Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques

Yi Zuo*
Institutes of Innovation for Future Society
Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Japan
Email: zuo@coi.nagoya-u.ac.jp

Katsutoshi Yada
Faculty of Commerce
Kansai University
Yamate-cho, Suita-shi, Osaka, Japan
Email: yada@kansai-u.ac.jp

A B M Shawkat Ali
School of Science and Technology
The University of Fiji
Saweni, Queens Rd, Lautoka, Fiji
Email: shawkata@unifiji.ac.fj

***Abstract*—Over the past decades, prediction of costumers' purchase behavior has been significantly considered, and completely recognized as one of the most significant research topics in consumer behavior researches. While we attempt to measure response of purchase intention to the contextual factors such as customers' age, gender and income, product price and sale promotion, most of business models are basing on a linear equation to estimate weight of these factors due to the linear equation is not only intuitive for other academics to compare and replicate but also luminous to explain the results for business practitioners. Nevertheless, comparing with other research fields (e.g. pattern recognition and text classification), the prediction methods for purchase behavior are overconcentration of the linear models, especially linear discriminant analysis and logistic regression analysis. On the other hand, as more and more information and communication technologies (ICT, e.g. POS and sensor) are introduced into retail, marketing and management to collect business data, the volumes of data are increasing in exponential growth. Analysis based on linear models are insufficient to satisfy the requirement of academics and practitioners any more, and machine learning techniques have been increasingly attracted us to conduct them as an alternative approach for knowledge discovery and data mining. With regard to these issues, this paper employs two representative machine learning methods: Bayes classifier and support vector machine (SVM) and investigates the performance of them with the data in the real world.**

***Index Terms*—Purchase Behavior, RFID, In-store Behavior, Machine Learning, Multivariable Normality Test**

I. Introduction

Prediction of costumers' purchase behavior has been dramatically investigated over half a century, and completely recognized as one of the most significant research topics in consumer behavior researches. While we attempt to measure response of purchase intention to the contextual factors such as customers' age, gender and income, product price and sale promotion [1][2], most of business models are basing on a linear equation to estimate weight of these factors due to the linear equation is not only intuitive for other academics to compare and replicate but also luminous to explain the results for business practitioners. Nevertheless, comparing with other research fields (e.g. pattern recognition and text classification), the prediction methods for purchase behavior are overconcentration of the linear models, especially linear discriminant analysis [3] and logistic regression analysis [1]. Including the limitation of being difficult to get more predictive accuracy, the prediction results are also easy to converge towards the tradition conclusions, which would lead to a lack of revealing customers' characteristic and diversity. On the other hand, due to the linear models are only able to distribute following monotonic increasing or decreasing either, it is unavailable to exactly represent complex relation (exc. linear separability) between purchase behavior and explanatory variables. Furthermore, as more and more information and communication technologies (ICT, e.g. POS and sensor) are introduced into retail, marketing and management to collect business data [4][5][6][7], the volumes of data are increasing in exponential growth. Analysis based on linear models are insufficient to satisfy the requirement of academics and practitioners any more, and machine learning methods have been increasingly attracted us in the last two decades, to conduct them as an alternative approach for knowledge discovery and data mining.

With regard to these issues, this paper employs two representative machine learning methods: Bayes classifier [8] and support vector machine (SVM) [9], and investigates the performance of them with the data in the real world. The data are collected by a shopping path research in supermarket which is one of the hottest research topic in consumer behavior. In conformity to our previous study [10], SVM is employed to compare with Bayes classifier and other linear models in order to demonstrate the variations of purchase intention over stay time. In addition, a measurable and cumulative factor - stay time is introduced to association with the age, which is extracted from the in-store behavior data. Compared with the traditional forecast models, such as linear regression analysis and Bayes classifier, SVM provides a significant improvement in the forecasting accuracy of purchase behavior (from around 80% to approximately 90%). Second, multivariate analysis is applied to statistically process the massive amount of data on customers' stay time to make kernel selection easier for the classification task of the SVM. Basically, multivariate or multivariable normality testing yields the insight information of data. The omnibus normality test proposed by Doornik and Hansen [11] is adopted within the SVM theory to choose the automatic kernel selection for consumer purchasing behavior extraction.

978-1-5090-5753-5/16 $31.00 © 2016 IEEE
DOI 10.1109/APWC.on.CSE.2016.13

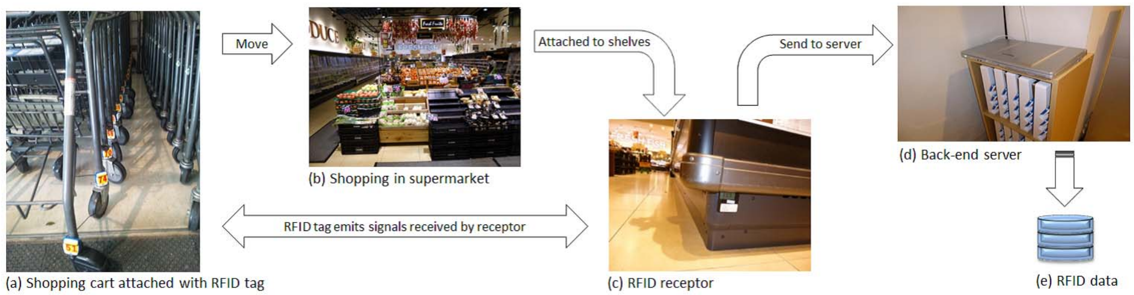


Fig. 1. Overview of the collection procedure for RFID data on customer shopping in a supermarket

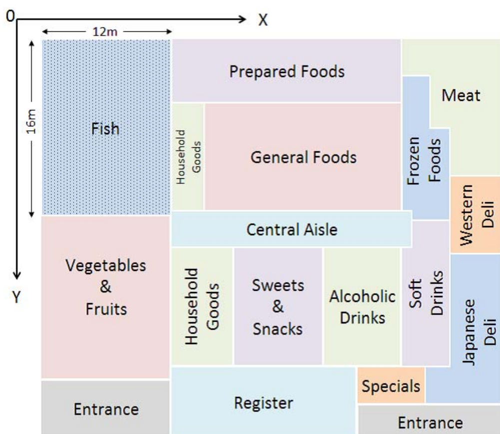


Fig. 2. Layout of the supermarket

The rest of the paper is organized as follows. In Section II, the framework of our RFID system and the preprocessing stage of RFID data are presented. The SVM and kernel trick are explained in Section III. Experimental results and multivariate omnibus normality tests are reported in Section IV. The results and conclusions are summarized at the end of this paper, in Section V.

II. Overview of System and RFID Data

A. Collection of RFID Data

We now demonstrate how this system can be used with actual customer in-store behavior data, using customer movement data gathered at a mid-sized supermarket in Japan. In this experiment, the shopping carts used by customers were equipped with RFID tags (Fig. 1(a)) which made it possible to precisely track in-store customer movements. The RFID system consists of 5 steps as shown in Fig. 1.

(a) RFID tags with unique IDs are individually attached to the shopping carts.
(b) As the customers walk through the aisles with this cart, the RFID tag emits a signal each second, and the information about the cart's position within the store can then be expressed as coordinates (x, y).
(c) These signals are received and sent to the back-end server via an RFID receptor located at the bottom (or the top) of the shelves.
(d) In the back-end server, a tracking system is employed to identify the signals and save them as raw data.
(e) By using an additional preprocessing system in the back-end server, the raw data are transformed into RFID data in XML form.

In addition to customers' movement data, floor layouts and purchasing histories were also gathered. The floor layout within the store was divided into 16 sections (Fig. 2). Some of those sections (e.g. Household Goods and General Foods) had subsections, and there were 28 subsections in total.

To record the customer's trip, the layout is reproduced into a picture with x and y coordinates on the scale of 15.7 pixels per meter. When the customer passes a certain area of the supermarket with a shopping cart equipped with an RFID tag, the information about her position can be received by the RFID receptor around the shelves and then transformed into a pixel point in our dataset using the matching floor layout. The RFID tag number attached to the shopping cart, the shopping date, time stamp, x and y coordinates of that time stamp, section of that coordinate and elapsed time are recorded. Table I shows the sample data obtained using our RFID system.

When the customer comes to the checkout register and makes a purchase, data are recorded and entered into our dataset. The dataset contains the shopping details as shown in Table II. Data include customer name, shopping date, type of item purchased, volume and unit price, totaling seven columns in the table.

We define this process from the time the customer enters the store until the purchase is completed as a basic unit of a customer's in-store behavior and assign a unique ID to identify it. In addition, by using this ID, the customer's purchase behavior obtained from the POS data are then linked to in-store behavior. After pre-processing the RFID and POS data, the data contained 5661 shopping units (sale transactions), with which 2847 customers could be tracked.

B. Measuring Range of the Fish Department

The experiment was carried out in a typical supermarket in Japan. In contrast to the previous studies, we focused on in-store customer behavior within a certain area instead of the

TABLE I
RFID DATA OF THE MOVEMENT

Customer Name	RFID Tag No.	Date	Time	X	Y	Selling Area	Elapsed Time
Anna	T001	2009/05/11	12:01:12	91	542	Entrance	1
			⋮				
Anna	T001	2009/05/11	12:03:51	79	87	Fish	1
Anna	T001	2009/05/11	12:03:52	85	88	Fish	1
Anna	T001	2009/05/11	12:03:53	86	89	Fish	2
Anna	T001	2009/05/11	12:03:55	87	87	Fish	1
Anna	T001	2009/05/11	12:03:56	95	88	Fish	1
Anna	T001	2009/05/11	12:03:57	99	88	Fish	1
Anna	T001	2009/05/11	12:03:58	98	88	Fish	10
Anna	T001	2009/05/11	12:04:08	92	88	Fish	1
Anna	T001	2009/05/11	12:04:09	91	89	Fish	1
			⋮				
Anna	T001	2009/05/11	12:12:05	319	511	Register	1

TABLE II
DETAIL OF THE POS DATA

Customer Name	Date	Time	Item Name	Item Category	Volume	Amount
Anna	2009/05/11	12:12:30	Cabbage	Vegetable	1	150
Anna	2009/05/11	12:12:30	Banana	Fruit	1	198
Anna	2009/05/11	12:12:30	Sashimi	Fish	2	596
Anna	2009/05/11	12:12:30	Pork	Meat	1	232

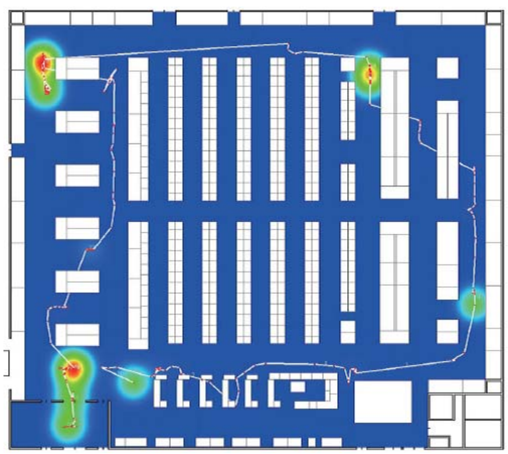


Fig. 3. Visualization of stay time from shopping trip [12]

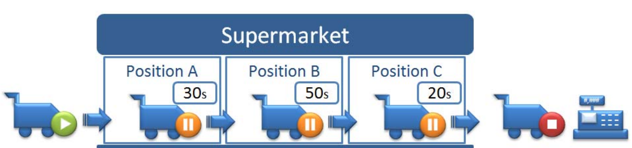


Fig. 4. Sample calculation of stay time

whole supermarket. Since fish is featured much more prominently on the Japanese plate, we selected the fish department as the experimental object. The measuring range, with a length of 16 meters and a width equals 12 meters, is represented as the shaded pattern in Fig. 2.

C. Definition of stay time

In this section, we explain the definition of stay time, i.e., the amount of time customers spent in the fish department. For a given customer, the RFID tag tracks the shopping trip from the time of entering the store to making a purchase at the checkout register. Fig. 3 shows a density estimation of stay time using one customer's shopping trip [12]. The figure was drawn using line segments to connect coordinate points of the customer's trip, which was tracked per second, and is combined with the stay time distribution expressed by the density estimation method. From this figure, we can see that using only the customer's shopping trip can make it difficult to know how a customer spend time in a certain area.

In this section, we explain the definition of stay time. For instance, Anna stayed 30 s in position A (x_A, y_A), next moved to position B (x_B, y_B) and stayed 50 s, and then moved to position C (x_C, y_C) and stayed 20 s (see Fig. 4). There were 100 s of total stay time in Anna's shopping. Thus, if Anna remained in position (x_i, y_i) for t_i seconds, then the time spent by Anna in the supermarket is expressed as follows:

$$T = \sum_{i=0}^{n} t_i \tag{1}$$

where the notation t_i denotes the "Elapsed Time" shown in Table I. Furthermore, making an addition to Eq. (1), only if she comes into the fish selling area is the position accepted. Therefore, the stay time T_{Fish} that the customer spent in the fish department is defined as follows:

$$T_{Fish} = \sum_{i=0}^{n} t_i, \tag{2}$$

$$t_i = \begin{cases} \text{Elapsed Time}, & \text{if position in fish department.} \\ 0, & \text{otherwise.} \end{cases}$$

By using Eq. (2), each individual customer's stay time spent in the fish department is calculated.

III. Methodology

A. Support Vector Machine

The support vector machine (SVM) belongs in the supervised learning theory group, which is comparatively very effective for classification, regression and clustering tasks. Compared to other learning algorithms, an SVM can effectively handle high-dimensional data space due to its unique kernel component. Different kernel functions can easily generate a set of decision functions, even when the number of dimensions is greater than the total number of samples. In the data modeling phase, the SVM identifies the small number of data points, called support vectors (SVs), that are closest to the hyperplane. Therefore, the SVM acts in the learning space as a memory-efficient learning algorithm. In this section, we briefly summarize SVM theory. Let us consider l, an independent and identically distributed sample: $(x_1, y_1), \cdots, (x_l, y_l)$, where x_i for $i = 1, \cdots, l$ and $y_i = \{+1, -1\}$ is the class label for data point x_i. Our aim is to find a decision function f with the property $f(x_i) = y_i, \forall i$.

$$y_i[(\mathbf{w} \cdot \mathbf{x_i}) + b] \geq 1, \quad \forall i. \tag{3}$$

Basically, a separating hyperplane often does not exist. To allow for the possibility of examples violating Eq. (3), Vapnik introduced the slack variables ξ_i

$$\xi_i \geq 0, \quad \forall i \tag{4}$$

to get

$$y_i[(\mathbf{w} \cdot \mathbf{x_i}) + b] \geq 1 - \xi_i, \quad \forall i. \tag{5}$$

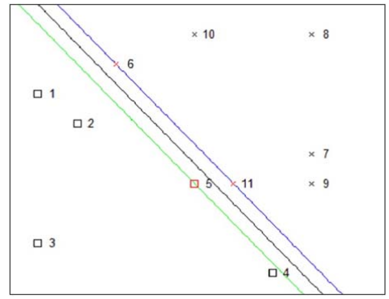
Therefore, the optimization problem becomes:

$$\tau(\mathbf{w}, \xi) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + \gamma \sum_{i=1}^{l} \xi_i. \tag{6}$$

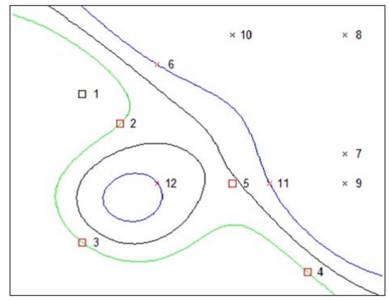
The constraints are described in Eqs. (4) and (6). Vapnik then introduced the Lagrange multiplier α_i and used the Kuhn-Tucker theorem of optimization theory. Therefore, we can calculate the weight vector following Eq. (7) with nonzero coefficients α_i only where the corresponding example $(\mathbf{x}_i, y_i)$ precisely meets the constraint Eq. (5).

$$\mathbf{w} = \sum_{i=1}^{l} y_i \alpha_i \mathbf{x}. \tag{7}$$

These nonzero coefficients are called *support vectors*(SVs); SVM then ignores the remaining data vectors in the testing phase of the unseen instances. As a result, the SVM can provide a faster solution compared to other learning algorithms.



(a) Linear classification



(b) Nonlinear classification

Fig. 5. Typical classifications generated by applying different kernel tricks

The constraint Eq. (5) is satisfied automatically (with $\xi_i = 0$), and it does not appear in expansion Eq. (7). The coefficients α_i are found by solving the following quadratic programming problem. Maximize

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \tag{8}$$

subject to

$$0 \leq \alpha_i \leq \gamma, \quad i = 1, \cdots, l \quad and \quad \sum_{i=1}^{l} \alpha_i y_i = 0. \tag{9}$$

Therefore the decision function can be written as

$$f(\mathbf{x}) = \mathbf{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i \cdot (\mathbf{x}_i \cdot \mathbf{x}_j) + b\right]. \tag{10}$$

To obtain better general decision surfaces, one can first nonlinearly transform a set of input vectors $x_1, \cdots, x_l$ into a high-dimensional feature space. Therefore the final decision function becomes:

$$f(\mathbf{x}) = \mathbf{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i \cdot K(\mathbf{x}_i \cdot \mathbf{x}_j) + b\right] \tag{11}$$

where $K(\mathbf{x}_i \cdot \mathbf{x}_j)$, called the "kernel,"is the most important component of SVM theory.

B. Linear/Nonlinear Classification

In the field of classification, the purpose of statistical learning algorithms is to construct a hyperplane from the observed

TABLE III
COMPARISON OF FORECAST METHODS

Forecast Method	Accuracy	Accuracy($P = 1$)
Linear Discriminant Analysis	81.09%	81.29%
Logistic Regression Analysis	80.75%	81.01%
Baye Classifier	81.49%	97.85%
Support Vector Machine	90.63%	99.12%

data points, and separate as many as possible into two different categories. To facilitate this purpose, SVM provides a set of hyperplanes including two margin hyperplanes, individually. The optimal hyperplane is the one that has the largest distance between two margin hyperplanes. As shown in Fig. 5, the black line denotes the optimal hyperplane, and is labeled maximum-margin hyperplane. The green and blue lines denote the two margin hyperplanes for either category, which can be called support vectors.

The thought notion of an optimal hyperplane was first suggested by Vapnik and applied to linear classification (Fig. 5(a)). He subsequently proposed an algorithm for nonlinear classification (refer to Fig. 5(b)) by using the kernel trick to maximize the gap between margin hyperplanes, which is the most important aspect of SVM theory. In addition to being applied to linear classification, several typical kernels applied in nonlinear classification have been proposed in SVM as follows:

- Linear kernel: $K(\mathbf{x}_i \cdot \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)$
- Polynomial: $K(\mathbf{x}_i \cdot \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d$
- Gaussian radial basis function: $K(\mathbf{x}_i \cdot \mathbf{x}_j) = exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j) \parallel^2$, for $\gamma > 0$, sometimes parametrized using $\gamma = \frac{1}{2}\sigma^2$
- Hyperbolic tangent: $K(\mathbf{x}_i \cdot \mathbf{x}_j) = tanh(\kappa \mathbf{x}_i \cdot \mathbf{x}_j + c)$, for some (not every) $\kappa > 0$ and $c > 0$

IV. Experimental Setup

A. Explanation of Variables

The SVM procedure mentioned in Section III is a binary classification to separate the customers into 2 classes. The response variable is the purchase behavior, defined as a Boolean variable 0/1, which denotes the unpurchased and purchased modes. Two explanatory variables are also employed in the SVM. One is age, which denotes the demographic characteristic of the customers, and the other is stay time, which denotes their behavioral attributes.

B. Accuracy Comparison

In this section, we compare the forecast accuracy of SVM mentioned in Section III-A with that of linear discriminant analysis, logistic regression analysis and a Bayes classifier.

Our experimental data were recorded from May 11,2009 to June 15, 2009. As such, we selected the data from May 11,2009 to June 10, 2009 as the training data (including 4776 sale transactions), assigning the remaining 5 days' worth of data as the testing data (including 885 sale transactions).

Table III shows the results of each forecast method. The SVM algorithm was implemented in Matlab [1]. The forecast procedures of linear discriminant analysis and logistic regression analysis were encoded using the programming language R[2]. The results of Bayes classifier is were taken from one of our previous studies [8]. The column of "Accuracy" denotes the hit ratio for the whole data set, both in the purchased and non-purchased groups. The forecast accuracy of SVM was only lightly higher than that of the other models. In the column of "Accuracy($P = 1$)" which denotes the hit ratio only for the data in the purchase state, SVM was much more accurate than the linear classification methods, lightly higher than Bayes classifier.

C. Multivariate Omnibus Normality Test

When the observed data are applied to an SVM to find an appropriate kernel trick for mapping the observed data, the multivariate normality test becomes necessary. For this issue, Doornik and Hansen suggested a convenient version of the omnibus test for normality [11].

If the input has n observations along a p-dimensional vector, they suggested that a $p \times n$ matrix $X' = \{x_1, x_2, \cdots, x_n\}$ be applied. The mean and covariance are $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'$, respectively. Then, create a matrix V the variances on the diagonal:

$$V = diag(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_p^2), \tag{12}$$

and form the correlation matrix $C = V^{-\frac{1}{2}} S V^{-\frac{1}{2}}$. Define the $p \times n$ matrix $Y' = y_1, \cdots, y_n$ from the transformed observations:

$$y_i = H\Lambda^{-\frac{1}{2}} H' V^{-\frac{1}{2}}(x_i - \bar{x}), \tag{13}$$

with $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$, the matrix with the eigenvalues of C on the diagonal. The columns of H are the corresponding eigenvectors, such that $H'H = I_p$ and $\Lambda = H'CH$. Using population values for C and V, a multivariate normal can thus be transformed into an independent standard normal; using only approximated sample values. Now, we can compute the univariate skewness and kurtosis for each of the p-transformed vectors of n observations. Defining $B_1' = (\sqrt{b_{11}}, \cdots, \sqrt{b_{1p}})$, $B_2' = (\sqrt{b_{21}}, \cdots, \sqrt{b_{2p}})$ and $\imath$ as a p-vector of ones, the test statistic:

$$E_p^a = \frac{bB_1'B_1'}{6} + \frac{n(B_2 - 3\imath)'(B_2 - 3\imath)}{24} \underset{a}{\widetilde{\ }} \chi^2(2p). \tag{14}$$

The proposed multivariate statistic is:

$$E_p = Z_1'Z_1 + Z_2'Z_2\chi^2(2p) \underset{app}{\widetilde{\ }} \chi^2(2p) \tag{15}$$

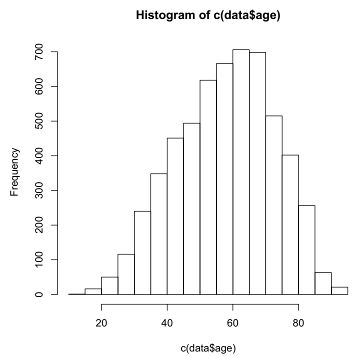
where $Z_1' = (z_{11}, \cdots, z_{1p})$ and $Z_2' = (z_{21}, \cdots, z_{2p})$ are determined by Eqs. (16) and (17) given in Appendix D. After transformation of the data to approximated and independently standard normals, the univariate test was applied to each dimension.

[1] The procedures and results of SVM are expressed in Appendix C.

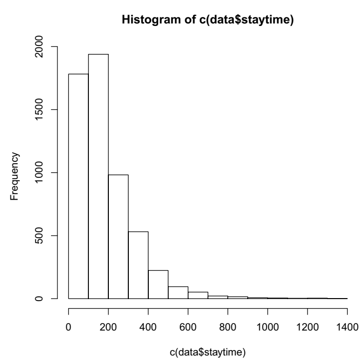
[2] The procedures and results of linear discriminant analysis and logistic regression analysis are expressed in Appendix A and Appendix B, respectively

TABLE IV
SVM PERFORMANCES WITH DIFFERENT KERNEL TRICKS

Kernel Type	Linear	Polynomial (p=2,⋯.5)				RBF (σ=0.2⋯..1.0)				
Parameter Value	-	p=2	p=3	p=4	p=5	σ=0.2	σ=0.4	σ=0.6	σ=0.8	σ=1.0
Accuracy	81.21%	89.32%	89.67%	89.47%	89.79%	89.42%	89.90%	90.02%	90.14%	90.63%
Modeling (sec)	42.73	166.34	198.84	268.80	266.20	2.01	1.65	1.31	1.25	1.15
Evaluating (sec)	0.04	0.03	0.03	0.01	0.01	0.08	0.09	0.08	0.14	0.09



(a) Age



(b) Stay time

Fig. 6. Frequency of explanatory variables

This research employed multivariate analysis to statistically process the massive amount of customer behavior data to facilitate kernel selection during the classification task of the SVM. A multivariate or multivariable normality test examined the data characteristics. We use this information in our experiment to choose the linear or non-linear kernel for SVM.

Before applying the multivariate normality test to our data, we constructed a bar graph to see the nature of individual attributes. As shown in Fig. 6, the age variable was normally distributed (Fig. 6(a)), but stay time was not (Fig. 6(b)). Therefore, the multivariate test was necessary to investigate on our data.

According to the results shown in Table IV, three types of kernelsv(Linear, Polynomial and RBF) were employed to test our data (Multivariate Omnibus Normality Test Performance was implemented in Matlab). If the kernel used was a Gaussian radial basis function (RBF), the optimal hyperplane could be constructed with parametrization using $\sigma = 1.0$. We observed from the experiments that the classification accuracy could improve with the higher RBF kernel parameters. However, we stopped at $\sigma = 1.0$ to follow the classification style of the SVM application.

V. CONCLUSIONS

In this paper, we have suggested a method for extracting consumer purchasing behavior. Utilizing RFID data acquired from individuals in a Japanese supermarket, we examined several important methodological issues related to the use of RFID data in support vector machines (SVMs) to predict purchasing behavior.

First, we provided a time perspective on shopping in a certain area instead of the entire grocery store. In contrast to shopping paths, customer stay time can help us improve our understanding of in-store behavior within a small range, which is one of the most important factors impacting one's purchasing decision. Stay time was also a meaningful variable for the retailers, enabling them better understand the purchasing behavior of specific items, rather than of the sales amount for the whole store. Second, we used SVM to apply to forecast purchase behavior, which was independent of the distribution and relationship of variables even though they were linear or nonlinear aspect of variables. In the numerical example, SVM demonstrated better forecasting performance related to linear discriminant analysis, logistic regression analysis and even Bayes classifier. Finally, because the variables of age and stay time had different distributions, we conducted a multivariate omnibus normality test on the data, and an optimal hyperplane was constructed when the RBF was applied as the kernel trick. We hope to build upon this work to the highest accuracy level possible for consumer behavior extraction.

ACKNOWLEDGMENT

This work was supported in part by MEXT Strategic Project to Support the Fomation of Research Bases at Private Universities (FY2014-2019) and MEXT Grant-in-Aid for Scientific Research (A) Grant Number 16H02034 (FY2016-2021).

REFERENCES

[1] Peter M. Guadagni and John D. C. Little, "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, Vol.2, No.3, pp.203-238, 1983.
[2] Sunil Gupta, "Impact of Sales Promotions on When, What, and How Much to Buy", *Journal of Marketing Research*, Vol.25, No.4, pp.342-355, 1988.

[3] Thomas S. Robertson and James N. Kennedy, "Prediction of Consumer Innovators: Application of Multiple Discriminant Analysis", *Journal of Marketing Research*, Vol.5, No.1, pp. 64-69, 1968.
[4] Herb Sorensen, "The Science of Shopping", *Marketing Research*, Vol.15, pp.30-35, 2003.
[5] Jeffrey S. Larson, Eric T. Bradlow and Peter S. Fader, "An Exploratory Look at Supermarket Shopping Paths", *International Journal of Research in Marketing*, Vol.22, No.4, pp.395-414, 2005.
[6] Sam K. Hui. Eric T. Bradlow and Peter S. Fader, "Testing Behavioral Hypotheses using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior", *Journal of Consumer Research*, Vol.36, No.3, pp.478-493, 2009.
[7] Katsutoshi Yada, "String Analysis Technique for Shopping Path in a Supermarket", *Journal of Intelligent Information Systems*, Vol.36, No.3, pp.385-402, 2011.
[8] Yi Zuo and Katsutoshi Yada, "Application of Bayesian Network Sheds Light on Purchase Decision Process basing on RFID Technology", *2013 IEEE 13th ICDM*, pp.242-249, 2013.
[9] Vladimir Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
[10] Yi Zuo, "Prediction of consumer purchase behavior using Bayesian network: an operational improvement and new results based on RFID data", *Int. J. Knowledge Engineering and Soft Data Paradigms*, Vol. 5, No. 2, pp.85-105, 2016.
[11] Jurgen A. Doornik and Henrik Hansen, "An Omnibus Test for Univariate and Multivariate Normality", *Oxford Bulletin of Economics and Statistics*, Vol.70, Issue Supplement s1, pp.927-939, 2008.
[12] Shinya Miyazaki, Takashi Washio and Katsutoshi Yada, "Analysis of Residence Time in Shopping Using RFID Data – An Application of the Kernel Density Estimation to RFID", *2011 IEEE 11th International Conference on Data Mining*, pp.1170-1176, 2011.

APPENDIX A
LINEAR DISCRIMINANT ANALYSIS

R Console

```
> result<-glm(Purchase~Age+Time)
> summary(result)

Coefficients:
             Estimate   Std.Error  t value  Pr(>|t|)
(Intercept)  6.950e-01  2.232e-02  28.161   <2e-16 ***
Age         -3.831e-05  3.684e-04  -0.104   0.917
Time         7.957e-04  3.775e-05  24.978   <2e-16 ***
------------
Signif.   0 '***'  0.001 '**'  0.01 '*'  0.05 '.'  0.1 ' '
```

APPENDIX B
LOGISTIC REGRESSION ANALYSIS

R Console

```
> result<-glm(Purchase~Age+Time, family=binomiial)
> summary(result)

Coefficients:
             Estimate   Std.Error  t value  Pr(>|t|)
(Intercept)  2.756e-01  1.651e-01  -2.445   0.0145 *
Age         -5.739e-03  2.695e-03  -1.673   0.0943 .
Time         1.227e-02  6.821e-04  23.393   <2e-16 ***
------------
Signif.   0 '***'  0.001 '**'  0.01 '*'  0.05 '.'  0.1 ' '
```

APPENDIX C
SUPPORT VECTOR MACHINE

Matlab Command Window

```
Number of variables: 2
Sample size: 5661
------------------------------------------------------------------
MV omnibus test statistic: 3807.942819
Equivalent degrees of freedom: 4.000000
P-value associated to the Royston's statistic: 0.000000
Lower critical value associated to the Royston's statistic: 0.484419
Upper critical value associated to the Royston's statistic: 11.143287
With a given significance = 0.050
Data analyzed do not have a normal distribution.
------------------------------------------------------------------
```

APPENDIX D
MULTIVARIATE OMNIBUS NORMALITY TEST

The transformation for the skewness $\sqrt{b_1}$ into z_1 is as follows:

$$
\begin{aligned}
\beta &= \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \\
\omega^2 &= -1+2(\beta-1)^{\frac{1}{2}}, \\
\delta &= \frac{1}{[log(\sqrt{\omega^2})]^{\frac{1}{2}}}, \\
y &= \sqrt{b_1}\left[\frac{\omega^2-1}{2}\frac{(n+1)(n+3)}{6(n-2)}\right]^{\frac{1}{2}}, \\
z_1 &= \delta log\left[y+(y^2+1)^{\frac{1}{2}}\right].
\end{aligned}
\tag{16}
$$

The kurtosis b_2 is transformed from a gamma distribution to χ^2, which is then translated into standard normal z_2 using the Wilson-Hilferty cubed root transformation as follows:

$$
\begin{aligned}
\delta &= (n-3)(n+1)(n^2+15n-4), \\
a &= \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta}, \\
c &= \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta}, \\
k &= \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta}, \\
\alpha &= a+b_1c, \\
\chi &= (b_2-1-b_1)2k, \\
z_2 &= \left[\left(\frac{\chi}{2\alpha}\right)^{\frac{1}{3}} - 1 + \frac{1}{9\alpha}\right].
\end{aligned}
\tag{17}
$$