average of purchase price per point is

relatively high in both genres. From Table 1, it is considered that the influence of Lady's wear, club was higher loyalty because the average of purchase price per item is relatively high in both genres.

In Table 4, the partial regression coefficient of point retention value was negative, and the partial regression coefficient of point balance worked positive. This is consistent with the fact that the proportion of loyal customers increased in the immediate-type where points are used for a short period of time and used immediately without saving points. However, the partial regression coefficient of Point retention value is – 0.0694, the partial regression coefficient of point balance is 0.0804, both partial regression coefficients can be said to be slight compared with partial regression coefficient in purchase by category. This suggests that the use period and balance of points do not significantly influence the classification between loyal customers and general customers classified by RFM analysis, and purchasing behavior greatly affects the classification.

6   Conclusion

In this study, we clarified the characteristic of point usage phenomenon in an EC site history data. It can be said that there is no big difference in the tendency of point use period in loyal and general customers classified by RFM analysis. However, there were differences between loyal customers and general customers depending on the point savings to consumption. This is due to the point data used in this study being linear and continuous point data at stores on the Internet. In the previous research [4], it is reported According to consumers in linear and continuous point programs continue to keep points until they reach a certain high point balance. In addition, another research [5], it is suggested that promoting the use of points in linear and continuous point programs of shops on the Internet does not lead to an improvement in customer loyalty. The conclusion of this study supports this.

The following three issues will be addressed in the future.

Firstly, it is a research on the point use tendency with respect to commodity price. In this research, we focused on the usage period of points, but the way of point is usage for each product unit price, the amount of points used for commodity price, was excluded from the research. It is possible to use only the fraction of the item price as a situation when using points. Also, in the case of shops on the Internet, many settlements are made through credit cards rather than cash. For this reason, we do not use fractional numbers, so we can think of patterns that always use the full amount. If the change in the utilization rate of points with respect to the unit price of a product is clarified, it will be possible to propose measures to promote the use of points from retailers and to improve sales.

Secondly, it is a research on point use trends in retailers in other forms. The subject of this research was a single item mail order and a retail business limited on the Internet. There are also multi-channel type retailers combining mall-type shops, Internet shops and real stores for retailers who only sell in real stores and retailers on the Internet. It is necessary to promote generalization of research by clarifying the use tendency of points in these customers.

Finally, it is a research on the tendency of goods bought at points. In this research, we do not consider that points affect the determination of purchased items themselves. For example, because the point's expiration date expires, it may be considered that the item is selected according to the point amount. Also, if the loyalty program is not provided, it can be inferred that the point cannot be acquired originally, and the customer neglects the point itself rather than cash. Whether items purchased at points are different from those purchased for cash is left as future research subjects.

References

1. Aoki, A., Sasaki, I.: A study of the effects of point program and discount system in Japanese retailers. Melco J. Manag. Account. Res. **4**(2), 3–16 (2015). (in Japanese)
2. Nakagawa, H.: Consumers' perceived benefit from price discounts and offering points: the moderating role of promotional benefit level. J. Behav. Econ. Financ. **8**, 16–29 (2015). (in Japanese)
3. Marcus, C.: A practical yet meaningful approach to customer segmentation. J. Consum. Mark. **15**(5), 494–504 (1998)
4. Smith, A., Sparks, L.: Reward redemption behaviour in retail loyalty schemes. Br. J. Manag. **20**(2), 204–218 (2009)
5. Nakagawa, H., Ono, J.: The effects of loyalty programs online: a comparative study on real shops. J. Acad. Soc. Direct Mark. **15**, 5–32 (2016). (in Japanese)