Knowledge Discovery in a Direct Marketing Case using Least Squares Support Vector Machines

S. Viaene,[1,*] B. Baesens,[1] T. Van Gestel,[2] J. A. K. Suykens,[2]
D. Van den Poel,[3] J. Vanthienen,[1] B. De Moor,[2] G. Dedene[1]
[1] *K.U. Leuven, Department of Applied Economic Sciences, Naamsestraat 69, B-3000 Leuven, Belgium*
[2] *K.U. Leuven, Department of Electrical Engineering ESAT-SISTA, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[3] *Ghent University, Department of Marketing, Hoveniersberg 24, B-9000 Ghent, Belgium*

We study the problem of repeat-purchase modeling in a direct marketing setting using Belgian data. More specifically, we investigate the detection and qualification of the most relevant explanatory variables for predicting purchase incidence. The analysis is based on a wrapped form of input selection using a sensitivity based pruning heuristic to guide a greedy, stepwise, and backward traversal of the input space. For this purpose, we make use of a powerful and promising least squares support vector machine (LS-SVM) classifier formulation. This study extends beyond the standard recency frequency monetary (RFM) modeling semantics in two ways: (1) by including alternative operationalizations of the RFM variables, and (2) by adding several other (non-RFM) predictors. Results indicate that elimination of redundant/irrelevant inputs allows significant reduction of model complexity. The empirical findings also highlight the importance of frequency and monetary variables, while the recency variable category seems to be of somewhat lesser importance to the case at hand. Results also point to the added value of including non-RFM variables for improving customer profiling. More specifically, customer/company interaction, measured using indicators of information requests and complaints, and merchandise returns provide additional predictive power to purchase incidence modeling for database marketing.

* Author to whom correspondence should be addressed; e-mail: Stijn.Viaene@econ.kuleuven.ac.be.

Contract grant sponsor: Flemish government (Research Council K.U. Leuven: GOA-Mefisto; FWO-Flanders: ICCoS and AN-MMM; IWT: STWW Eureka).
Contract grant sponsor: Belgium government.
Contract grant number: IUAP-IV/02, IV-24.
Contract grant sponsor: European Commission.
Contract grant number: TMR, ERNSI.

I. INTRODUCTION

The main objective of this paper involves the detection and qualification of the most relevant variables for repeat-purchase modeling in a direct marketing setting. This knowledge is believed to vastly enrich customer profiling and thus contribute directly to more targeted customer contact.

The empirical study focuses on the purchase incidence, i.e., the issue whether or not a purchase is made from any product category offered by the direct mailing company. Standard recency frequency monetary modeling semantics underlie the discussed purchase incidence model.[1] This binary (buyer versus nonbuyer) classification problem is being tackled in this paper by using least squares support vector machine (LS-SVM) classifiers. LS-SVMs have recently been introduced in the literature[2] and excellent benchmark results have been reported.[3] Having constructed an LS-SVM classifier with all available predictors, we engage in an input selection experiment. Input selection has been an active area of research in the datamining field for many years now. A compact, yet highly accurate model may come in very handy in (on-line) customer profiling systems. Furthermore, by reducing the number of input features, both human understanding and computational performance can often be vastly enhanced.

Section II elaborates on some response modeling issues including a literature review and description of the data set. In Section III, we discuss the basic underpinnings of LS-SVMs for binary classification. The input selection experiment and corresponding results are presented and discussed in Section IV.

II. THE RESPONSE MODELING CASE FOR DIRECT MARKETING

A. Response Modeling in Direct Marketing

Cullinan is generally credited for identifying the three sets of variables most often used in response modeling: recency, frequency, and monetary (RFM).[1,4,5] Since then, the literature has provided so many uses of these three variable categories, that there is overwhelming evidence both from academically reviewed studies as well as from practitioners' experience that the RFM variables are an important set of predictors for modeling mail-order repeat purchasing. However, the beneficial effect of including other variables into the response model has also been investigated.

For mail-order response modeling, several alternative problem formulations have been proposed based on the choice of the dependent variable. The first category is purchase incidence modeling.[6] In this problem formulation, the main question is whether a customer will purchase during the next mailing period, i.e., one tries to predict the purchase incidence within a fixed time interval (typically half a year). Other authors have investigated related problems dealing with both the purchase incidence and the amount of purchase in a joint model.[7,8] A third alternative perspective for response modeling is to model interpurchase time through survival analysis or (split-)hazard rate models which model whether a purchase takes place together with the duration of time until a purchase occurs.[9,10] Table I provides a summary of contributions with regard to the three alternative problem formulations. We observe that the (first) purchase incidence formulation is clearly the most popular in the existing literature.[11] Moreover, most studies include many predictors, even though only a minority includes all categories.

This paper focuses on the first type of problem, i.e., purchase incidence modeling. This choice is motivated by the fact that the majority of previous research in the direct marketing literature focuses on the purchase incidence problem.[12,13] Furthermore, this is exactly the setting that mail-order companies are typically confronted with. They have to decide whether or not a specific offering will be sent to a (potential) customer during a certain mailing period. Given a tendency of rising mailing costs and increasing competition, we can easily see an increasing importance for response modeling.[14] Improving the targeting of the offers may indeed counter these two challenges by lowering nonresponse. Moreover, from the perspective of the recipient of the (direct mail) messages, mail-order companies do not want to overload consumers with catalogs. The importance of response modeling to the mail-order industry is further illustrated by the fact that the issue of improving targeting was among the top three concerns with 73.5% of the catalogers in the sample mentioned in Ref. 15.

B. The Data Set

From a major Belgian mail-order company, we obtained data on past purchase behavior at the order-line level, i.e., we know when a customer purchased what quantity of a particular product at which price as part of what order. This allowed us, in close cooperation with domain experts and guided by the extensive literature, to derive all the necessary purchase behavior variables for a total sample size of 5,000 customers, of which 37.94% represent buyers. For each customer, these variables were measured for the period between 1 July 1993 and 30 June 1997. The goal is to predict whether an existing customer will repurchase in the observation period between 1 July 1997 and 31 December 1997 using the information provided by the purchase behavior variables. This problem boils down to a binary classification problem: Will a customer repurchase or not? Notice that the focus is on customer retention and not on customer acquisition.

The recency, frequency, and monetary variables have then been modeled as described in detail in Ref. 11. We used two time horizons for all RFM variables. The Hist horizon refers to the fact that the variable is measured for the period between 1 July 1993 and 30 June 1997. The Year horizon refers to the fact that the variable is measured over the last year. Including both time horizons allows us to check whether more recent data are more relevant than historical data. All RFM variables are modeled both with and without the occurrence of returned merchandise, indicated by R and N in the variable name, respectively. The former is operationalized by including the counts of returned merchandise in the variable values, whereas in the latter case these counts are omitted. Taking

Table I. Literature review of response modeling papers.

	Independent Variable					Dependent Variable			Context of Application		
Reference	R	F	M	Other Behavioral Variables	Socio-Demographic Variables	Binary	Binary & Amount	Binary & Timing	Fund-Raising	Catalog (general)	Catalog or Specialty Mailing
Berger and Magliozzi (1992)[16]	X	X	X		X	X				X	
Bitran and Mondschein (1996)[17]	X	X	X				X			X	
Bult and Wittink (1996)[18]	X		X		X		X		X		
Bult (1993)[19]			X	X	X	X					X
Bult et al. (1997)[20]	X	X	X	X	X	X			X		
Gönül and Shi (1998)[21]	X	X				X				X	
Kaslow (1997)[22]	X	X	X	X	X			X		X	
Levin and Zahavi (1998)[7]	X	X	X	X	X	X	X				X
Magliozzi and Berger (1993)[23]	X	X	X			X				X	
Magliozzi (1989)[24]	X	X	X			X					X
Thrasher (1991)[25]	X	X		X	X	X					X
Van den Poel & Leunis (1998)[10]	X	X		X	X			X			X
Van den Poel (1999)[11]	X	X	X	X	X	X				X	
Van der Scheer (1998)[8]			X		X		X		X		
Zahavi and Levin (1997)[13]	X		X	X	X	X					X

into account both time horizons (Year versus Hist) and inclusion versus exclusion of returned items (R versus N), we arrive at a 2×2 design in which each RFM variable is operationalized in four ways.

For the recency variable, many operationalizations have already been suggested. In this paper, we define the recency variable as the number of days since the last purchase within a specific time window (Hist versus Year) and in- or excluding returned merchandise (R versus N).[4] Recency has been found to be inversely related to the probability of the next purchase, i.e., the longer the time delay since the last purchase the lower the probability of a next purchase within the specific period.[1]

In the context of direct mail, it has generally been observed that multibuyers (buyers who already purchased several times) are more likely to repurchase than buyers who only purchased once.[4,26] Although no detailed results are reported because of the proprietary nature of most studies, the frequency variable is generally considered to be the most important of the RFM variables [12]. Bauer suggests to operationalize the frequency variable as the number of purchases divided by the time on the customer list since the first purchase.[4] We choose to operationalize the frequency variable as the number of purchases made in a certain time period (Hist versus Year) while in- or excluding returned merchandise (R versus N).

In the direct marketing literature, the general convention is that the more money a person has spent with a company, the higher his/her likelihood of purchasing the next offering.[27] Nash suggests to operationalize monetary value as the highest transaction sale or as the average order size.[12] Levin and Zahavi propose to use the average amount of money per purchase.[27] We model the monetary variable as the total accumulated monetary amount of spending by a customer during a certain time period (Hist versus Year) while in- or excluding returned merchandise (R versus N). Additionally, we include the natural logarithmic transformation (ln) of all monetary variables as a means to reduce the skewness of the distributions.

Apart from the RFM variables, we also included nine other customer profiling inputs.[11] The type and frequency of contact which customers have with the mail-order company may yield important information about their future purchasing behavior. The GenInfo and GenCust are binary customer/company interaction variables indicating whether the customer asked for general information (respectively, filed general complaints). Since customer (dis)satisfaction may not only be revealed by general complaints but also by returning items, we included two extra variables. The RetMerch variable is a binary variable indicating whether the customer has ever returned an item that was previously ordered from the mail-order company. The RetPerc variable measures the total monetary amount of returned orders divided by the total amount of spending. The Ndays variable models the length of the customer relationship in days. It is commonly believed that consumers/households with a longer relationship with the company have a higher probability of repurchase than households with shorter relationships. IncrHist and IncrYear are operationalizations of a behavioral loyalty measure. We propose to perform a median split of the length of the relationship (time since the household became a customer). This enables us to compare the number of purchases (i.e., frequency) between the first and last half of the time window. The following formula is used:

$$\frac{\text{purchases second half} - \text{purchases first half}}{\text{purchases first half}} \tag{1}$$

When the above measure is positive, this may give us an indication of increasing loyalty by that customer to the (mail-order) company, and *ipso facto* satisfaction with the current level of service. Remember that the suffix Hist reflects that the whole purchase history is used, whereas in the case of the suffix Year, only transactions from the last year are included. The ProdclaT respectively ProdclaM variables represent the total (T) [respectively, mean (M)] forward-looking weighted product index. The weighting procedure represents the "forward-looking" nature of a product category purchase, derived from another sample of data.

Table II gives an overview of the variables discussed above. Notice that all missing values were handled by the mean imputation procedure[28] and that all predictor variables were normalized to zero mean and unit variance prior to their inclusion in the model.[29]

III. LEAST SQUARES SVM CLASSIFICATION

A. LS-SVMs for Binary Classification

Given a training set $\{x_i, y_i\}_{i=1}^{N}$ with input data $x_i \in \mathbb{R}^n$ and corresponding binary class labels $y_i \in \{-1, +1\}$, the SVM classifier, according to Vapnik's original formulation,[30–33] satisfies the following conditions:

$$\begin{aligned} w^T\varphi(x_i) + b &\geq +1 \quad \text{if } y_i = +1 \\ w^T\varphi(x_i) + b &\leq -1 \quad \text{if } y_i = -1 \end{aligned} \tag{2}$$

Table II. A listing of all inputs (both RFM and non-RFM) included in the direct marketing case.

Recency	Frequency	Monetary	Other
RecYearR	FrYearR	MonHistR	ProdclaT
RecYearN	FrYearN	MonHistN	ProdclaM
RecHistR	FrHistR	MonYearR	GenCust
RecHistN	FrHistN	MonYearN	GenInfo
		ln(MonHistR)	Ndays
		ln(MonHistN)	IncrHist
		ln(MonYearR)	IncrYear
		ln(MonYearN)	RetMerch
			RetPerc

which is equivalent to:

$$y_i\left[w^T\varphi(x_i) + b\right] \geq 1 \qquad i = 1, \ldots, N \tag{3}$$

The nonlinear function $\varphi(\cdot)$: $\mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$ maps the input space to a high dimensional (and possibly infinite dimensional) feature space. In primal weight space the classifier then takes the form:

$$y(x) = \text{sign}\left[w^T\varphi(x) + b\right] \tag{4}$$

however, it is never evaluated in this form. One defines the optimization problem as:

$$\min_{w, b, \xi} \mathscr{T}(w, \xi) = \frac{1}{2}w^T w + c\sum_{i=1}^{N} \xi_i \tag{5}$$

subject to:

$$\begin{aligned} y_i\left[w^T\varphi(x_i) + b\right] \geq 1 - \xi_i \qquad i = 1, \ldots, N \\ \xi_i \geq 0 \qquad i = 1, \ldots, N \end{aligned} \tag{6}$$

The variables ξ_i are slack variables which are needed to allow misclassifications in the set of inequalities (e.g., due to overlapping distributions). The positive real constant c should be considered as a tuning parameter in the algorithm. For nonlinear SVMs, the QP-problem and the classifier are never solved and evaluated in this form. Instead, a dual space formulation and representation are obtained by applying the Mercer condition (see Refs. 30–33 for details).

Vapnik's SVM classifier formulation was modified by Suykens and Vandewalle[2] into the following LS-SVM formulation:

$$\min_{w, b, e} \mathscr{T}(w, e) = \frac{1}{2}w^T w + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2 \tag{7}$$

subject to the equality constraints:

$$y_i\left[w^T\varphi(x_i) + b\right] = 1 - e_i, \qquad i = 1, \ldots, N \tag{8}$$

This formulation now consists of equality instead of inequality constraints and takes into account a squared error with a regularization term similar to ridge regression. The solution is obtained after constructing the Lagrangian:

$$\mathscr{L}(w, b, e; \alpha) = \mathscr{T}(w, e) - \sum_{i=1}^{N} \alpha_i\left\{y_i\left[w^T\varphi(x_i) + b\right] - 1 + e_i\right\} \tag{9}$$

where α_i are the Lagrange multipliers. After taking the conditions for optimality, one obtains the following linear system[2]:

$$\left[\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega + \gamma^{-1}I \end{array}\right]\left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \mathbf{1} \end{array}\right] \tag{10}$$

where $Z = [\varphi(x_1)^T y_1; \ldots; \varphi(x_N)^T y_N]$, $Y = [y_1; \ldots; y_N]$, $\mathbf{1} = [1; \ldots; 1]$, $\alpha = [\alpha_1; \ldots; \alpha_N]$, $\Omega = ZZ^T$, and Mercer's condition[2] is applied within the Ω matrix:

$$\begin{aligned} \Omega_{ij} &= y_i y_j \varphi(x_i)^T \varphi(x_j) \\ &= y_i y_j K(x_i, x_j) \end{aligned} \tag{11}$$

For the kernel function $K(\cdot, \cdot)$ one typically has the following choices:

$K(x, x_i) = x_i^T x,$ (linear kernel)

$K(x, x_i) = \left(x_i^T x + 1\right)^d,$ (polynomial kernel of degree d)

$K(x, x_i) = \exp\left\{-\|x - x_i\|_2^2 / \sigma^2\right\},$ (radial basis function (RFB) kernel)

$K(x, x_i) = \tanh\left(\kappa x_i^T x + \theta\right),$ (multilayer perceptron (MLP) kernel)

where d, σ, κ, and θ are constants. Notice that the Mercer condition holds for all $\sigma \in \mathbb{R}^+$ and $d \in \mathbb{N}$ values in the RBF and the polynomial cases, but not for all possible choices of κ and θ in the MLP case. The LS-SVM classifier is then constructed as follows:

$$y(x) = \operatorname{sign}\left[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right] \tag{12}$$

Note that the matrix in (10) is of dimension $(N + 1) \times (N + 1)$. For large values of N, this matrix cannot easily be stored, such that an iterative solution method for solving it is needed. A Hestenes–Stiefel conjugate gradient algorithm is suggested in Ref. 34 to overcome this problem. Basically, the latter rests upon a transformation of the matrix in (10) to a positive definite form.[34] A straightforward extension of LS-SVMs to multiclass problems has been proposed in Ref. 35, where additional outputs are taken to encode multiclasses as is often done in classical neural network methodology.[29] A drawback of LS-SVMs is that sparseness is lost due to the choice of a 2-norm. However, this can be circumvented in a second stage by a pruning procedure which is based upon removing training points guided by the sorted support value spectrum.[36]

B. Calibrating the RBF LS-SVM Classifier

All classifiers were trained using RBF kernels.[3] Estimation of the generalization ability of the RBF LS-SVM classifier is then realized by the following experimental setup[3]:

(1) Set aside $\frac{3}{4}$ of the data for the training set and the remaining $\frac{1}{4}$ for testing, respecting the original class distribution.
(2) Perform 10-fold cross validation on the training data for each (σ, γ) combination from the initial candidate tuning sets Σ and Γ typically chosen as follows:

$$\Sigma = \{0.5, 5, 10, 15, 25, 50, 100, 250, 500\} \cdot \sqrt{n}$$

$$\Gamma = \{0.01, 0.5, 1, 10, 50, 100, 500, 1000\} \cdot \frac{1}{N}$$

The square root $\sqrt{n}$ of the number of inputs n is introduced, since $\|x - x_i\|_2^2$ in the RBF kernel is proportional to n and the factor $\frac{1}{N}$ is introduced such that the misclassification term $\gamma\sum_{i=1}^{N} e_i^2$ is normalized with the size of the data set.

(3) Choose optimal (σ, γ) from the initial candidate tuning sets Σ and Γ by looking at the best cross validation performance for each (σ, γ) combination.

(4) Refine Σ and Γ iteratively by means of a grid search mechanism to further optimize the tuning parameters (σ, γ). In our experiments, we repeated this step three times.

(5) Construct the LS-SVM classifier using the total training set for the optimal choice of the tuned hyperparameters (σ, γ).

(6) Assess the generalization ability by means of the independent test set.

Following the procedure outlined above, one obtained the results depicted in Table III. The optimized RBF LS-SVM classifier, trained on $\frac{3}{4}$ of the data set, achieves a percentage correctly classified on the training data of 77.54% with $\sigma = 13.75$ and $\gamma = 1.50$. Performance on the independent test data amounts to 74.48% correctly classified. We contrasted these results with those obtained using a linear kernel for the LS-SVM classifier. As can be observed from Table III, the percentage correctly classified drops to 76.26% on the training set and to 73.76% on the independent test set.

IV. THE INPUT SELECTION EXPERIMENT

A. Input Selection in a Nutshell

Input selection is a commonly adhered technique to reduce model complexity. The goal is to find a reduced coordinate system that allows one to project the data on a more compact representation. The general assumption underlying this operation and justifying it, is that the studied data approximately lie within the bounds of this reduced space. As such, models with fewer inputs are capable of improving both human understanding and computational performance. Moreover, elimination of redundant and/or irrelevant inputs may also improve the predictive power of an algorithm.[37] Selecting the best subset of a set of n predictors is a nontrivial problem. This follows from the fact that the optimal input subset can only be obtained when the input space is exhaustively searched. When n inputs are present, this would imply the need to evaluate 2^n-1 input subsets. Unfortunately, as n grows, this very quickly becomes computationally infeasible. For that reason, heuristic search procedures are often preferred. Input selection can then either be performed as a preprocessing step, indepen-

Table III. Classification accuracy of the optimized RFB LS-SVM classifier versus an LS-SVM optimized using a linear kernel.

Classification Accuracy	LS-SVM RBF Kernel	LS-SVM Linear Kernel
Training (3750 observations)	77.54%	76.26%
Test (1250 observations)	74.48%	73.76%

dent of the induction algorithm, or explicitly make use of it. The former approach is termed filter, the latter wrapper.[38] Filter methods operate independently of the learning algorithm. Undesirable inputs are filtered out of the data before induction commences. Focus[39] and Relief[40] are well-known filter methods. Wrapper methods make use of the actual target learning algorithm to evaluate the usefulness of inputs. Typically, the input evaluation heuristic that is used is based upon inspection of the trained parameters and/or comparison of predictive performance under different input subset configurations. Input selection is then often performed in a sequential fashion, e.g., guided by a best-first input selection strategy. The backward selection scheme starts from a full input set and stepwise prunes input variables that are undesirable. The forward selection scheme starts from the empty input set and stepwise adds input variables that are desirable. Hybrids of the above also exist.

B. Wrapping the Optimized LS-SVM Classifier

Input selection effectively starts at the moment the LS-SVM classifier has been constructed on the full set of n available predictors. The input selection procedure is based upon a (greedy) best-first heuristic, guiding a backward search mechanism through the input space.[38] The mechanics of the implemented heuristic for assessing the sensitivity of the classifier to a certain input are quite straightforward. We apply a strategy of constant substitution in which an input is perturbed to its mean while all other inputs keep their values and compute the impact of this operation on the performance of the obtained LS-SVM classifier without reestimation of the LS-SVM parameters α and b. This assessment is done using the separate pruning set to obtain an unbiased estimate of the change in classification accuracy of the constructed classifier. The pruning set consists of 1250 observations that were randomly selected from the training set of 3750 observations. Figure 1 provides a concise overview of the different steps of the experimental procedure.

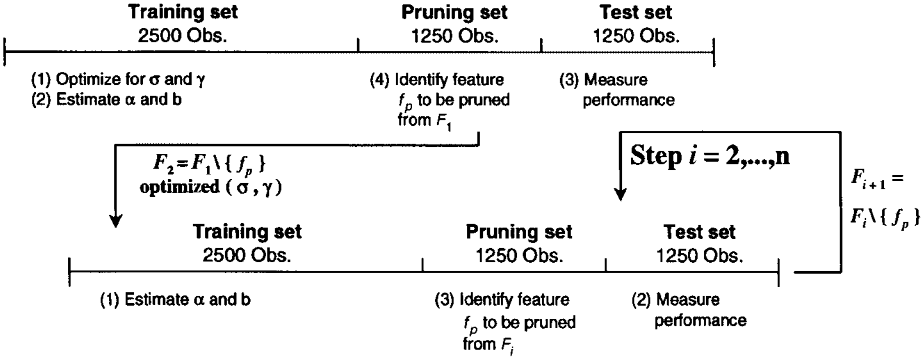

Figure 1.

Starting with a full input set F_1, all n inputs are pruned sequentially, i.e., one by one. The first input f_p to be removed, is determined at the end of Step 1 [task (4)]. After having removed this input from F_1, the reduced input set $F_2 = F_1 \setminus \{f_p\}$ is used for subsequent input removal. At this moment, an iteration of identical Steps i is started, in which, in a first phase, the LS-SVM parameters α and b are re-estimated on the training set [task (1) of Step i], however, without recalibration for σ and γ, and the generalization ability of the classifier is quantified on the independent test set [task (2) of Step i]. Notice that the originally optimized γ and σ values obtained in task (1) of Step 1 remain unchanged during the entire input selection phase. Again, input sensitivities of the resulting classification model (without re-estimation of α and b) are assessed on the pruning set to identify the input to which the classifier is least sensitive when perturbed to its mean [task (3) of Step i]. This input is then pruned from the remaining input subset and disregarded for further analysis. The pruning procedure is thereupon resumed with a reduced input set, until all inputs are eventually removed. Once all inputs have been pruned, the preferred reduced model is then determined by means of the highest pruning set performance.

Table IV summarizes the empirical findings of the pruning procedure for the RFM case. Observe how the suggested input selection method allows significant reduction of model complexity (from 25 to 9 inputs) without any significant degradation of the generalization behavior on the independent test set. The test set performance amounts to 73.92% for the full model and 73.52% for the reduced model.

The order of input removal as depicted in Table V, provides further insight into the relative importance of the predictor categories (cf. Table II). The reduced model consists of the nine inputs that are underlined in Table V. This reduced set of predictors consists of frequency, monetary, and other (non-RFM) variables. It is especially important to note that the reduced model includes information on returned merchandise. Furthermore, notice the absence of the recency component in the reduced input set. Inspection of the order of removal of inputs, while further pruning this reduced input set, highlights the relative importance of the frequency variables. More specifically, the last two variables to be removed belong to this predictor category. Note that an input set consisting of only these two inputs, still yields a percentage correctly classified at 72.00% on the test set. Results also point to the beneficial effect of including the

Table IV. Empirical assessment of the RBF LS-SVM classifiers for the full and reduced models.

Classification Accuracy	Full Model	Reduced Model
Training (2500 observations)	77.36%	76.04%
Pruning (1250 observations)	76.72%	77.20%
Test (1250 observations)	73.92%	73.52%
Number of Inputs	25	9

Table V. Order of input removal. Each input is qualified by its category with r, f, m and o respectively standing for recency, frequency, monetary and other (cf. Table II).

Pruning Steps									
1–5		6–10		11–15		16–20		21–25	
RetPerc	o	ProdclaM	o	RecHistN	r	FrYearN	f	MonYearR	m
ln(MonHistN)	m	MonHistR	m	IncrHist	o	ln(MonHistR)	m	MonYearN	m
RecHistR	r	IncrYear	o	RecYearR	r	MonHistN	m	GenInfo	o
Ndays	o	ln(MonYearR)	m	RecYearN	r	GenCust	o	FrHistR	f
ProdclaT	o	ln(MonYearN)	m	FrHistN	f	RetMerch	o	FrYearR	f

non-RFM customer profiling variables GenInfo and GenCust for improving predictive accuracy. They underline that customer/company interaction variables, here measured by indicators of information requests and complaints, provide additional predictive power to purchase incidence modeling for database marketing.

V. CONCLUSION

In this paper, we applied an LS-SVM based input selection wrapper to a real-life direct marketing case involving the modeling of repeat-purchase behavior. Based on a thorough review of the literature, we extended the well-known recency, frequency, monetary (RFM) framework (1) by using alternative operationalizations of the original variables, and (2) by including several additional behavioral variables. The sensitivity based, stepwise input selection method, constructed as a wrapper around the LS-SVM classifier, allows significant reduction of model complexity without degrading predictive performance. The empirical findings highlight the role of frequency and monetary variables in the reduced model, while the recency variable category seems to be of somewhat lesser importance within the response model. Results also point to the beneficial effect of including non-RFM customer profiling variables for improving predictive accuracy. More specifically, customer/company interaction, measured by indicators of information requests and complaints, and merchandise returns provide additional predictive power to purchase incidence modeling for database marketing.

This work was partly carried out at the Leuven Institute for Research on Information Systems (LIRIS) of the Department of Applied Economic Sciences of the K.U. Leuven in the framework of the KBC Insurance Research Chair. This work was partly carried out at the ESAT laboratory and the Interdisciplinary Center of Neural Networks ICNN of the K.U. Leuven. S. Viaene, holder of the KBC Insurance Research Chair, and B. Baesens are both Research Assistants of LIRIS. T. Van Gestel is a research assistant with the Fund for Scientific Research Flanders (FWO-Flanders), J. Suykens is a postdoctoral researcher with the Fund for Scientific Research Flanders (FWO-Flanders), B. De Moor is a senior research associate with the Fund for Scientific Research Flanders (FWO-Flanders). D. Van den Poel is an assistant professor at the Department of

Marketing at Ghent University. J. Vanthienen and G. Dedene are Senior Research Associates of LIRIS.

References

1. Cullinan GJ. Picking them by their batting averages' recency-frequency-monetary method of controlling circulation. Direct Mail/Marketing Association, New York, manual release 2103 edition, 1977.
2. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293–300.
3. Van Gestel T, Suykens JAK, Baesens B, Viaene S, Vanthienen J, Dedene G, De Moor B, Vandewalle J. Benchmarking least squares support vector machine classifiers. Technical Report 00-37, ESAT-SISTA, K.U. Leuven, Leuven, Belgium, 2000.
4. Bauer A. A direct mail customer purchase model. J Direct Marketing 1988;2(3):16–24.
5. Kestnbaum RD. Quantitative database methods. In: The direct marketing handbook. New York: McGraw Hill; 1992. p 588–597.
6. Bult JR. Target selection for direct marketing, PhD thesis, Groningen University, 1993.
7. Levin N, Zahavi J. Continuous predictive modeling: a comparative analysis. J Interactive Marketing 1998;12(2):5–22.
8. Van der Scheer HR. Quantitative approaches for profit maximization in direct marketing, PhD thesis, Groningen University, 1998.
9. Dekimpe MG, Degraeve Z. The attrition of volunteers. European J Operat Res 1997;98:37–51.
10. Van den Poel D, Leunis J. Database marketing modeling for financial services using hazard rate models. Internat Rev Retail, Distribution and Consumer Res 1998; 8(2):243–257.
11. Van den Poel D. Response Modeling for Database Marketing using Binary Classification, PhD thesis, K.U. Leuven, 1999.
12. Nash EL. Direct marketing: strategy, planning, execution. Third Ed. New York: McGraw Hill; 1994.
13. Zahavi J, Levin N. Issues and problems in applying neural computing to target marketing. J Direct Marketing 1997;11(4):63–75.
14. Hauser B. List Segmentation. In: The direct marketing handbook. New York: McGraw-Hill; 1992. p 233–247.
15. DMA. Statistical fact book 1998. Twentieth Ed. New York: Direct Marketing Association; 1998.
16. Berger P, Magliozzi T. The effect of sample size and proportion of buyers in the sample on the performance of list segmentation equations generated by regression analysis. J Direct Marketing 1992;6(1):13–22.
17. Bitran GR, Mondschein SV. Mailing decisions in the catalog sales industry. Manage Sci 1996;42(9):1364–1381.
18. Bult JR, Wittink DR. Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising. Int J Res Marketing 1996;13:215–226.
19. Bult JR. Semiparametric versus parametric classification models: an application to direct marketing. J Marketing Res 1993;30:380–390.
20. Bult JR, Van der Scheer H, Wansbeek T. Interaction between target and mailing characteristics in direct marketing, with an application to health care fund raising. Int J Res Marketing 1997;14:301–308.
21. Gönül F, Shi MZ. Optimal mailing of catalogs: a new methodology using estimable structural dynamic programming models. Manage Sci 1998;44(9):1249–1262.
22. Kaslow GA. A microeconomic analysis of consumer response to direct marketing and mail order, PhD thesis, California Institute of Technology, 1997.

23. Magliozzi TL, Berger PD. List segmentation strategies in direct marketing. Omega Int J Manage Sci 1993;21(1):61–72.
24. Magliozzi TL. An empirical investigation of regression meta-strategies for direct marketing list segmentation models, PhD thesis, Boston University, 1989.
25. Thrasher RP. Cart: a recent advance in tree-structured list segmentation methodology. J Direct Marketing 1991;5(1):35–47.
26. Stone B. Successful direct marketing methods. Chicago: Crain books; 1984.
27. Levin N, Zahavi J. Segmentation analysis with managerial judgment. J Direct Marketing 1996;10(3):28–47.
28. Little RJA. Regression with missing x's: a review. J Amer Statist Assoc 1992;87(420):1227–1230.
29. Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995.
30. Cristianini N, Shawe-Taylor J. An introduction to support vector machines. Cambridge, UK: Cambridge University Press; 2000.
31. Smola A. Learning with kernels, PhD thesis, Technical University, Berlin, 1999.
32. Vapnik V. The nature of statistical learning theory. New York: Springer-Verlag; 1995.
33. Vapnik V. Statistical learning theory. New York: John Wiley & Sons; 1998.
34. Suykens JAK, Lukas L, Van Dooren P, De Moor B, Vandewalle J. Least squares support vector machine classifiers: a large scale algorithm. In: 14th European Conference on Circuits Theory and Design, Stresa, Italy, 1999. p 839–842.
35. Suykens JAK, Vandewalle J. Multiclass least squares support vector machines. In: 10th International Joint Conference on Neural Networks, Washington DC, 1999.
36. Suykens JAK, Lukas L, Vandewalle J. Sparse least squares support vector machine classifiers. In: 9th European Symposium on Artificial Neural Networks, Bruges, Belgium; 2000. p 37–42.
37. Bellman RE. Adaptive control processes. Princeton: Princeton University Press; 1961.
38. John G, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem. In: Machine Learning: Proc Eleventh Int Conf, San Francisco, California, 1994. p 121–129.
39. Almuallim H, Dietterich TG. Learning with many irrelevant features. In: Ninth Nat Conf on Artificial Intelligence, Anaheim, California. Menlo Park, CA: AAAI Press; 1991. p 547–552.
40. Kira K, Rendell LA. The feature selection problem: Traditional methods and a new algorithm. In: Tenth Nat Conf on Artificial Intelligence, San Jose, California. Menlo Park, CA: AAAI Press; 1992. p 129–134.