「정보시스템연구」 제26권 제4호
한국정보시스템학회
2017년 12월, pp. 17~37

http://dx.doi.org/10.5859/KAIS.2017.26.4.17

An Application of Support Vector Machines to Customer Loyalty Classification of Korean Retailing Company Using R Language

Nguyen, Phu-Thien* · Lee, Young-Chan**

〈Content〉

Ⅰ. Introduction

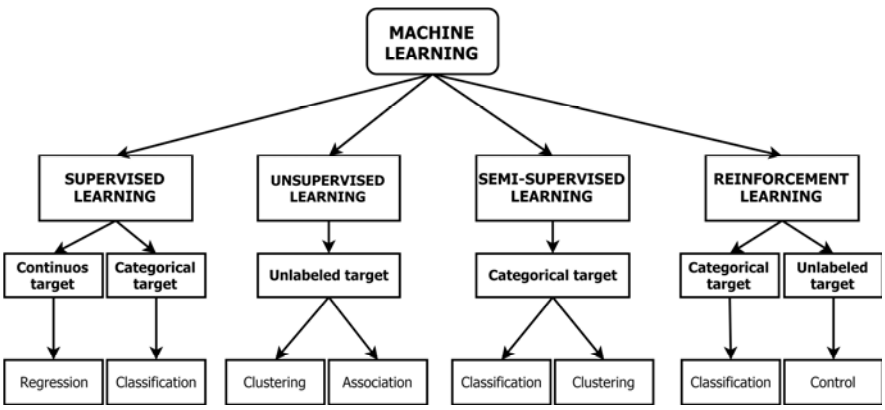
Machine learning and data analytics, which more and more popular on every area of life, are the trend of future. From IBM's lab, Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "machine learning" in 1959. According to Samuel, machine learning is the field of study that gives "computer the ability to learn without being explicitly programmed". Blum and Mitchell (1998) suggests a bit more detail definition: "A computer program is said to learn from experience E with respect to

* International Business Cooperative Course, Graduate School of Dongguk University, thiennguyenphu@dongguk.ac.kr
** Dept. of Business Administration, Dongguk University, chanlee@dongguk.ac.kr, corresponding Author

some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." Although there are many different definition of machine learning, they have a common point that is giving machines "intelligence" to solve tasks. Recently, machine learning divided into 4 main categories: Supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. While supervised learning "learns" from labeled data to construct model and predict new data, unsupervised learning detect patterns to clustering unlabeled data into meaning group, semi-supervised learning uses both supervised and unsupervised learning to construct model and reinforcement learning maps situations to actions to maximize a numerical reward signal. Today Machine Learning categories are shown in <Figure 1>

Customer loyalty is center of Customer Relationship Management (CRM), exploring CRM data to find potential customers is important task of every firm and organization, from small and medium businesses like a coffee shop to large corporations like Apple or Microsoft. In business, if the task of making new customer is as silver valuable, then keeping the old ones is gold. There are many benefits to keep long-term customers stay, including the fact that they are more likely to introduce your business to others via word-of-mouth recommendation, they are also more likely to purchase other products from your firm.

In the retail industry, it seems as though companies are constantly faced with the issue of trying to find new customers. Most of companies are obsessed with making sure their advertising, displays, and pricing all "scream out" to attract new business. This focus on pursuing new customers to increase sales is certainly prudent and necessary, but, at the same time, it can wind up hurting them. Therefore, the company focus really should be



<Figure 1> Machine learning main categories

on the loyal customers - the 20 percent of clients who currently are their best customers. Because the loyalty is important, there is a need of distinguishing customers into loyal and general groups.

In this study, we used a popular data mining algorithm called Support Vector Machines (SVMs) with the support of combination between Random Forest (RF) and Recursive Feature Elimination (RFE) algorithms, applied to “marketing personalization” use case of machine learning to state which factors can affect to customer loyalty of a Korea retailing firm. Although SVMs have shown excellent performance in a wide range of research areas, from bio-informatics, beat and text recognition, face authentication, customer churn prediction and so on, but there are only a few application of SVMs on CRM database to classify customers. We also used R, the most popular Artificial Intelligence (AI)/Data Analysis tool for the last several years, as developing language. According to a statistical poll carried in May 2017 on kdnuggets.com website, which shown in <Table 1> below, although Python has firstly caught up with R in Analytic/Data science tools population, R is still a large usage tool in machine learning area.

With flexible and large supporting Application Programming Interface (APIs); R is being used to run machine learning algorithms in both science fields and application technology. R programing language was lead the population ranking for many last years and became favorite development platform of many Artificial Intelligence/Data Mining developers. Above advantages of SVMs and R programing language are reasonable to apply these technologies into classifying loyal customers.

<Table 1> Top Analytics/Data Science Tools in 2017 KDnuggets Poll

Tool	2017 % Usage	% change 2017 vs 2016
Python	52.60%	15.00%
R language	52.10%	6.40%
SQL language	34.90%	-1.80%
RapidMiner	32.80%	0.70%
Excel	28.10%	-16.00%
Spark	22.70%	5.30%
Anaconda	21.80%	37.00%
Tensorflow	20.20%	195.00%
scikit-learn	19.50%	13.00%
Tableau	19.40%	5.00%
KNIME	19.10%	6.30%

With this model, we are trying to find out which factors impact to the customer loyalty on Korean retailing firm, and also to find whether these factors are related to loyalty by a particular formula? The result will be then compared with other machine learning to evaluate model performance.

Ⅱ. Literature Review

2.1 Customer loyalty in Korean retailing industry

Generally, retailing company need to be communicating with these customers on a regular basis method such as telephone, mail, email, social media, etc. These people are the ones who can and should influence the company buying and merchandising decisions. A loyal customer always feels better when the company solicit their input and showing them how much you value it. The more the company do for them, the more they will recommend the company to others. Positive word of mouth is “gold” for business. In terms of customer loyalty, Net Promoter Score (NPS) still acts as popular and simplest method to measure with only one question. Since many other researchers assert that customer loyalty is a multidimensional this method has started a long-term argument. Lin and Wang (2006) constructed a model to address that perceived value, trust, habit, and customer satisfaction are significant factors in determining customer loyalty; So et al. (2007) confirmed the positive impact of logistics service quality and relationship orientation on customer loyalty through customer satisfaction in electronic commerce; Keiningham et al. (2007) examined different customer satisfaction and loyalty metrics and evaluated their relationship to customer retention; Lee (2005) suggested that user satisfaction, learning cost, transaction fee, and reputation positively related to customer loyalty in online stock trading company; and Zaki et al. (2016) confirmed that single metrics alone cannot predict customer loyalty and consequently are unlikely to deliver actions to managers. In agreeing with these researchers, we suppose that customers’ loyalty-based behaviors are multidimensional and therefore a better measurement tool is required.

In addition, many researchers and academics agree that customer loyalty is an important part in the business as it is considered as the "back-bone of the business" (Gremler and Brown, 1996). By classifying customers, the company could benefit a lot by analyzing those loyalty patterns in its own market. The company should start by its completely loyal customers for better identification of its target market, thus increasing the demand on their products (Kotler and Armstrong, 2010). Moreover, there are some risk of misclassifying loyal customer. Marking loyal customer as non-loyal one not only leaks company's profit but also makes customers feel like they are not respected as they deserve. Vice versa, treating non-loyal customers as loyal ones might consume company's time and resource for nothing and thus, reduce company's performance.

2.2 Support Vector Machines (SVMs)

The SVMs central idea is to define a hyperplane that differentiate observations into two classes. Considering the example of classifying observations into positive class and negative class shown in <Figure 2>, suppose we have a set of positive and negative points has shown below. The task now is finding the function that defines the hyperplane with maximum width of margin.

Suppose we got a vector w that perpendicular to the hyperplane and a unknown vector u represents for vector from point O to an unknown observation. The projection of vector u on vector w always gives us a constant C or:

$\overline{w}.\overline{u}+b=0$ where b = -C (1)

So we have the classification rules:

$\overline{w}.\overline{u}+b \geq 0$ Then u is positive (2)

$\overline{w}.\overline{u}+b \leq 0$ Then u is negative (3)

With b is the bias depended on which the point O to be chosen.

Because the original dot of axis is freely chosen, for any known observation x, we also have:

$\overline{w}.\overline{x_+}+b \geq 1$ (4)

$\overline{w}.\overline{x_-}+b \leq 1$ (5)

Given yi such that y_i = 1 for positive observations and y_i = -1 for negative observations, the (4) and (5) functions then become:

$y_i(\overline{w}.\overline{x_i}+b)-1 \geq 0$ (6)

The formula (6) will get the value 0 when x is on the margin lines.

As we can see in <Figure 2>, the margin of hyperplane is the length of vector $\overline{x_+}-\overline{x_-}$ when this vector perpendicular to the hyperplane (or when α = 90^o). Since the vector w is normal vector of the hyperplane, the distance between 2 margins is:

$$margin=(\overline{x_+}-\overline{x_-})\frac{\overline{w}}{||w||}=\frac{2}{||w||} \quad (7)$$

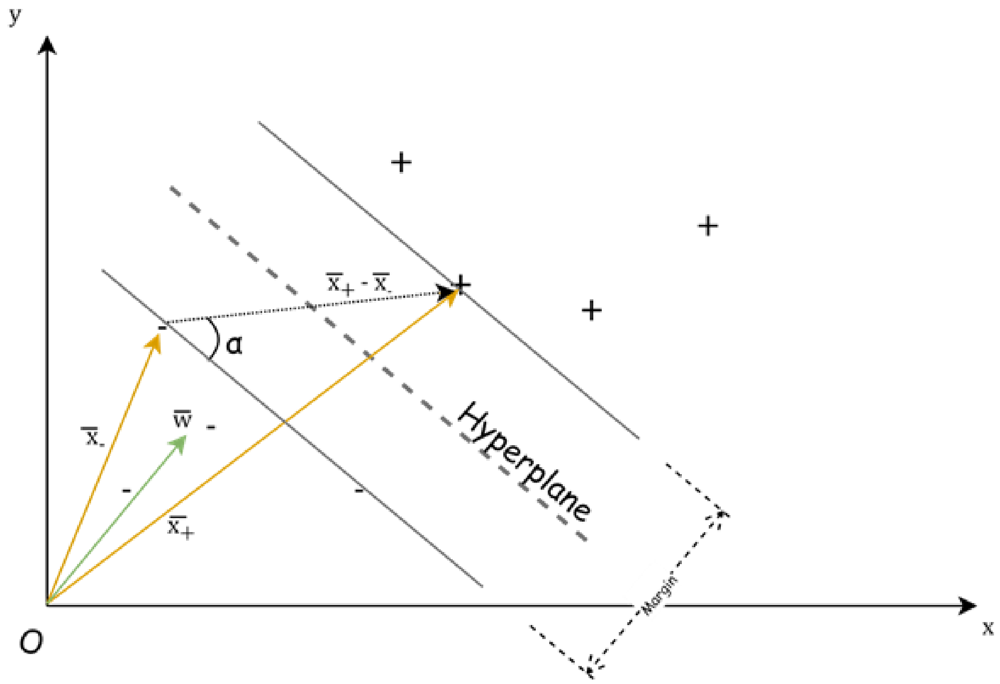
To maximize margin, we have to minimize ||w||. For further convenience, we can maximize (7) by finding minimum of $\frac{1}{2}||w||^2$ (8)

Minimize (8) subject to (6) with Lagrange Multiplier Expression we get:

$$\overline{w}=\sum_i \alpha_i y_i \overline{x_i} \quad \& \quad \sum_i \alpha_i y_i = 0$$

This result shows us that vector w is a linear sum of some (or all) vector x. It means we can find vector w using information that given from known observations x.

Known for its high accuracy and good performance, SVMs is a widely used



<Figure 2> Understanding SVMs classification algorithm by example

classification tool in text classification (Joachims, 1998), facial expression recognition (Michel, 2003), and health industry (Leslie, 2002), object detection (Felzenszwalb et al., 2010), predicting freshmen student attrition (Delen, 2010), determining whether a food can be consumed (Kim et al., 2017) etc. In economic industry, Hung et al. (2006) applied data mining to manage churn in mobile telecom, Farquad et al. (2009) extracted rules from SVMs to apply to churn prediction in bank credit cards, Cui and Curry (2005) used SVMs to make marketing prediction, Coussement and Van den Poel (2008) used SVMs approach in detecting customer churn in subscription services. Min and Lee (2005) used SVMs and Principal Component Analysis (PCA) to predict bankruptcy of Korean largest credit guarantee organization, etc.

2.3 Other binary classification methods

2.3.1 Logistic Regression (LR)

In the logistic regression approach, the relationship between the final decision and the independent variables is expressed as:

$$p = \frac{1}{e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j)}} \qquad (9)$$

This is called the logistic response function where p is denoted as probability of acceptance of the j^{th} item, x_j is vector of descriptors measured for the j^{th} item, β_j is vector of coefficients and β_0 acts as an intercept term. Then, instead of calculating the probability, we look at a different measurement of chance that particular observation belongs to specific class. This chance is called odds. The odds of belonging to class 1 is defined as the ratio of the probability of belonging to class 1 to the probability of belonging to class 0:

$$odds = \frac{p}{1-p} \quad (10)$$

Substituting (9) into (10) we have:

$odds = e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j}$ Or

$\ln(odds) = \beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j$

This is called logit and takes value from -∞ to ∞, thus, our final formulation of the relation between the response and the predictors uses the logit as the dependent variable and models it as a linear function of the j predictors (Shmueli, 2008).

LR was widely applied on businesses to predict business failure (Hui and Sun, 2011), to predict small business performance (Wood, 2006), to modeling small-business credit score (Bensic et al., 2005), etc. Known for simplicity and fast computation speed, LR is suitable for analyzing high-dimensions dataset and for statistical modeling with binary outcome (Tu, 1996). It is also less prone to over-fitting problem due to low variance.

2.3.2 Discriminant Analysis (DA)

DA is a statistical technique used to classify an observation into one of several priori groupings dependent upon the observation's individual characteristics (Altman, 1968). The original dichotomous discriminant analysis was developed by Ronald Fisher in 1936 as the basis for improved separation of observations into classes. The purpose of discriminant analysis is to find the linear combination of ratios which best discriminates between the groups which are being classified.

DA was used in many other research areas such as predicting bankruptcy (Altman, 1968), predicting business failure (Edward, 1972) or being applied on face recognition (Wenyi et al., 1998), etc. The strength of DA is ability of predicting outcome with small number of observations and large predictors. In other hand, it also has some weaknesses: (i) DA requires objects in various classes is balanced; (ii) DA is sensitive to over-fitting problem; (iii) DA is only applicable to linear classification problems.

2.3.3 Random Forest

The first idea of RF was developed by Ho (1995) using ensemble learning, a method of using multiple learning algorithms to obtain better predictive performance than any of the constituent learning algorithms alone. In 2006, Breiman and Cutler made their own Random Forest trade mark. They developed an extension of RF by combining “bagging” idea, which was suggested by Breiman (1996), and random selection of features Breiman (2001), which was introduced by Ho, in order to construct a collection of decorrelated decision trees with controlled variance.

The power of RF is that not does not require input variables to be handled before using: RF can work well with binary input, categorical input, or numerical input without any scaling requirement. Moreover, RF is versatility, simplicity and widely supported. We can use RF for regression tasks, classification or prediction task, and even clustering missions. As all other techniques, RF has some drawback: First, the RF models are very hard to interpret; And second, RF models might take large size of memory and quite slow to evaluate. Even so, RF is still a favorite algorithm of many data scientists.

2.3.4 Random Forest – Recursive Feature Elimination (RF-RFE)

Not only Artificial Intelligence/Machine Learning but also predictive analytics science, features selection is a one of most important parts. This task refers to the process of identifying a subset of most important variables among whole set of dataset variables. This subset must have predicting outcome ability, maybe not as good as using every variable but need to be good enough to use and to run faster. There are three objectives of feature selection task: (i) improving the prediction performance of the predictors, (ii) providing faster and more cost-effective predictors, and (iii) providing a better understanding of the underlying process that generated the data (Guyon and Elisseeff, 2003).

In this study, we used Recursive Feature Elimination (RFE) algorithms to extract predictors going to be used for constructing model. The idea of RFE is repeatedly construct a sub model, with an algorithm, and choose the best or worst performing predictor. Then the RFE algorithm sets this predictor aside and repeats the process with the rest of the features. The looping process will be applied until all predictors in the dataset are examined. The moment that the features eliminated will be used to rank them. Since the performance of RFE depends heavily on the type of algorithm that is used for predictors ranking, we decided to choose Random Forest (RF) algorithm to rank predictors. The combination of RF and RFE has been proved by Granitto et al. (2006). According to Granitto, RF has more efficient performance with small features set when comparing to SVMs algorithms, and with high dimension dataset, RF also generates similar ranking performance. Louw & Steel (2006) used RFE to extract Kernel Fisher Discriminant Analysis (KFDA) algorithm and improved that the performance of KFDA can be improved markedly by performing variable selection by means of recursive feature elimination. Other researchers, Johannes et al.

introduced a new algorithm called Reweighted Recursive Feature Elimination (RRFE) based on RFE theory and applied on risk stratification of cancer patients (2010); Hu et al. (2013) compared PCA with RFE and so on. These stuffs shows that RFE should be a optimized algorithm for finding the best performing subset of predictors.

2.4 Customer Relationship Management (CRM)

In the last several years CRM has grown to be one of the major trends in marketing, both in academia and in practice. Kumar (2010) addressed that CRM extends the concept of database marketing to a customer level in order to develop profitable company-to-customer relationships. Hosseini et al. (2010) assessed the customer loyalty using data mining approach and RFE model on CRM. Goodhue et al. (2002) used six company experiences in particular to illustrate three CRM target that companies aim for to get business benefits. These papers shows that CRM is a good resource for any business manager to dig and understand their customers.

CRM is not just technologies, it also refers to strategies, practices that organizations use to manage and interact with customers throughout customer lifecycle. Teo et al. (2006) and He et al. (2004) classified CRM framework into operational and analytical based on architecture point of view. According to these researchers, operational CRM refers to automation of business processes while analytical CRM refers to the analysis of customer characteristics and behaviors.

In retailing industry, CRM has more power. A good connection between retailers and customers is key to make customers visit store, brand loyalty and increase sales conversions. An accurate retailing CRM database can entice customers come to store, give them reason to visit, stay and purchase. By applying machine learning on retailing CRM data, we are going to analyze customer data within the analytical CRM framework to find valuable information hidden in the data collected by a Korean retailing firm.

Ⅲ. Research Design and Analysis

3.1 Dataset

For the purpose of this study, dataset from a Korean retailing firm was used. It is a small part of whole CRM database, stored one year data. The CRM database records every customers’ information and activities, including gender, age, purchase habit, type of purchase goods, or even discount sensitivity. It contains 2000 customer’s records with 40 fields. The input dataset was sorted by first purchase time, so in oder to ensure subjectivity, we randomly merger all records before using it. The predictors used in constructing models are shown in <Table 2>.

In this dataset, the field “CUST_GRADE” indicates how is customers loyal to company brand, it grades customers into “One time”, “Normal”, “Great”, “Excellent”, “VIP”, “Affiliate Employee”, and “Employee”. Within this study, we picked up only “VIP” level to mark as loyal customers. The reason why did we classified like this is that in marketing, loyal customers represent no more than 20 percent of customer base, but often make up more than 50 percent of company sales. 17.1 percent of “VIP” customer in this dataset is most appropriate to mark as loyal customers.

3.2 Research Procedure

This study carefully follows data mining procedure introduced by Shmueli (2008) that is shown in <Figure 3>. The main stages are described as below:

Step 1: Data Preprocessing

The collected raw data was messy and needed to be cleaned. The missing value cells were checked, classified and treated differently (for missing categorical cell, whole records will be deleted; for missing numerical cells, missing value will be replaced by column's mean value). Then, the columns were narrowed down by observing box-plot and frequency table visualization method. Since some algorithms require numerical input variables, we also converted categorical variables into numerical variables. After cleaning, from 2000 records with 40 fields, the dataset has been reduced to 1822 records with 27 fields and ready to be used.

Step 2: Variables selection

This step’s goal is finding a subset of the original variables set to include in model construction. Most of previous studies use Principal Component Analysis (PCA) (Min and Lee, 2005), Linear Discriminant Analysis (LDA) (Song et al., 2010), or Generalized Discriminant Analysis (GDA) (Stuhlsatz et al., 2012) etc, to extract the significant features. In this study, we decided to use RFE algorithms due to its excellent performance that was shown in previous studies.

Cleaned data then was processed with RF-RFE algorithm: It ranks features in most significant to less significant order, we selected 7 predictors at the top of ranking table to construct our model. The reason is that we had tried with other numbers of predictors and found out that a set of 7 predictors is the best choice.

The Features Ranking Table in <Table 3> shows the rank of each input variables that was scored by RF-RFE algorithm.

Step 3: Model Fitting.

<Table 2> Seven most important variables

Variable name	Description	Summary
TOP	Customer rating based on purchased amount	0.00696, 0.00938~9.99803
M	Total purchased money from beginning of customer	1886340, 1886650 ~ 69155120
REV_PER_VISIT	The average amount of money purchased per visit of each customer.	22656, 26690 ~ 3149350
MOST_EXP_GOOD_PREF_YN	Determine whether this customer prefers expensive goods or not.	Expensive good prefer (13.8%), Expensive goods do not prefer (86.2%)
API	The average number of purchases that customer makes per year (ex: 64.6 means this customer makes ~65 purchases per year in average)	1,.61 2.34 ~ 217
MULTI_BRANCH_VISIT_YN	Determine whether customer also visits other branches or not.	One branch visit (77.3%), Multi-branch visit(19.9%), Unidentified (2.9%)
F	Number of visits per year of each customer.	1, 2~227
CUST_GRADE (outcome)	Customer loyalty level was graded by company	VIP (17.1%), Affiliate employee (1%), Employee (0.8%), Great (39.4%), Normal (1.1%), One times (0.3%), Excellent (40.35%)

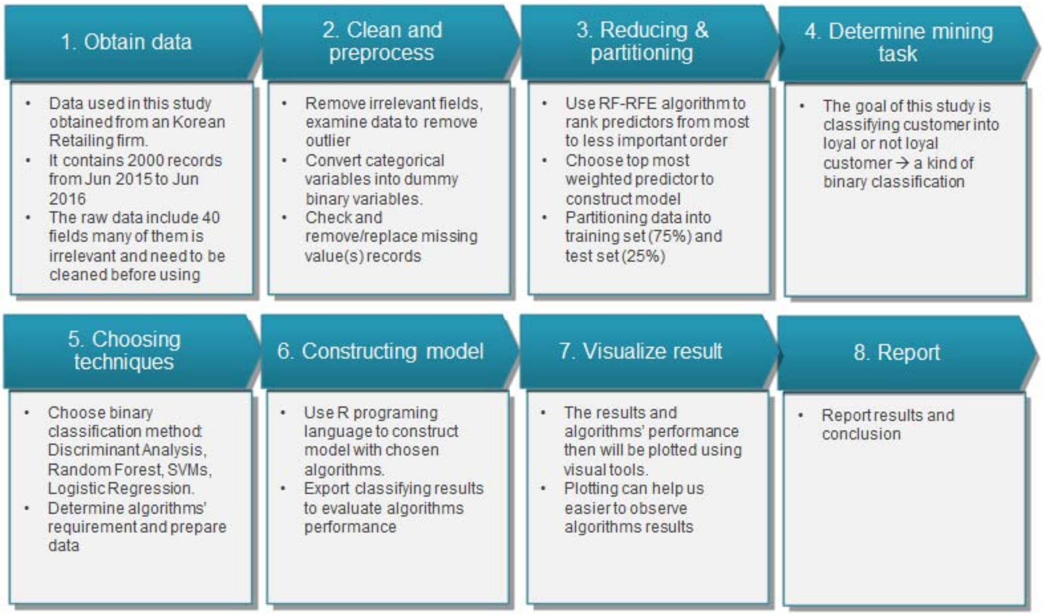
Using features selected in step 2, we constructed SVMs model with R programing code via e1071 package. Other machine learning classification algorithms which are bundled into other packages, are also used to compare with SVMs' performance.

Step 4: Reporting results

In this step, we extracted confusion matrix of each algorithm in step 3 and compare them. As we expected, the SVMs model results better classifying performance than others. With 455 records in test dataset, SVMs has successfully classified 435 and only 20 records were misclassified, equivalent to 95,6% success rate. The summary of confusion matrices and algorithms' performance is shown in <Table 4>.

3.3 Result Analysis

Before fitting models, we partitioned the dataset into training set (75%, 1365 observations) and test set (25%, 455 observations).



<Figure 3> Detailed research procedure

Nevertheless, SVMs model can be trained with a lot of various parameters, so it requires a method of determining best parameters to use. There are some vary method to do this task, in this study, we utilized k-fold cross-validation method, according to practical guideline to SVMs using grid-search and cross-validation which was suggested by Hsu et al. (2004). Fortunately, there are many packages in R can be used to perform cross-validation, we can do this with both “caret” or “e1071” packages, it was fast and easy to perform. We chose the “tune” function that was packed into “e1071” package with 10-fold cross-validation to determine the best optimal parameters. We also used RBF kernel to construct SVMs model. Among 3 SVMs non-linear kernels, RBF kernel maps the observations to higher dimension space, with this technique, always exist a hyperplane to differentiate samples. Sigmoid operates similar to RBF but for some certain parameters , polynomial kernel takes longer time to train model and requires more hyper-parameters than RBF (Min & Lee, 2005). Thus, we used RBF kernel as default to construct SVMs model. Using these parameters, the SVMs model has successfully classified 435/455 observations when applied to test dataset, equivalent to 95.6% accuracy.

In LR, because LR algorithm use logit function to classify observations, and logit function is a continuos value, so we needed to indicate the cut-off value. In this case, we use default cut-off value of 0.5, because with our point of view, loyal customer class is as important as normal customer class. For training set, among 1365 observations, LR algorithm has successfully classified 1243 and misclassified 122 observations (91% success). Tested with test dataset, LR model provided even more better performance: among 455 observations, 417 records was successfully classified, equivalent to 91.6% success.

<Table 3> RF-RFE Features Ranking

Ranking	Variables	RMSE
1	TOP	0.2963
2	M	0.2942
3	REV_PER_VISIT	0.272
4	MOST_EXP_GOOD_PREF_YN	0.266
5	API	0.2638
6	MULTI_BRANCH_VISIT_YN	0.2643
7	F	0.2612
8	TOP_CLASS	0.2576
9	AGE_CODE	0.2575
10	BIZ_AREA	0.2569
11	BRANCH_CODE	0.2565
12	HOME_PRODUCT	0.2568
13	NUM_ITEM_CATEGORY	0.2553
14	NORM_GOOD_PERF_TYPE_YN	0.2541
15	PURCHASE_RULE_YN	0.2538
16	COSMETIC	0.253
17	FOOD	0.2524
18	LPL	0.2526
19	GOODS	0.2524
20	KIDS	0.2508
21	WOMEN_WEAR	0.2505
22	MEN_WEAR	0.251
23	HOME_APPLIANCE	0.2519
24	ONLINE_MEMBER_YN	0.251
25	GENDER	0.2505
26	FURNITURE	0.2503

In case of DA model, 1235/1365 training set's observations was successfully classified (90.5% success). The success rate on test dataset is similar to training set: 90.36% compare to 90.5%.

In previous researches, RF algorithm often acts as good performance machine learning algorithm even if the training set has not many labelled observations, but in this study, the RF model has less accuracy than any other classification method. The collected result shown in <Table 4> indicates that Machine Learning algorithms have successfully classified retailing customers into loyal and normal groups. In this stage, comparisons was made between performance of SVMs and other algorithms. The evaluation was performed by comparing error rate in confusion matrices (that shown in <Table 4>). Within this dataset, at least ~90% customer records were exactly classified by algorithms and SVMs performed better than others in classifying loyal customers (95.6%). This study lead us to three following understanding: First, similar to other areas, machine learning and SVMs can be totally applied on Korean Retailing Industry to predict customers' behaviors. Second, in retailing industry, customer loyalty is multi-dimensional factor that is impacted by many purchasing information. We addressed that loyal retailing customer often prefers expensive goods, come more frequently and pay more amount of money to purchase. Third, as shown in <Figure 4> SVMs seems to be better than other classification algorithms in classifying loyal customers of Korean retailing firm with CRM database.

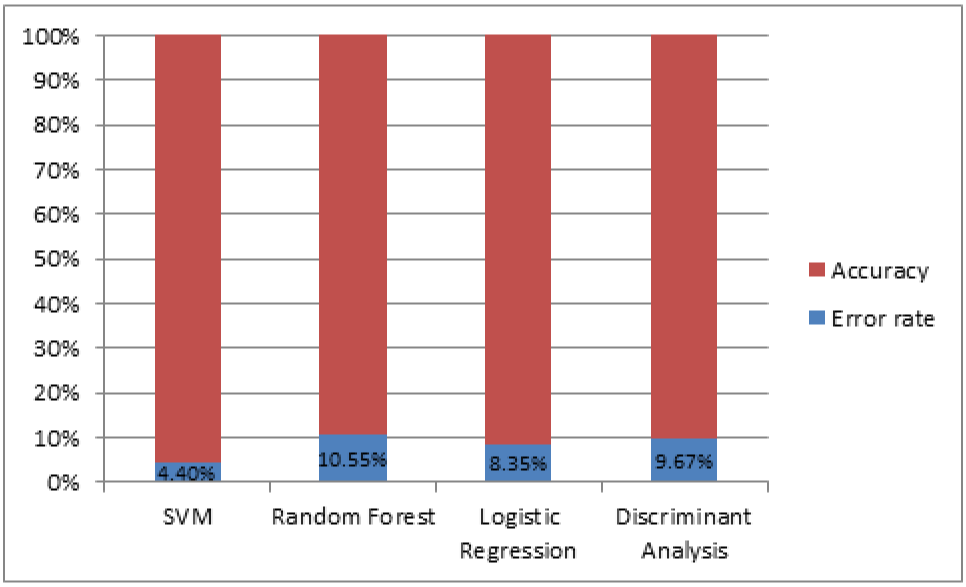
Ⅳ. Conclusion

4.1 Summary and Implications

In Korea, most being used methods to classifying customer are RFM (Recency, Frequency and Monetary) and NPS (Net Promoter Score). While, as discussed earlier, NPS is one dimension measurement and unreliable, RFM also has limitations and risks: (i) RFM scoring puts specific business, product and marketing events aside. It cannot measure the short-term customers behavior. (ii) RFM bypass lower-scoring customers instead of nurturing them. (iii) RFM model is too granular. Our effort contributes to customer loyalty management and classification in the following three ways: First, the proposed model integrates multi dimensions of CRM data. By using multi dimensions of CRM data when assessing customer loyalty, we indirectly proved the unreliability of NPS, which uses just one dimension to assess customer loyalty; We also could classify and measure a complex quantitative factor like customer loyalty more precisely. Second, we suggested a new approach of classifying customer based on loyalty. Data mining is favorite technique which has been proved appropriate to apply to customer behavior classification and prediction. Third, by comparing SVMs model to other machine learning classification method, we suggested a better method for classifying customer based on loyalty using CRM data, which contains many thousands data records. Using this study, researchers could build their own model on other domains (finance, health industries…) or other

<Table 4> Confusion matrices

Algorithms	Predicted	Actual		Performance
		Not loyal	Loyal	
SVMs	Not loyal	371	20	95.60%
	Loyal	0	64	
Random Forest	Not loyal	354	31	89.45%
	Loyal	17	53	
Logistical Regression	Not loyal	361	10	91.65%
	Loyal	28	56	
Discriminant Analysis	Not loyal	359	35	90.33%
	Loyal	9	52	



<Figure 4> Algorithms' performance chart

developing platforms (Python, RapidMiner, XLMiner, ect.). Further researchers also can have an overview about the performance of different classification algorithms for selecting a suitable one to apply on their studies.

4.2 Limitations and Future's Direction

Although successfully classified retailing customers based on loyalty, this model still has some limitations: (i) Loyalty need to be continuously developed by keeping in touch with customer. Ball et al. (2004) described the role of communication in explaining customer loyalty. According to that, the effect of communication on loyalty is important and interesting even more than other factors. The models in this study could not measure the effect of communication on customer loyalty. Developing a tool for measuring business-ustomer communicating effect before applying ML will be our next interesting research topic. (ii) Since data scheme collected by each CRM software is different to others, it is difficult to apply the data cleaning step on CRM data collected by other CRM software. To deal with this limitation, we need a stage, in which users input their CRM data and map their data's fields with our system's built in fields before constructing model.

References

Altman, E. I., "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, 1968, pp. 589-609.

Ball, D., Coelho, P. S., and Machas, A., "The role of communication and trust in explaining customer loyalty: An extension to the ECSI model," *European Journal of Marketing*, Vol. 38, No. 9/10, 2004, pp. 1272-1293.

Bensic, M., Sarlija, N., and Zekic Susac, M., "Modelling small business credit scoring by using logistic regression, neural networks and decision trees," *Intelligent Systems in Accounting, Finance and Management*, Vol. 13, No. 3, 2005, pp. 133-150.

Blum, A., and Mitchell, T., "Combining labeled and unlabeled data with co-training," *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, ACM, 1998.

Breiman, L., "Bagging predictors," *Machine Learning*, Vol. 24, No. 2, 1996, pp. 123-140.

Breiman, L., "Random forests," *Machine Learning*, Vol. 45, No. 1, 2001, pp. 5-32.

Coussement, K., and Van den Poel, D., "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Systems with Applications*, Vol. 34, No. 1, 2008, pp. 313-327.

Cui, D., and Curry, D., "Prediction in marketing using the support vector machine," *Marketing Science*, Vol. 24, No. 4, 2005, pp. 595-615.

Deakin, E. B., "A discriminant analysis of predictors of business failure," *Journal of Accounting Research*, Vol. 10, 1972, pp. 167-179.

Delen, D., "A comparative analysis of machine learning techniques for student retention management," *Decision Support Systems*, Vol. 49, No. 4, 2010, pp. 498-506.

Dudyala, A. K. and Ravi, V., "Predicting credit card customer churn in banks using data mining," *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 1, 2008, pp 4-28.

Farquad, M. A. H., Ravi, V., and Raju, S. B., "Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards," *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Vol. 5908, 2009, pp. 390-397.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D., "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, 2010, pp. 1627-1645.

Fisher, R. A., "The use of multiple measurements in taxonomic problems," *Annals of Human Genetics*, Vol. 7, No.

2, 1936, pp. 179-188.

Goodhue, D. L., Wixom, B. H., and Watson, H. J., "Realizing business benefits through CRM: hitting the right target in the right way." *MIS Quarterly Executive*, Vol. 1, No. 2, 2002, pp. 79-94.

Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F., "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems*, Vol. 83, No. 2, 2006, pp. 83-90.

Gremler, D. D., and Brown, S. W., "Service loyalty: its nature, importance, and implications," *Advancing Service Quality: A Global Perspective*, Vol. 5, 1996, pp. 171-181.

Guyon, I., and Elisseeff, A., "An introduction to variable and feature selection," *Journal of Machine Learning Research*, Vol. 3, 2003, pp. 1157-1182.

He, Z., Xu, X., Huang, J. Z., and Deng, S., "Mining class outliers: concepts, algorithms and applications in CRM," *Expert Systems with Applications*, Vol. 27, No. 4, 2004, pp. 681-697.

Ho, T. K., "Random decision forests," *Proceedings of the Third International Conference on*, 1995, pp. 278-282.

Hosseini, S. M. S., Maleki, A., and Gholamian, M. R., "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Systems with Applications*, Vol. 37, No. 7, 2010, pp. 5259-5264.

Hsu, C. W., Chang, C. C., and Lin, C. J., "*A practical guide to support vector classification*," 2003.

Hu, C., Wang, J., Zheng, C., Xu, S., Zhang, H., Liang, Y., Bi, L., Fan, Z., Han, B., and Xu, W., "Raman spectra exploring breast tissues: Comparison of principal component analysis and support vector machine recursive feature elimination," *Medical Physics,* Vol. 40, No. 6, 2013, pp. 063501.

Hung, S. Y., Yen, D. C., and Wang, H. Y., "Applying data mining to telecom churn management," *Expert Systems with Applications,* Vol. 31, No. 3, 2006, pp. 515-524.

Joachims, T., "Text categorization with support vector machines: Learning with many relevant features," *Machine Learning: ECML-98*, 1998, pp. 137-142.

Johannes, M., Brase, J. C., Fröhlich, H., Gade, S., Gehrmann, M., Fälth, M., Sültmann, H., and Beißbarth, T., "Integration of pathway knowledge into a reweighted recursive feature elimination approach for risk stratification of cancer patients," *Bioinformatics,* Vol. 26, No. 17, 2010, pp. 2136-2144.

Keiningham, T. L., Cooil, B., Aksoy, L., Andreassen, T. W., and Weiner, J., "The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet," *Managing Service Quality: An International Journal*, Vol. 17, No. 4, 2007, pp. 361-384.

Kim, S. A., Kim, J. W., Won, D. Y., and Choi, Y. R., "A Halal Food Classification Framework Using Machine Learning Method for Enhancing Muslim Tourists," *The Journal of Information Systems*, Vol. 26, No. 3, 2017, pp. 273-293.

Kotler, P., and Armstrong, G., *Principles of Marketing*. Pearson education, 2010.

Kumar, V., *Customer Relationship Management.* John Wiley & Sons, Ltd, 2010.

Lee, M. H., "Loyalty of On-line Stock Trading Customers," *The Journal of Information Systems*, Vol. 14, No. 2, 2005, pp. 155-172.

Leslie, C., Eskin, E., and Noble, W. S., "The spectrum kernel: A string kernel for SVM protein classification," *Pacific Symposium on Biocomputing*, Vol. 7, 2002, pp. 566-575

Li, H., and Sun, J., "Empirical research of hybridizing principal component analysis with multivariate discriminant analysis and logistic regression for business failure prediction," *Expert Systems with Applications*, Vol. 38, No. 5, 2011, pp. 6244-6253.

Lin, H. H., and Wang, Y. S., "An examination of the determinants of customer loyalty in mobile commerce contexts," *Information & Management*, Vol. 43, No. 3, 2006, pp. 271-282.

Louw, N., and Steel, S. J., "Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination," *Computational Statistics & Data Analysis*, Vol. 51, No. 3, 2006, pp. 2043-2055.

Michel, P., and El Kaliouby, R., "Real time facial expression recognition in video using support vector machines," *Proceedings of the 5th International Conference on Multimodal Interfaces*. ACM, 2003, pp. 258-264

Min, J. H., and Lee, Y. C., "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert Systems with Applications*, Vol. 28, No. 4, 2005, pp. 603-614.

Samuel, A. L. "Some studies in machine learning using the game of checkers," *IBM Journal of Research and Development*, Vol. 3, No. 3, 1959, pp. 210-229.

Shmueli, G., Patel, N. R., and Bruce, P. C., *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*. John Wiley & Sons, 2008.

So, S. H., Ryu, I., Cho, G., and Park, Y. S., "Structural Relationships of Logistics Service Quality, Relationship Orientation, Customer Satisfaction and Customer Loyalty in Electronic

Commerce," *The Journal of Information Systems*, Vol. 16, No. 4, 2007, pp. 107-129.

Song, F., Mei, D., and Li, H., "Feature selection based on linear discriminant analysis," *Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on*, Vol. 1, 2010, pp. 746-749.

Stuhlsatz, A., Lippel, J., and Zielke, T., "Feature extraction with deep neural networks by a generalized discriminant analysis," *IEEE transactions on neural networks and learning systems*, Vol. 23, No. 4, 2012, pp. 596-608.

Teo, T. S., Devadoss, P., and Pan, S. L., "Towards a holistic perspective of customer relationship management (CRM) implementation: A case study of the Housing and Development Board, Singapore," *Decision Support Systems*, Vol. 42, No. 3, 2006, pp. 1613-1627.

Tu, J. V., "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, Vol. 49, No. 11, 1996, pp. 1225-1231.

Wood, E. H., "The internal predictors of business performance in small firms: A logistic regression analysis," *Journal of Small Business and Enterprise Development*, Vol. 13, No. 3, 2006, pp. 441-453.

Zaki, M., Kandeil, D., Neely, A., and McColl-Kennedy, J. R., *The Fallacy of the Net Promoter Score: Customer Loyalty Predictive Model*, Cambridge Service Alliance, University of Cambridge, 2016.

Zhao, W., Chellappa, R., and Krishnaswamy, A., "Discriminant analysis of principal components for face recognition," *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, IEEE, 1998.

응위엔푸티엔 (Nguyen, Phu-Thien)

동국대학교 일반대학원 경영학 석사학위를 취득하였다. 주요 관심분야는 Business Analytics, 인공지능, 전자상거래, 핀테크 등이다.

이 영 찬 (Lee, Young-Chan)

서강대학교 경영학사, 동 대학원에서 경영학 석사 및 박사학위를 취득하였다. 현재 동국대학교 경주캠퍼스 경영학부 교수로 재직하고 있으며, Annals of Management Science, The Open Operational Research Journal의 Editorial Board, 한국경영학회 MIS 분야 대표 편집위원으로 활동 중이다. 주요 관심 분야는 기업성과측정, 데이터마이닝, 복잡계 이론, 다기준의사결정 등이다.

<Abstract>

An Application of Support Vector Machines to Customer Loyalty Classification of Korean Retailing Company Using R Language

Nguyen, Phu-Thien · Lee, Young-Chan

Purpose

Customer Loyalty is the most important factor of customer relationship management (CRM). Especially in retailing industry, where customers have many options of where to spend their money. Classifying loyal customers through customers' data can help retailing companies build more efficient marketing strategies and gain competitive advantages. This study aims to construct classification models of distinguishing the loyal customers within a Korean retailing company using data mining techniques with R language.

Design/methodology/approach

In order to classify retailing customers, we used combination of support vector machines (SVMs) and other classification algorithms of machine learning (ML) with the support of recursive feature elimination (RFE). In particular, we first clean the dataset to remove outlier and impute the missing value. Then we used a RFE framework for electing most significant predictors. Finally, we construct models with classification algorithms, tune the best parameters and compare the performances among them.

Findings

The results reveal that ML classification techniques can work well with CRM data in Korean retailing industry. Moreover, customer loyalty is impacted by not only unique factor such as net promoter score but also other purchase habits such as expensive goods preferring or multi-branch visiting and so on. We also prove that with retailing customer's dataset the model constructed by SVMs algorithm has given better performance than others. We expect that the models in this study can be used by other retailing companies to classify their customers, then they can focus on giving services to these potential vip group. We also hope that the results of this ML algorithm using

R language could be useful to other researchers for selecting appropriate ML algorithms.

Keywords: Support Vector Machines, SVMs, Customer Relationship Management, CRM, Recursive Feature Elimination, RFE, Random Forest, RF, R Language, Loyalty, Korean Retailing, Customer Classification

* 이 논문은 2017년 11월 13일 접수, 2017년 11월 28일 1차 심사, 2017년 12월 18일 게재 확정되었습니다.