

18th International Conference on Knowledge-Based and Intelligent
Information & Engineering Systems - KES2014

Consumer purchasing behavior extraction using statistical learning theory

Yi Zuo^{a,*}, A B M Shawkat Ali^b, Katsutoshi Yada^{a,c}

^a*Data Mining Laboratory, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8086, Japan*

^b*School of Engineering and Technology, Central Queensland University, Rockhampton, QLD 4702, Australia*

^c*Faculty of Commerce, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8086, Japan*

Abstract

Consumers classification is one of the most important task in the retail sector. RFID (Radio Frequency IDentification) - A wireless non-contact technology is made easier to classify the consumers' in-store behavior, recently. This paper presents an extraction of consumer purchasing behavior using statistical learning theory SVM (Support Vector Machine). In this research, we present our recent investigation outcome on the consumers shopping behavior in a Japanese supermarket using RFID data. We observe that it is possible to express the individual difference of consumers how are they spending time (we call it stay time in this paper) on shopping in a certain area of the supermarket. The contribution of this research is in two folds: we employ a SVM model on dealing with the RFID data of the consumer in-store behaviour firstly, as compared with other forecast model such as linear regression analysis and bayesian network, SVM provides a significant improvement in the forecasting accuracy of purchase behaviour (from 81.49% to 88.18%). Secondly, the kernel trick is adopted inside the SVM theory to choose the appropriate kernel for consumer purchasing behavior extraction.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of KES International.

Keywords: RFID; Consumer Behavior; Support Vector Machine; Kernel Trick; Classification

1. Introduction

Since the first POS (Point Of Sale) system has been applied in a supermarket in 1970s, POS data has been increasingly recognized to be a key strategic resource and considerably investigated to understand purchase behavior of customers. Most of business models in this field are being applied to examine and analyze the purchase behavior in order to lead the sales increase^{1,2}. The POS data can provide a most direct and effective approach to help the manager and retailer understand the purchase behavior of their customers. However, the utility of POS data has been indicated in supermarket's success, it still sheds no light on the in-store behavior of customers.

* Corresponding author. Tel.: +81-6-6368-1228 (ex. 4534) ; fax: +81-6-6330-3304.

E-mail address: zuoyisaki@yahoo.co.jp

A wireless non-contact technology named RFID (Radio Frequency IDentification) has brought a new perspective on this situation. In the early 2000s, the first in-store behavior experiment is carried out by Sorensen Associates³. As an object attached with a small RFID tag, the movements can be identified and tracked automatically. One of the main advantages of RFID technology in the research field of marketing is that it can accurately capture the in-store behavior of the customers in a supermarket. In Sorensen's research, the position of shopping carts and baskets with a RFID tag attached are used as surrogates for the position of the customers and transmitted their position information to the back-end server as the RFID data. Basing on the RFID data, Larson et al.⁴ present an exploratory work and they identified 14 canonical path types as the typical of the grocery store travel. However, they do not take account of the affect of in-store behavior on purchase decision, this work familiarizes us with the existence of RFID data. On the other hand, a significant problem is pointed out by Yada⁵. Regarding customer shopping path analyses, one of the most important information is missing on time spent in each section. However, the component ratio of time in each section is used as a class label to represent classification rules in his paper, a quantity analysis of stay time how to cause a change in purchase trend is still necessary.

This paper employs SVM model to resolve this issue. The original SVM algorithm is invented by Vapnik⁶ and his team in the end of last century at Royal Holloway College, London. The contribution of this research is in two folds: we develop a SVM model on dealing with the RFID data of the consumer in-store behavior first time. As compared with other forecast model such as linear regression analysis and bayesian network, SVM provides a significant improvement in the forecasting accuracy of purchase behavior (from 81.49% to 88.18%). Secondly, multivariate analysis is applied to statistically process the massive amount of customers' stay time to make easier kernel selection for the classification task with SVM. Basically kernel trick provides the insight information of data instead of computing the mapping explicitly. The appropriate kernel selection can generate an insightful decision function, which makes sense of the consumer purchasing behavior extraction.

The remained paper is organized as follows. In Section 2, a review of literatures and models on research topic are discussed briefly. In Section 3, the framework of our RFID system and the preprocessing stage of RFID data are represented. Accuracy comparison of SVM and its application to purchase decision-making are explained in Section 4 and 5. The results and conclusion are summarized at the end of this paper in Section 6.

2. Recent advancement on research topic

2.1. Literature review of RFID data

In the early 2000s, an in-store behavior research is carried out in a supermarket in Western United States by Sorensen Associates³. With the RFID tags attached to the bottom of shopping carts, they emit the signal of position which is received by receptors throughout the whole supermarket. Each sale transaction is extended to a shopping tracks in 5-second interval, and recorded as the RFID data into the database.

Basing on Sorensen's RFID data, Larson et al.⁴ presented an exploratory work to identify a total of 14 canonical path types as the typical of the grocery store travel. However, the affect of in-store behavior on purchase decision has not been empirically investigated until recently. This work familiarizes us with the existence of RFID data and also leaves to us examining many business implications for store managers in detail.

In Hui et al.'s work⁷, they estimated an integrated model on RFID data, which records in-store path of each customer and also be linked to POS data for the items purchased in their shopping path. Due to the improved understanding of customer in-store behavior, implications of their research for store layout decisions are briefly discussed. Also, a set of behavioral hypotheses on customers' visit, shop and purchase decision are established and tested, in which "As a consumer spends more time in the store, she becomes more likely to be in a shopping mode when in a particular zone." is suggested in their paper.

In one of Yada's works⁵, character string analysis technique is applied to analysis the shopping path in order to seek visit patterns of customers who purchase a large number of items. Moreover, the ratio of stay time in each section is used as an explanatory attribute to separate the customers into "high-volume" ones or "low-volume" ones. However, the researches above have indicated the usability of RFID data, such as the stay time obtained from RFID data is only used as a clustering indicator. Even though its positive effect on purchase behavior has been proved qualitatively, a quantity analysis of stay time how to cause a change in purchase trend is still necessary.

2.2. Literature review of forecast method

2.2.1. Linear discriminant analysis

As the purchase behavior is a typical 0/1 binary variable, linear discriminant analysis is one of the most typical models used to predict⁸. In linear discriminant analysis, the observation variable Y is approximated with the independent variables $\{X_1, X_2, \dots, X_n\}$ in a linear combination and the coefficients are generated from a regression analysis as follows:

$$Y = \alpha_0 + \sum_{i=1}^n \alpha_i X_i \quad (1)$$

where the notation α_i is the coefficient parameter of model and determined by minimizing the sum of squares error.

2.2.2. Logistic regression analysis

Logistic regression analysis is one type of linear regression analysis which is used extensively medical and social science fields for predicting binary variable¹. In logistic regression analysis, the observation variable Y is approximated with the independent variables $\{X_1, X_2, \dots, X_n\}$ as follows:

$$\text{logit}(Y) = \ln\left(\frac{\Pr(Y)}{1 - \Pr(Y)}\right) = \beta_0 + \sum_{i=1}^n \beta_i X_i \quad (2)$$

where the $\Pr(Y)$ is the output probability of variable Y between 0 and 1. The coefficient parameter β_i is the same as Section 2.2.1 Logistic regression analysis is also applied in marketing researches for predicting a customer's tendency to purchase a product.

2.2.3. Bayesian network

Bayesian network (BN) is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph. In one of our previous studies⁹, BN is employed to predict purchase behavior. The probability theory of BN is based upon the bayes' rule, and if two observed events have the relation like $A \rightarrow B$ can be expressed as follows:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}. \quad (3)$$

$\Pr(A)$ and $\Pr(A|B)$ denote the prior probability of event A and the posterior probability of event A when A according to the event B , respectively. $\Pr(B|A)$ denotes the likelihood function. The denominator $\Pr(B)$ is equal to $\sum_i \Pr(B|A = a_i)\Pr(A = a_i)$ which denotes the marginal distribution in all states of event A . By using Eqn. (3), the prior probability $\Pr(A)$ is revised to $\Pr(A|B)$ by multiplying the likelihood $\Pr(B|A)$. Furthermore, due to BN can represent non-linear and probabilistic estimation among a set of variables, a revision to the traditional understanding of the stay time that the longer stay time not always having a positive effect on purchasing is demonstrated.

2.2.4. Support vector machine

Support Vector Machine (SVM) belongs in the supervised learning theory group which is comparatively very effective for classification, regression and clustering tasks. Comparing with other learning algorithms it could effectively handle high dimensional data space due to its unique ingredient of kernel. Different kernel functions can easily generate a set of decision functions even when the number of dimensions are greater than the total samples. In the data modeling phase it learns a very little amount of data points those are close to data separating hyperplane and we call support vectors. Therefore SVM acts in the learning space as a memory efficient learning algorithm.

Let us consider l i.i.d. sample: $(x_1, y_1), \dots, (x_l, y_l)$, where x_i for $i = 1, \dots, l$ and $y_i = \{+1, -1\}$ is the class label for data point x_i . To obtain a better general decision surfaces, one can first nonlinearly transform a set of input vectors $x_1, \dots, x_l$ into a high-dimensional feature space, and the decision function f that can be written as

$$f(\mathbf{x}) = h(\mathbf{x}) + b \quad (4)$$

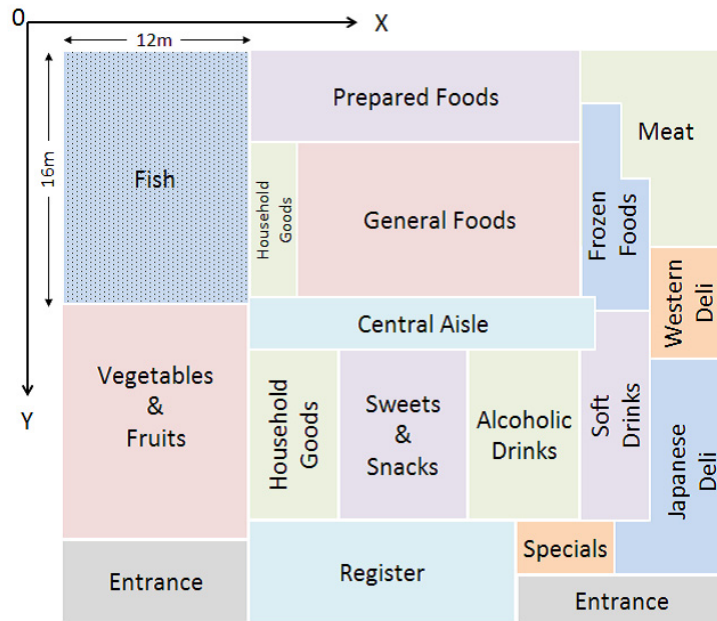


Fig. 1. Layout of the supermarket

where $h(\mathbf{x}) = \sum_{i=1}^l y_i \alpha_i (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}))$. By using the kernel trick, the inner product can be replaced by $K(\mathbf{x}_i \cdot \mathbf{x})$. Therefore the final decision function become as follows:

$$f(\mathbf{x}) = \sum_{i=1}^l y_i \alpha_i K(\mathbf{x}_i \cdot \mathbf{x}) + b \quad (5)$$

where $K(\mathbf{x}_i \cdot \mathbf{x})$ is the most import ingredient in SVM theory which is called kernel. Among all of hyperplanes, the best hyperplane ($f(\mathbf{x}) = 0$) can be found when the distance between two margin hyperplanes ($f(\mathbf{x}) = -1; f(\mathbf{x}) = 1$) is maximized. In this paper, SVM is applied to customers classification depending on the value of Eqn. (5) to separate into 2 classes as follows:

$$y_i = \begin{cases} 1 & f(\mathbf{x}_i) \geq 1, \\ -1 & f(\mathbf{x}_i) \leq -1. \end{cases}$$

3. Overview of system and RFID data

3.1. Collection of RFID data

In addition to customers' movement data, floor layouts and purchasing history data were also gathered. The floor layout within the store was divided into 16 sections (Fig. 1). Some of those sections (e.g. Household Goods and General Foods) had subsections, and in total there were 28 subsections.

In order to record the trip of customer, the layout is reproduced into a picture from x and y coordinates on the scale of 15.7 pixels per meter. While the customer passes a certain area of the supermarket with a shopping cart attached RFID tag, the information of her position can be received by RFID receptor around the shelves and be transformed to a pixel point into our dataset using the floor layout matching. RFID tag number attached to shopping cart, shopping date, time stamp, x and y coordinates of that time stamp, section of that coordinate and elapsed time are recorded, and Table 1 shows the sample data obtained using our RFID system.

Table 1. RFID data of the movement

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

Table 2. Detail of the POS data

Customer No.	Date	Time	Item Name	Item Category	Volume	Amount
Lucy	2009/05/11	12:12:30	Cabbage	Vegetable	1	150
Lucy	2009/05/11	12:12:30	Banana	Fruit	1	198
Lucy	2009/05/11	12:12:30	Sashimi	Fish	2	596
Lucy	2009/05/11	12:12:30	Pork	Meat	1	232

When the customer comes to the checkout register and purchases, the POS data what she has bought is recorded and transformed into our dataset. The dataset is the shopping details as shown in Table 2, and the columns denote customer name, shopping date and checkout time, category of the purchased item, its volume and unit price, respectively.

From the customer coming into the entrance until the purchase is completed, we define this process as a basic unit of customer's in-store behavior and give a unique ID to identify it. Also, by using this ID, the customer's purchase behavior obtained from POS data is linked to her in-store behavior. After pre-processing of the RFID data and POS data, we get 6883 shopping units(sale transactions) from the experiment, in which there are 2847 customers be tracked.

3.2. Measuring range of fish selling area

The experiment was carried out in a typical supermarket in Japan. Comparing to the previous studies, we focus on the customer in-store behavior in a certain area instead of the whole supermarket. Since fish is featured much more prominently on the Japanese plate, the fish selling area is selected as the experiment object. And, the measuring range where length equals 16 metres and width equals 12 metres is shown as the shadowy pattern in Fig. 1.

3.3. Definition of stay time

In this section, we explain the definition of the stay time for the customers how they spent in fish selling area. For given a customer, her shopping trip which is tracked from her coming into the entrance until coming to the checkout register to purchase is tracked by the RFID tag.

Fig. 2 shows a density estimation of stay time using one customer's shopping trip¹⁰. The figure has been drawn using line segments to connect coordinate points of the customer's trip tracked per second, and is combined with the stay time distribution expressed by the density estimation method. In Miyazaki et al.'s study¹⁰, they reveal that using only the shopping trip is difficult to know how does customer spend time in a certain area.

Thus, the analysis of stay time is required. If customer Lucy remained in position (x_i, y_i) is t_i seconds, then the time spent by Lucy in the supermarket is expressed as follows:

$$T = \sum_{i=0}^n t_i \quad (6)$$

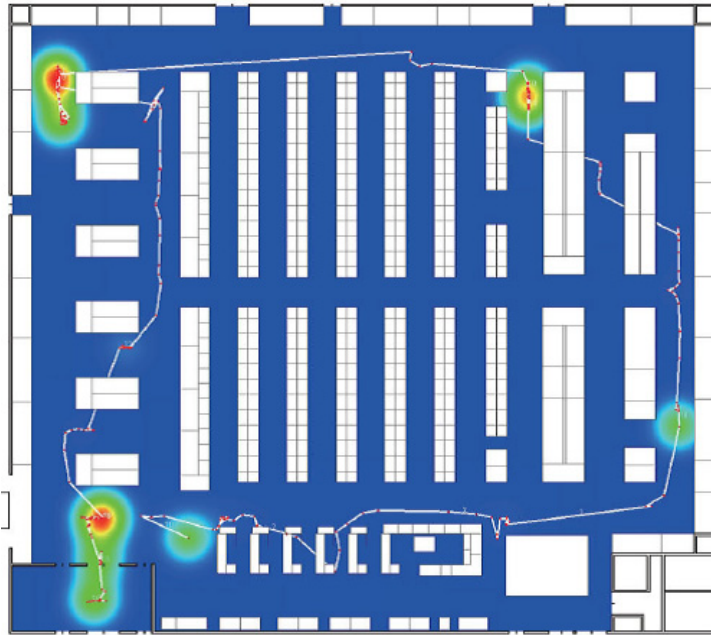


Fig. 2. Visualization of stay time from shopping trip¹⁰

where the notation t_i denotes the “Elapsed Time” shown in the Table 1. And making an addition to Eqn. (6), only if she comes into the fish selling area, this position is accepted. Therefore, the stay time T_{Fish} of customer spent in fish selling area is defined as follow:

$$T_{Fish} = \sum_{i=0}^n t_i, \quad t_i = \begin{cases} \text{Elapsed Time, if position in fish selling area.} \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

By using the Eqn. (7), the stay time is calculated for the individual customer spent in the fish selling area, respectively.

4. Experimental setup

4.1. Variable explanation

A successful decision on purchase is basically depending on the customer response to the products or supermarkets. In order to understand the process of purchase decision, customers are classified into homogeneous groups in which they have the similar purchase behavior and characteristics. This can also serve to clarify the variations of purchase behavior among different groups.

In most cases, gender and age are nature to be recognized as the customers characteristics. However, both of them have no new ideas (at least when they are used to predict), they are still powerful and widely used in nowadays researches to separate customers into clusters. In this study, 5661 sale transactions are collected and only 535 of them are male (<10%). As this type data having minority sample data can cause a extreme skewness so as to impair the generalizing capability of classifier, we only use age as the explanatory variable. Another explanatory variable is stay time which can denote the in-store behavior of customers. In contrast to the previous studies about stay time in which the length of time is used as the classification attribute, the aim of this paper is to demonstrate the change of purchase intention over time. Furthermore, in addition to the age, the variations (sensitivities) of this change in different age brackets can be are illustrated from the customer perspective. For all the 5661 sale transactions, the stay time is calculated by Eqn. (7) for the fish area, individually.

Table 3. Comparison of forecast method

Forecast Method	Accuracy	TPR ($P = 1$)
Linear Discriminant Analysis	81.09%	81.29%
Logistic Regression Analysis	80.75%	81.01%
Bayesian Network	81.49%	97.85%
Support Vector Machine*	88.18%	98.66%

*The results of Support Vector Machine is predicted by using gaussian radial basis function with $\sigma = 0.4$.

Table 4. SVM performances with different kernel trick

Kernel Type	Linear	Polynomial ($d = 2 \cdots 5$)				RBF ($\sigma = 0.2 \cdots 1.0$)				
Parameter Value	-	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$\sigma = 0.2$	$\sigma = 0.4$	$\sigma = 0.6$	$\sigma = 0.8$	$\sigma = 1.0$
Training Error	0.1992	0.1135	0.1098	0.1077	0.1197	0.1085	0.1075	0.1092	0.1089	0.1089
Modeling(sec)	0.83	1.40	3.10	12.24	108.97	2.62	2.50	2.43	2.40	2.42
Evaluating(sec)	0.04	0.07	0.07	0.07	0.06	0.14	0.13	0.11	0.11	0.11

The response variable is the purchase behavior defined as a boolean variable F/T, which denotes unpurchased mode and purchased mode (purchased in fish area or not), respectively. The procedure of SVM mentioned in Section 2.2.4 is a typical binary classification. By employed age that denotes the demographic characteristic of customers and stay time denotes the behavioral attribute, SVM can separate the customers into 2 classes.

4.2. Accuracy comparison

In this section, we compare the forecast accuracy of SVM with that of linear discriminant analysis, logistic regression analysis and bayesian network mentioned in Section 2.2.

Since the period of experiment data is from May 11, 2009 to June 15, 2009, we select one month from May 11, 2009 to June 10, 2009 as the training data (including 4776 sale transactions), and the remained 5 days is the testing data (including 885 sale transactions).

Table 3 shown the results of each forecast method. The algorithm of SVM is implemented in programming language R with “kernlab” package, also the forecast procedures of linear discriminant analysis and logistic regression analysis are encoded using the programming language R with “MASS” package. The results of bayesian network is one of our previous studies⁹. The column of “Accuracy” denotes the hitting ratio on the whole data both in purchased and non-purchased. And, the forecast accuracy of SVM is much higher than other models. In the column of “TPR ($P = 1$)” which denotes the hitting ratio on the data only in purchase state (True Positive Rate - how many positive results predicted correctly to all positive samples in the test data), SVM also has shown a much higher accuracy than linear and other classification methods.

4.3. Kernel selection

When the observed data is applied into SVM, in order to find an appropriate kernel to map the observed data, several typical kernels tested in linear/nonlinear classification are 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: $K(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, for $\gamma > 0$. Sometimes parametrized using $\gamma = \frac{1}{2}\sigma^2$.

According to Table 4, these three kinds of kernels (linear, polynomial and RBF) are employed to test on our data, also polynomial and RBF are tested by different hyperparameters (degree d of polynomial kernel is $d = 2 \cdots 5$; variance σ of RBF kernel is $\sigma = 0.2 \cdots 1.0$).

When the linear kernel is used as $K(\cdot)$, the classification results are shown in Fig. 3(a). And from this figure, it is obvious that the observed data is not linearly separable. All of the data is assigned into purchase state.

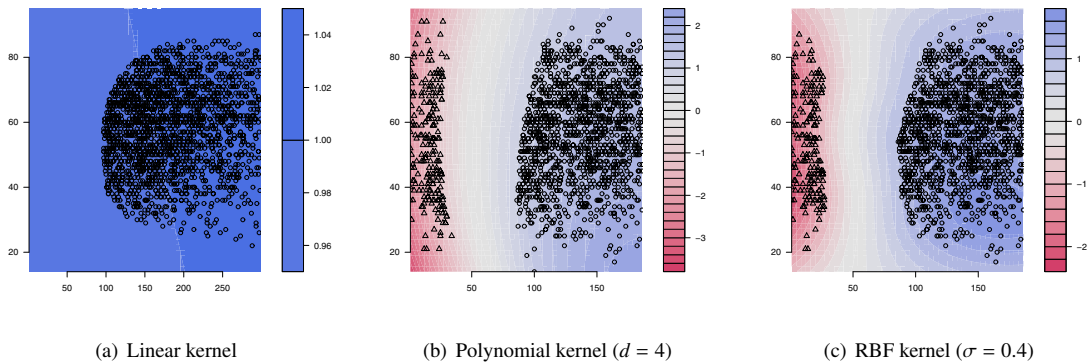


Fig. 3. Results of the classification by applying different kernel tricks

In polynomial kernel, if degree $d = 1$, polynomial kernel is the same as linear kernel that hyperplane is a level surface. With the increasing adjustment of d value by 1 to $d = 5$, the training error is much lower than linear kernel shown in Table 4. Fig. 3(b) showed the classification results of polynomial kernel used $d = 4$, and two classes are sufficiently separated by a smooth curved hyperplane.

Gaussian radial basis function (RBF) is one of the most efficient kernel, which is widely used in a variety of classification analysis. As shown in the Table 4, the training error of RBF can reach almost the same accuracy as polynomial, but only costing short calculation time (Calculation time of polynomial is exponential growth with d increasing; Calculation time of RBF is minor change even in a certain large σ). Furthermore, when RBF is parametrized using $\sigma = 0.4$, the training error is minimum in all of the testing. And, the optimal hyperplane constructed with $\sigma = 0.4$ is shown in Fig. 3(c), which is irregular margin plane. However, the classification accuracy can be improved with the higher RBF kernel parameter values σ , we stop at $\sigma = 1.0$ to follow the classification style of SVM application.

5. Implications for business

5.1. Probabilistic outputs of decision function

As mentioned in Section 2.2.4, after decision function Eqn. (5) replacing by kernel function, it becomes easier to calculate the discriminant value. Fig. 4 shown a curved surface by using RBF kernel with $\sigma = 0.4$. However, this figure can be intuitive to represent the features of predicting basing on the explanatory variables, it is still difficult to interpret for managers and retailers how to understand the discriminant value. Furthermore, if curved surface is drawn by using other kernel e.g. polynomial, the discriminant value would be infinite value so that the results of SVM provide no meaningful information from the study of customer in-store behavior except the high accuracy of customers classification.

Regarding this issue, a calibrated value of output from decision function i.e. posterior probability, is required which can also be applicable to business decision-making implications of purchase behavior. Platt¹¹ proposed a sigmoid function basing on the decision values in case of classification. As bayes' rule is also applied, the posterior probability of purchase behavior ($P = 1$) is rewritten from Eqn. (5) as follow:

$$\Pr(P = 1|f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \quad (8)$$

where A and B are parameters that can be calculated by minimizing the likelihood.

5.2. Decision-making implications of purchase probability

By using the Eqn. (8), the change of purchase probability can be represented well as cumulative distribution. Here, we set age brackets in 10-age interval to separate customers into homogeneous groups, and the mean age of

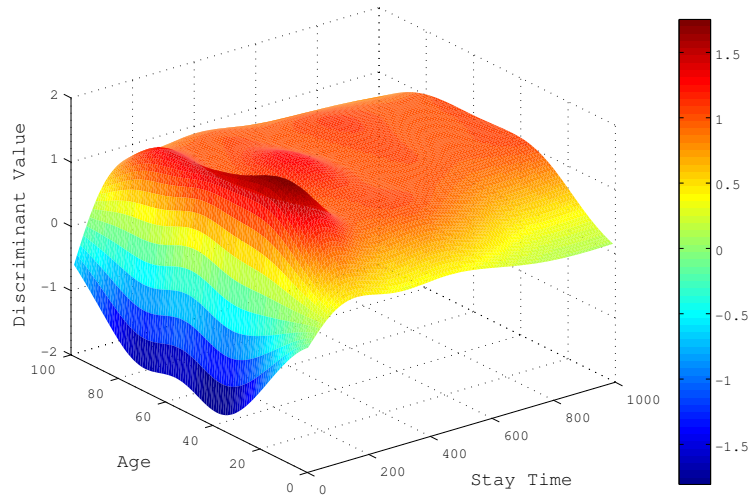


Fig. 4. Discriminant value of decision function

each group is used as surrogate to demonstrate the change of purchase intention over time. As shown in Fig. 5, the solid lines denote the posterior purchase probability of each age bracket and the dashed line denotes the prior probability (78.23%) of purchase. However, in most of the predicting models the decision threshold is 50%. In this case (prior probability > 50%), the prior probability can tell the managers and retailers a meaningful information that the customers who have a strong purchase intention rather than to classify the customers who purchased and who not purchased.

Firstly, to get an overview of Fig. 5, the results of SVM supported our previous studies⁹, that stay time has non-monotonic effect on purchase probability. Additionally, the results of SVM can also show much more details which can be analyzed quantitatively. However, junior customers have a higher purchase probability in the initial time, they need to spent much more time when the posterior purchase probability exceeds prior probability. 50s customers (purple line) are the first to come into purchase mode after 76 sec. 10s customers (blue line) are the last to come into purchase mode after 115 sec. After that, the purchase probability of each customer would reach their peak around 140 sec to 180 sec. However, the purchase intention of all the customers become weak, only 10s (blue line) and 20s (green line) come into unpurchase mode after 264 sec and 735 sec, respectively.

Therefore, we suggest managers to improve sales promotion e.g. point of purchase (POP) advertising which can attract customers' attention to spent much more time so as to make them be more potential to come into purchase mode. Meanwhile, managers also need to take account of the junior customers as a special case, sales promotion that supplements advertising should persuade them to come into purchase before the inflection point.

6. Conclusions

An extraction of consumer purchasing behavior is suggested in this paper. Via an investigation in Japanese supermarket basing on RFID data, we examined several important methodological issues related to the use of RFID data to predict purchase behavior by using the Support Vector Machine (SVM). At first, we provided a time perspective on shopping in a certain area instead of the whole grocery store. In contrast with the shopping paths, the stay time can help us to improve the understanding of customer in-store behavior in a small range, which is also one of the most important factor to effect the purchasing decision. It was also meaningful for the retailers, since they can get hold of purchase behavior of some special items better, rather than the sales amount of the whole store. Secondly, we employ SVM to apply to customers classification and purchase behavior forecast, which was not depending on the distribution and relation of variables even though they are linear or non-linear. Through the numerical example, SVM

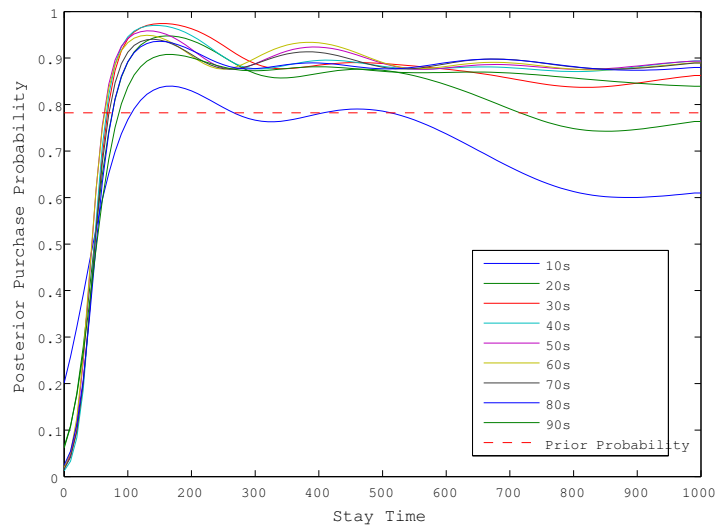


Fig. 5. Posterior probability of purchase among different age brackets

shown a higher forecast performance, especially in predicting customers in purchase mode than linear discriminant analysis, logistic regression analysis and even bayesian network. Finally, by using the probabilistic output of SVM, we suggested business implications to assist retailers and managers to understand the process of purchase behavior and improve the purchase decision of their customers via measuring stay time.

As a continuation of this research our future aim is to reach the highest accuracy level for consumer behavior extraction. Also, we desire to generate a new kernel function which can be customized into mapping various of actual situation from a business perspective.

Acknowledgements

This work was supported in part by MEXT Strategic Project to Support the Fomation of Research Bases at Private Universities (FY2009-2013) and MEXT Grant-in-Aid for Young Scientists (B) Grant Number 25780277.

References

1. Guadagni, P.M., Little, J.D.C.. A logit model of brand choice calibrated on scanner data. *Marketing Science* 1983;**2**(3):203–238.
2. Gupta, S.. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research* 1988;**25**(4):342–355.
3. Sorensen, H.. The science of shopping. *Marketing Research* 2003;**15**:30–35.
4. Larson, J.S., Bradlow, E.T., Fader, P.S.. An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing* 2005;**22**(4):395–414.
5. Yada, K.. String analysis technique for shopping path in a supermarket. *Journal of Intelligent Information Systems* 2011;**36**(3):385–402.
6. Vapnik, V.N.. *The Nature of Statistical Learning Theory*. New York, NY: Springer-Verlag New York Inc; 1995.
7. Hui, S.K., Bradlow, E.T., Fader, P.S.. Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of Consumer Research* 2009;**36**(3):478–493.
8. Robertson, T.S., Kennedy, J.N.. Prediction of consumer innovators: Application of multiple discriminant analysis. *Journal of Marketing Research* 1968;**5**(1):64–69.
9. Zuo, Y., Yada, K.. Application of bayesian network sheds light on purchase decision process basing on rfid technology. *2013 IEEE 13th International Conference on Data Mining Workshops* 2013;:242–249.
10. Miyazaki, S., Washio, T., Yada, K.. 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 Workshops* 2011;:1170–1176.
11. Platt, J.C.. Probabilities for sv machines. In: Smola, A., Bartlett, P., Schoelkopf, B., Schuurmans, D., editors. *Advances in Large Margin Classifiers*. Cambridge, MA: MIT Press; 2000, p. 61–73.