Integrated Decision Model for Direct Marketing: A Case Study on Catalog Mailing

Jen-Ying Shih†
Graduate Institute of Global Business and Strategy
National Taiwan Normal University, Taipei, Taiwan
Tel: (+886) 2-7734-3303, Email: jyshih@ntnu.edu.tw

Jin-Ru Huang
Graduate Institute of Global Business and Strategy
National Taiwan Normal University, Taipei, Taiwan
Tel: (+886) 2-7734-3303, Email: lenahuang032@gmail.com

Abstract. In recent years, with the advancement of information and communication technology and prosperous development of logistics, retailers heavily rely on virtual channels to provide better service for customers and conduct direct marketing campaigns for attracting more sales. Following the trend, consumers have changed their shopping behavior, which also increases the market scale of virtual channels. Virtual channels not only help retailers obtain a large number of customers to increase business growth, but also bring a lot of uncertainty and various expenditures. This study will probe into customer response and return behavior in direct marketing of a virtual retailer by data mining techniques. The data set covers customer demographic data, methods of payment, and transaction behavior data. Logistic regression analysis was used to choose relevant variables, and then two prediction models were constructed by backpropagation neural networks for classification tasks. Finally, the decision tree algorithm was applied to reveal several explicit business rules. The result shows that whether customers using discount is a significant factor in both models. Moreover, attributes, such as living area, frequency, recency, value, return ratio, preferences of payment and product category, also show importance when applying decision tree algorithms to explain the prediction models.

Keywords: direct marketing, data mining, customer response, return behavior.

1. INTRODUCTION

Non-store retailing via virtual channels has existed for a long time because of its convenience for consumers and efficiency for sellers' operation. With the advancement of information and communication technology and prosperous development of logistics, it has been getting popular in recent years. Catalog shopping has been one of virtual channels for retailers to provide product information in the fine printed paper format or digital format for their customers to attract their attention and thus place purchase orders via mail, fax, phone call, web or smart phone app. In the view points of the marketing function, catalog plays the following three roles: 1) demonstrating corporate (or brand) image represented in the elaborated documents, 2) providing sales promotion via discount coupons contained in catalogs, and 3) enhancing customers' affection toward the retailers by delivering catalogs to them (Yohn, 2015). Retailers often adopt differentiated strategies to operate their catalog business, which results in emphasizing different roles of catalog in terms of cost-benefit consideration.

In the first role, retailers focus on design of catalog to convey their corporate images, and the products they choose to display on catalog are also based on such logic, which usually makes the issue volume of the catalog mailed to customers is large and the number of pages is few. In this role, the catalog is a kind of advertising activities. As to the second role, retailers adopt precise marketing strategies to mail the catalogs with discount coupons to target customers who may place their orders after they receive the catalogs. In this role, the catalog is a kind of promotion activities and may be considered as one of the channels between retailers and customers. Retailers have to decide which set of customers they should choose to mail catalogs for maximizing net profits of catalog channel. The issue volume of the catalog is relatively small and the mailing targets are limited to those customers who have a high propensity to purchase. The reason is that the more customers they reach, the more costs (i.e., mailing costs and printing costs) they incur. For the third role, the catalog is regarded as a kind of customer services whose purpose is to maintain customer relationships by periodically mailing catalog. Hence, the mailing volume is usually large.

Based on the previous discussion, the second role of catalog needs to be deliberately discussed. Since the role of the catalog is regarded as a channel with sales promotion, direct marketing is required to effectively identify those potential customers. Therefore, a precise decision model for predicting response possibility of each customer toward a catalog is developed for the catalog marketing purpose.

Catalog shopping is one of the virtual channels of retailers, and the return costs should also be considered in the direct marketing decision process because virtual channels have higher return rates than physical channels. Return risk should be considered in the model. Therefore, the purpose of this research is to apply data mining techniques, including logistic regression (LR), neural networks (NN), and decision trees (DT), to propose an integrated decision model for direct marketing via mailing catalog with consideration of customers' return risk.

The organization of this research is structured as follows. Section 2 will give a brief introduction to the relevant literatures used for developing a direct marketing model in this research. In section 3, we will introduce the background of the case company in this study and the data set gathered from the case company. The model development process and research results will be described in section 4. Finally, section 5 will give some concluding remarks.

2. RELEVANT LITERATURE

In this section we will review the RFM theory and past relevant work on direct marketing to give a research background for developing a direct marketing model. In addition, the importance of customer return behavior on e-commerce will be discussed in this section.

Direct marketing has been considered as an efficient marketing approach in accessing potential customers with relatively low cost and an effective marketing strategy for capturing the target customers compared with mass marketing (Shih and Chen, 2014).

The RFM theory has been used by a firm to measure each customer's consumption behavior during an observation period, for example, 1 year. In RFM theory, R denotes the recency, which means the time period from the last purchase date of a customer till now; F indicates the consumption frequency of a customer during the observation period; M represents the total monetary value of consumption during the observation period. A firm usually classifies its customer into 125 categories (5*5*5) based on the three dimensions (i.e., R, F, and M) and dividing five groups along each dimension. This arrangement can be applied on grouping customers and target the groups with the highest response rate to conduct marketing campaigns. Since the theory was proposed by Bult and Wangsbeek (1995), it has invoked a lot of interests on applying the theory for predicting customer response in direct marketing (Baesen et al., 2002; McCartyz and Hastak, 2007; Olson and Chae, 2012; Shih and Chen, 2014); therefore, the RFM model derived from this theory has also been considered as a critical technique for database marketing (Blatterg, 2008).

The simplicity of the RFM model makes it popular in direct marketing; however, it only considers limited RFM factors and ignores the other determinants for response prediction, which makes its performance in predicting customer response cannot achieve satisfactory level. Several researchers (e.g., Shih and Chen 2014; Baesens et al. 2002) extend the model to covering the other relevant variables, for example, demographic variables, for predicting customer response. For developing a complete model to covering more factors, data mining techniques has been applied for this purpose and has been expected to achieve better performance.

Many retailers provide their customers with the rights of returning items within *N* days (*N* are greater than or equal to the legal requirement) if they are unsatisfied with the products. A good return service policy has become one of the necessary requirements in a highly competitive market. However, some consumers may abuse the rights of returning items and bring excess return costs to firms. Therefore, businesses should consider the costs of inertial return behavior of some consumers in their direct marketing decision to avoid excessive amount of returning costs. A strict return policy may lower logistics expense incurred from return behavior and decrease customer deception caused by returning items, but it may also decrease customer satisfaction and loyalty (Bahn and Boyd, 2004; Mollenkopf et al., 2007). It is critical to consider inertia return behavior of some customers in direct marketing decision to avoid accessing these customers who may bring negative contributions to firms because handing returned items would generate a lot of costs in the returning process, including labor costs, handing expenses, transportation costs, storage costs and product damage costs. Handing return items will spend 2 to 5 times of costs in selling items and may lead to no marginal profits (Johnson, 2003). Gordon (2006) proposed that firms should conduct de-marketing for those customers with low lifetime profitability and low strategic value, which can focus their efforts on high value customers.

3. DATA SET

3.1 The case company

This study gathered data from a virtual retailer (hereafter the case company) and developed an integrated decision model for the case company to choose the target customers to mail catalog with the hope to promote some selected products to increase sales. The case company is a TV home shopping retailer; moreover, the channels it has used to reach customers cover not only TV channels, but also catalog mailing, web service, Apps on mobile phones, MOD programs, and telemarketing. Figure 1 shows the shopping process of customers in this case company. The case company mails promotional documents to the potential customers (process 1). After customers receive advertising or promotional documents (i.e., catalog), they will make decisions to determine whether they will order the products (process 2). After they place the order to the case company and receive the products (process 3), there will be two outcomes. These customers may be happy to own the products, or they may be unsatisfied with the shopping experience and decide to return the items they purchased (process 4). The catalog mailing strategy is to access those customers with high propensity to respond to the catalog mailing (process 2) and low possibility to return items (process 4), which requires to develop a mechanism to choose the target customers.

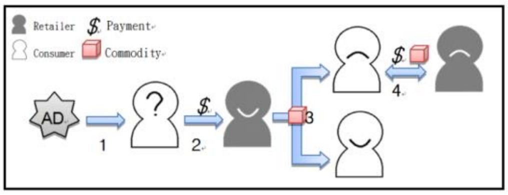

Figure 1: The shopping process

Data mining techniques have been identified as facilitators in the implementation of direct marketing (Shih and Chen, 2014). Therefore, this research attempts to assist the case company in developing a direct marketing model for identifying potential customers with a high tendency to place purchase orders and low possibility to return items when they receive the catalog. After the model scoring on these potential customers, the case company will mail the catalog to the customers with high scores and filter out the customers with high scores on returning behavior. We hope that this model can help capture potential customers of catalog channel and thus increase the case company's profits.

3.2 Sample and input variables

In the first stage of the integrated decision model, a response model is developed for predicting the possibility of each customer to respond to catalog mailing activities. We collected data from the case company, including who receives catalog, whether they respond to the catalog in one month (the target variable of the model), the transaction data of these customers in the previous 12 months, their demographic data, and their payment preference. Figure 2 shows the data window and the binary target variable in this study.

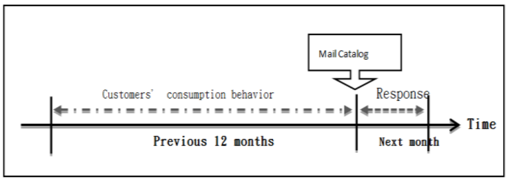

Figure 2: Data window and the binary target variable

The customers' consumption behavior in the previous 12 months was used for the coding of RFM variables (the concept can be referred to section 2 and their operational definitions are shown in Table 1.)

Table 1: RFM* Variable Coding Scheme

i	R_i	F_i	M_i
1	>= 10m**	<1	< $1,500
2	6 to 10m	1-3	$1,500-5,000
3	3 to 6m	3-8	$5,000-15,000
4	1 to 3m	8-13	$15,000-25,000
5	< 1m	>= 13	>=$25,000

* R: Recency; F: Frequency; M: Monetary Value
** m indicates the month.

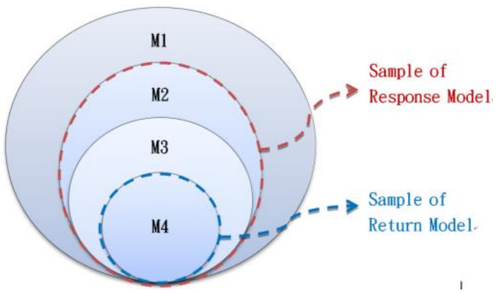

Figure 3: The sample selection

(Ps. M1 denotes all customers; M2 represents the customers who received the catalog; M3 indicates customers who respond to catalog mailing; M4 means the customers who placed the order and then returned items.)

The input variables are listed in Appendix Table A.1, which covers customer profile data, consumer behavior data, product category preference, and payment behavior. The sample selection of response model is shown in Figure 3.

In the second stage, a return prediction model is developed for forecasting the possibility of each customer's returning. The outcome of this model is used for filtering out the customers with high possibility of returning items. The input variables of this model are listed in Appendix Table A.2. The sample selection of response model is shown in Figure 3.

We adopted a 1:1 over-sampling approach, which was suggested by Ling and Li (1998) and Shih and Chen (2014) to avoid the problem that rare events would be ignored by the training model, to train the two models and design testing sets to evaluate their performances. Then, we split the sample into a 60% training set and a 40% testing set.

There are 493,165 customers (M2) who received catalogs from the case company, 20,521 customers (M3) respond to place order in catalog channel, among them 2,759 customers (M4) returned their purchases. Based on 1:1 over-sampling approach, there are 41,042 (20,521*2) records used for developing the customer response model, while there are 5,558 (2,759*2) records used for developing the return model.

4. MODEL DEVELOPMENT AND RESULTS

4.1 Model development process

In this study, Logistic regression analysis was used to choose relevant variables; then two prediction models were constructed by backpropagation neural networks for classification tasks, including identifying who will respond to the catalog mailing and who will return their purchase. Finally, the decision tree algorithm was applied to reveal several explicit business rules. The development process is shown in Figure 4.

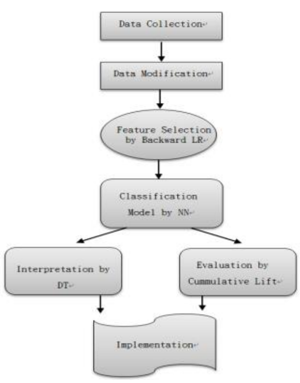

Figure 4: Model development process

4.2 Feature selection by backward logistic regression

This research adopts backward logistic regression approach to select key variables for explaining why a customer will response catalog mailing and why a customer will return the purchased item, and then the selected variables will be used as input data for developing neural network models to identify the target customers from the large scale of member data. The significant variables determined by backward regression are listed in Table 2 and Table 3, which covers living area, discount usage, etc. Totally, there are 22 significant variables found by the feature selection method for response model and 16 significant variables choosed by the feature selection method for return model.

Table 2: Variables selected by backward logistic regression for the response model

Variable	Chi-Square	Pr. > ChiSq
Consumption frequency of beauty products and material category	7.0265	0.0080
Consumption frequency of bedding category	8.0070	0.0047
Living Area	111.6542	<.0001
Whether using discount coupons	499.7769	<.0001
Consumption frequency of food category	48.0552	<.0001
Frequency (F)	33.9915	<.0001
Consumption frequency of home appliance category	6.0841	0.0136
Average return rate determined by return volume (%)	4.60841	0.0305
Average monetary value per order (log)	56.6124	<.0001
Value (log)	44.1940	<.0001
Membership seniority	38.5734	<.0001
Consumption frequency of mobile phone category	10.8017	0.0010
Monetary value	18.9061	0.0008
The average volume of commodities purchased per month in the past year	24.9011	<.0001
Preferred type of payment	24.5156	<.0001
Most preferred product category	94.4412	<.0001
Recency	45.1370	<.0001
Consumption frequency of body sculpting food	15.0358	0.0001

category		
Consumption frequency of beauty and SPA ticket category	12.2021	0.0005
Return volume in the past year	8.0516	0.0045
Consumption frequency of underwear category	10.9679	0.0009
Consumption frequency of apparel category	24.9443	<.0001

Table 3: Variables selected by backward logistic regression for the return model

Variable	Chi-Square	Pr. > ChiSq
Consumption frequency of bedding category	9.1927	0.0025
Living area	22.0880	0.0006
Whether using discount coupons	58.5953	<.0001
Consumption frequency of food category	10.7717	0.0010
Frequency (F)	25.6651	<.0001
Consumption frequency of jewelry category	9.3359	0.0022
Average monetary value per order (log)	25.5616	<.0001
The average volume of commodities purchased per month in the past year	24.1771	<.0001
Value (log)	11.7741	0.0006
Consumption frequency of boutique category	4.8759	0.0272
Preferred type of payment	34.1055	<.0001
Recency (R)	17.4350	0.0016
Average return rate determined by return amounts (%)	25.2944	<.0001
Consumption frequency of traveling product category	4.1171	0.0425
Consumption frequency of underwear category	8.2649	0.0040
Consumption frequency of women's apparel category	12.6340	0.0004
Consumption frequency of bedding category	9.1927	0.0025
Living area	22.0880	0.0006
Whether using discount coupons	58.5953	<.0001

4.3 Two classification models of neural networks

We applied backpropogation neural networks to train the two classification models with an architecture of three layers, including an input layer, a hidden layer and an output layer. In response model, the model architecture consists of an input layer with 22 units, a hidden layer with 20 units and an output layer with one unit. The number of input layer units is determined by backward logistic regression; the number of hidden layer units is determined by a series of trial experiments; the only one unit of the output layer represents the response of a customer. The model training process is shown in Figure 5. The average MSE of the training set is 0.205. The confusion matrix of the testing set is shown in Table 4, from which we can determine the accuracy rate of testing set is 0.297. Table 5 shows lift performance of the testing model. Under top 10% depth, the value of cumulative lift is 1.77.

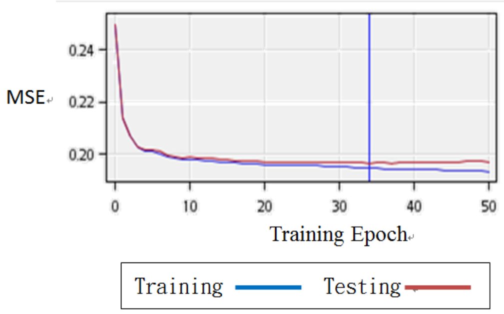

Figure 5: MSE of response model

Table4: Confusion matrix of the response model

		Actual Response		
		T	F	Total
Predicted Response	T	5,531	2,208	7,739
	F	2,675	5,998	8,673
Total		8,206	8,206	16,412

Table 5: Lift performance of response model

Depth	Lift	Cumulative Lift*
10	1.72229	1.77467
20	1.49574	1.67773
30	1.34470	1.59220
40	1.18537	1.50129
50	0.94878	1.40117
60	0.77467	1.30626
70	0.71133	1.22099
80	0.57491	1.14334
90	0.52439	1.07197
100	0.30000	1.00000

*the cumulative ratio of the percent of captured responses within each decile to the baseline percent response

In return model, the model architecture consists of an input layer with 16 units, a hidden layer with 15 units and an output layer with one unit to represent the return event. The average MSE of the training set is 0.192. The confusion matrix of the testing set is shown in Table 6, from which we can determine the accuracy rate of testing set is 0.256. Table 7 shows the lift performance of the testing model. Under top 10% depth, the value of cumulative lift is 1.83.

Table 6: Confusion matrix of the return model

		Actual Response		Total
		T	F	
Predicted Response	T	799	261	1,060
	F	306	844	1,150
Total		1,105	1,105	2,210

Table 7: Lift performance of the return model

Depth	Lift	Cumulative Lift*
10.00	1.80	1.83
20.00	1.71	1.79
30.00	1.56	1.70
40.00	1.29	1.60
50.00	1.00	1.49
60.00	0.76	1.37
70.00	0.53	1.26
80.00	0.40	1.15
90.00	0.35	1.07
100.00	0.33	1.00

*the cumulative ratio of the percent of captured responses within each decile to the baseline percent response

4.4 Decision tree for interpretation of the NN models

Although the neural networks can achieve better classification performance (Zahavi and Levin, 1997; Baesens et al., 2003; Bose and Chen, 2009), however, it cannot explain the mapping relationships between input variables and output variables explicitly, which looks like a black box and is unable to generate business rules for decision makers. Therefore, this study applied decision tree (Witten and Frank, 2005), C4.5, to interpret the neural network results. The important variables for explaining the response model and return model are listed in Table 8 and Table 9. Some explicit rules generated by the decision tree are shown in Appendix Table A.3 and A.4.

Table 8: Important variables determined by decision tree for the response model

Variable	# of nodes	Importance*
Whether using discount coupons	1	1
Frequency (F)	1	0.5746
Recency (R)	3	0.4182
Living area	3	0.2029
Most preferred product category	2	0.1704
Value (log)	2	0.1612
Average monetary value per order (log)	1	0.0848

* The relative importance of an input variable υ in subtree T is computed as $I(u;T) \propto \sqrt{\sum_{\tau \subset T} a(S_u,\tau)\Delta SSE(\tau)}$

where the sum is over nodes τ in T, and $s_υ$ denotes the primary or surrogate splitting rule using υ. $a(s_υ,τ)$ is the measure of agreement for the rule using υ in node τ. ΔSSE(τ) is the reduction in sum of square errors from the predicted values.

Table 9: Important variables determined by decision tree for the return model

Variable	# of nodes	Importance
Frequency (F)	1	1.0000
Whether using discount coupons	2	0.5546
Recency (R)	2	0.4031
Average return rate determined by the return amount (%)	3	0.3267
Value (log)	1	0.1592
Preferred type of payment	1	0.1172

* The relative importance of an input variable υ in subtree T is computed as $I(u;T) \propto \sqrt{\sum_{\tau \subset T} a(S_u,\tau)\Delta SSE(\tau)}$

where the sum is over nodes τ in T, and sυ denotes the primary or surrogate splitting rule using υ. a(sυ,τ) is the measure of agreement for the rule using υ in node τ. ΔSSE(τ) is the reduction in sum of square errors from the predicted values.

5. CONCLUSION

This research proposes an integrated decision model for determining the mailing lists of catalog marketing with the hope of maximizing the number of responses by several data mining techniques, including backward logistic regressions for feature selection, backpropogation neural networks for classification models, and C4.5 decision tree for explaining the classification results.

The result shows that whether customers using discount is a significant factor in both models. Moreover, attributes, such as living area, frequency, recency, value, return ratio, preferences of payment and product category, also show importance when applying decision tree algorithms to explain the prediction models. The importance order of the RFM concept is (1) F (2) R (3) M in this study. Future research work will focus on the comparison of the model with the RFM method and study the effect of return behavior on direct marketing decision.

REFERENCES

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003) Benchmarking state-of-the-art classification algorithms for credit scoring. *J. Oper. Res. Soc.* **54**, 627–635.

Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., and Dedene, G. (2002) Bayesian neural network learning for repeat purchase modelling in direct marketing. *Eur. J. Oper. Res.* **138**, 191–211.

Bahn, K. D. and Boyd, E. (2004). Information and its impact on consumers' reactions to restrictive return policies. *Journal of Retailing and Consumer Service*. 21(4), 415-423

Blattberg, R., Kim, B., and Neslin, S. (2008). *Database Marketing: Analyzing and managing customers*. New York: Springer.

Bose, I., and Chen, X. (2009) Quantitative models for direct marketing: A review from systems perspective. *Eur. J. Oper. Res.* **195**, 1–16.

Bult, J. R., and Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science*. **14(4)**, 378–394.

Gordon, I. (2006). Relationship demarketing: managing wasteful or worthless customer relationship. *Ivey Business Journal*, 70(4), 1-4

Johnson, B. (2003). Bringing it back: mastering the last frontier of corporate efficiency. *Materials Management and Distribution*, 48(7), 36

Ling, C.X., and Li, C. (1998) Data Mining for Direct Marketing: Problems and Solutions., in: *KDD*. pp. 73–79.

McCarty, J.A., and Hastak, M. (2007) Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. *J. Bus. Res*. **60**, 656–662.

Mollenkopf, D. A., Rabinovich, E., Laseter, T. M. and Boyer, K. K. (2007). Managing internet product retums: a focus on effective service operations. *Decision Sciences*, 38 (2), 215-50.

Olson, D.L., Chae, B.K., 2012. Direct marketing decision support through predictive customer response modeling. *Decis. Support Syst*. **54**, 443–451.

Rygielski, C., Wang, J.-C., and Yen, D.C. (2002) Data mining techniques for customer relationship management. *Technol. Soc*. **24**, 483–502.

Shin, J.-Y., and Chen, W.-H., 2014. A Data Mining Approach for Loan Marketing Response Model. *Proc. of The 15th Asia Pacific Industrial Engineering and Management Systems Conference*, Jeju, Korea. 1388-1393.

Witten, I.H., and Frank, E. (2005) *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Yohn, D.L., 2015. Why the Print Catalog Is Back in Style [WWW Document]. Harv. Bus. Rev. URL https://hbr.org/2015/02/why-the-print-catalog-is-back-in-style (accessed 7.20.15).

Zahavi, J., and Levin, N. (1997) Applying neural computing to target marketing. *J. Interact. Mark*. **11**, 5–22.

Appendix

Table A.1: Variables designed for developing the customer response model.

Variable Category	Symbol	Variable	Description
Target Variable	T	Response to catalog mailing	Designed as a binary variable, which was measured by whether customer place their order during the observation period (yes/no)
Customer Profile	CG	Gender	Customer gender (male/female)
	CA	Age	Customer age
	CL	Living area	Categorizing customer address into six areas

			in terms of living proximity
	CM	Membership seniority	Determined by time period from the customer's registration date to the mailing date
Consumption Behavior	BR	Recency (R)	Determined by the time period since the last purchase date to the mailing date
	BF	Frequency (F)	Consumption Frequency during one year
	BM	Monetary value (M)	Consumption amounts during one year
	BVal	Value	Value = M / R
	BVol	Volume of commodities purchased during one year	Volume of commodities purchased during one year
	BAV	The average volume of commodities purchased per month in the past year	BVol / 12
	BAM	Average monetary value per order	Bval / # of order in one year
Product Category Preference	PM	Most preferred product category	A total of 24 product categories defined in this concept, and the value of this variable is determined by the product category which has the highest consumption frequency
	PCi (i=1-24)	Consumption frequency of each product category	There are 24 product categories defined in this research, including fashion apparel and clothing accessory, jewelry, beauty and health, 3C, etc.
Payment Behavior	PW	Whether using discount coupons	Determined by whether the customer had used a discount coupon on their order (yes/no)
	PA	Amount of discount coupon used in one year	The accumulated amount of coupon used in one year
	PP	Preferred type of payment	Defined by the four types of payment (credit card, ATM, cash on delivery, and cash payment) which is determined by the largest cumulative amount of payment in one year
	PR	Redemption amount of reward points	Accumulated redemption amount of reward points used in payment in one year

Table A.2, A.3 and A.4 can be obtained upon request.