반응 모델링을 위한 이상탐지 기법

Novelty Detection Methods for Response Modeling

이형주*, 조성준**

* 서울대학교 산업공학과 (impatton@snu.ac.kr)
** 서울대학교 산업공학과 (zoon@snu.ac.kr)

Abstract

본 논문에서는 반응 모델링에서의 집단 불균형을 해소하기 위한 이상탐지 기법의 활용을 제안한다. DMEF4 데이터셋의 카탈로그 발송 작업에 대하여 두 가지의 이상탐지 기법, one-class support vector machine (1-SVM)과 learning vector quantization for novelty detection (LVQ-ND)을 적용하여 이진분류 기법들과 비교한다. 반응률이 낮은 경우에는 이상 탐지 기법들이 더 높은 정확도를 보인 반면, 반응률이 상대적으로 높은 경우에는 오분류 비용을 조정한 SVM 기법이 가장 좋은 성능을 보였다. 또한, 이상탐지 기법들은 발송비용이 낮은 경우에 높은 이익을 달성하였고, 발송비용이 높은 경우에는 SVM 모델이 가장 높은 이익을 달성하였다.

1. Introduction

Response modeling is usually formulated as a binary classification problem. The customers are divided into two classes, respondents and non-respondents. A classifier is constructed to predict whether a given customer will respond or not. Various classifiers have been employed such as logistic regression (LR), neural networks (NNs), and support vector machines (SVMs). From a mod-eling point of view, however, several difficulties arise (Shin and Cho, 2006; Zahavi and Levin, 1997). One of the most noticeable is a severe class imbalance resulting from a low response rate: typically less than 5% of customers are respondents (Gönül et al., 2000). A typical binary classifier will result in lopsided outputs to the non-respondent class (Kubat et al., 1997). In other words, the classifier will predict most or even all customers not to respond. Although the classification accuracy may be very high since a majority of customers are in fact non-respondents, that is not what we are interested in. We would like to construct a model which identifies a subset of customers that includes as many respondents and as few non-respondents as possible. Therefore, a balanced model is preferred although its accuracy may be lower than an unbalanced one.

There are a few balancing methods that can be used for imbalanced class problems such as under-sampling, over-sampling and cost-modifying methods (Domingos, 1999; Weiss, 2004). One can also employ a novelty detection approach (Japkowicz, 2001). In novelty detection, one of the classes, usually the majority class, is designated as normal while the other class as normal. A model learns the characteristics of the normal patterns in training data and detects novel patterns that are different from the normal ones (Bishop, 1994). In a geometric sense, the model generates a closed boundary around the normal patterns (Schölkopf et al., 2001). Real world applications include speaker identification (Gori et al., 1996), currency validation (Frosini et al., 1996; He et al., 2004) and machine fault detection (Tax and Duin, 2004). Various novelty detection methods have been proposed for such applications (Markou and Singh, 2003a, 2003b; Marland, 2003; Tax, 2001).

We propose to use novelty detection approaches for response modeling problems. In particular, one-class support vector machine (1-SVM) and learning vector quantization for novelty detection (LVQ-ND) (Lee and Cho, 2005) are considered for a catalogue mailing task with DMEF4[1]) dataset from the Direct Marketing Educational Foundation (DMEF). They are compared with two binary classifiers, logistic regression (LR) and support vector machine (SVM). It is shown that under a certain condition, the novelty detectors outperform the binary classifiers and that the novelty detection approaches can be viable solutions to the class imbalance in response modeling. In addition, a sensitivity analysis on the mailing cost is conducted for the response models.

The following section reviews novelty detection approaches for response modeling, 1-SVM and LVQ-ND. In Section 3, DMEF4 dataset and the experimental settings are described while the experimental results are presented in Section 4. Section 5 concludes this paper and discusses some

1) The Direct Marketing Association. Available at http://www.the-dma.org/dmef/dmefdset.shtml.

issues and future research directions.

2. Novelty Detection Approaches

We propose to use novelty detection approaches for response modeling. They have successfully alleviated the class imbalance problems from other domains (Japkowicz, 2001; Raskutti and Kowalczyk, 2004). They are especially useful when the class imbalance is extreme. The idea is to train a novelty detector exclusively with the normal patterns with unsupervised learning.

In response modeling, however, novel patterns are also available since there are always two classes. There have been approaches to utilize them. Gori et al. (1996) and Frosini et al. (1996), while training an auto-associative neural network (AANN), added a penalty term to the network error function so as to prevent novel patterns from being accepted. Tax and Duin (2004) proposed support vector data description (SVDD) to utilize novel data. SVDD defines a hypersphere with a minimal radius so that it surrounds as many normal patterns and at the same time as few novel patterns as possible. Recently, Lee and Cho (2005) proposed LVQ-ND which will be introduced later in this section.

Fig. 1 illustrates the decision boundaries generated by a binary classifier, a novelty detector trained only with one class, and a novelty detector trained with two classes. The classifier generates an open boundary between the two classes and will classify patterns in the upper-right corner as normal although they are not likely to belong to the normal class. That is because it is trained with the extremely underrepresented novel patterns. The novelty detector trained only with the normal class generates a closed boundary around the normal patterns, but cannot reject three novel patterns because it does not consider the novel patterns during training. On the other hand, the novelty detector trained with both classes generates a similar closed boundary except that it has been adjusted to reject the novel patterns.

Now we review two novelty detectors, 1-SVM and LVQ-ND. The former is trained only with the normal patterns, while the latter is trained with the normal patterns as well as novel patterns. In response modeling, either class can be designated as normal. We decided through preliminary experiments that the majority class, i.e. the non-respondent class, should be normal while the respondent class is novel. Therefore, the class labels were reversed with +1 for the non-respondents and -1 for the respondents. So 1-SVM and LVQ-ND are trained with the non-respondent patterns as normal. Given a new customer pattern, they perform classification by determining whether it belongs to the non-respondent class or not.

2.1 One-class Support Vector Machine (1-SVM)

1-SVM was proposed by Schölkopf et al. (2001) as a special case of SVM. 1-SVM finds a function that returns +1 for small regions containing most normal data and -1 for all other regions. A hyperplane **w** is defined to separate a fraction of patterns from the origin in a feature space by a maximal margin. An optimization problem can be considered as follows,

$$\min \frac{1}{2}\|\mathbf{w}\|^2 - \rho + \frac{1}{\nu N_R}\sum_{\mathbf{x}_i \in \mathbf{X}_R} \xi_i,$$
$$\text{s.t } \mathbf{w}^T \Phi(\mathbf{x}_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \cdots, N,$$

where $\nu \in (0, 1]$ is a trade-off parameter between the margin and the training error. The solution, which can be obtained analogously to SVM, satisfies sparsity, most of the Lagrangian multipliers being zero. Given a customer's pattern **x**, the decision function is expressed in terms of the expansion of the kernel functions:

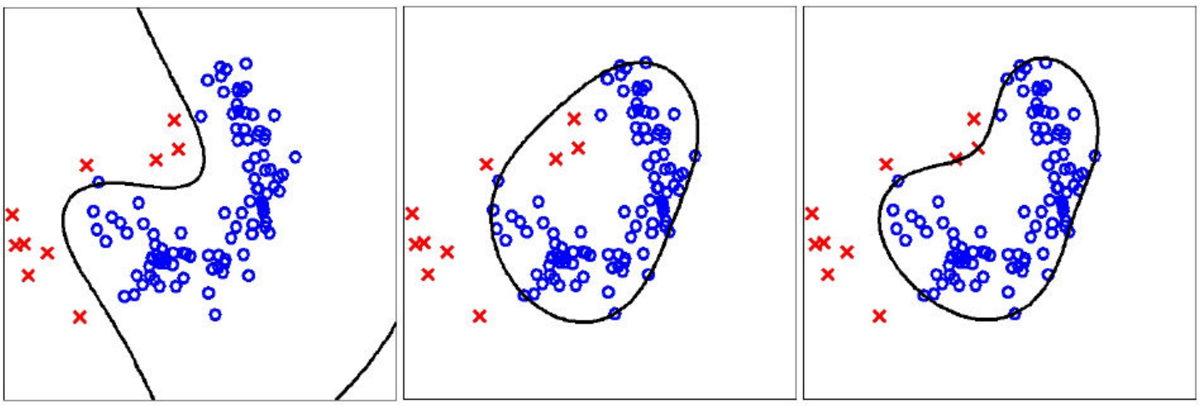


Fig. 1. Decision boundaries of a binary classifier (left), a novelty detector trained only with the normal patterns (middle) and a novelty detector trained with two classes of patterns (right): In a two-dimensional space, 100 normal patterns (circles) and 10 novel patterns (crosses) are generated. The solid curves represent the decision boundaries separating the two classes.

$$f(\mathbf{x}) = \mathrm{sign}\Big[\mathbf{w}^T\boldsymbol{\Phi}(\mathbf{x}) - \rho\Big] = \mathrm{sign}\Big[\sum_{\mathbf{x}_i \in \mathrm{SV}} \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho\Big].$$

If $f(\mathbf{x})=+1$, the customer is classified as a non-respondent, or as a respondent otherwise. The RBF kernel is used for 1-SVM.

2.2 Learning Vector Quantization for Novelty Detection (LVQ-ND)

LVQ-ND was recently proposed so as to utilize novel patterns in training codebook-based novelty detectors (Lee and Cho, 2005). We call it LVQ-ND since the codebook update rule resembles the original LVQ. Codebook methods such as k-means clustering applied to the normal data generates a set of codebooks (or cluster centers), $\mathbf{W}=\{\mathbf{w}_k|k=1,2,...,K\}$ which represents the normal data. The codebook $\mathrm{m}(\mathbf{x})$ of an input pattern $\mathbf{x}$ and the Voronoi region $\mathbf{S}_k$ of each codebook $\mathbf{w}_k$ are defined as follows,

$$\begin{aligned} &\mathrm{m}(\mathbf{x}) = \mathbf{w}_k \iff \mathbf{x} \in \mathbf{S}_k \\ &\iff \|\mathbf{w}_k - \mathbf{x}\|^2 < \|\mathbf{w}_l - \mathbf{x}\|^2, \ \forall l \neq k. \end{aligned}$$

With the novel patterns in the training set, the error function of LVQ-ND can be defined as

$$\begin{aligned} \bar{e}(\mathbf{x}) &= \frac{1}{N}\sum_i y_i \|\mathbf{x}_i - \mathrm{m}(\mathbf{x}_i)\|^2 \\ &= \frac{1}{N}\sum_k \Bigg[\sum_{\mathbf{x}_i \in \mathbf{S}_{R_k}} \|\mathbf{x}_i - \mathbf{w}_k\|^2 - \sum_{\mathbf{x}_i \in \mathbf{S}_{NR_k}} \|\mathbf{x}_i - \mathbf{w}_k\|^2\Bigg], \end{aligned}$$

where $\mathbf{S}_{Rk}=\mathbf{X}_R \cap \mathbf{S}_k$ and $\mathbf{S}_{NRk}=\mathbf{X}_{NR} \cap \mathbf{S}_k$. Minimizing the error function forces the codebooks to be located close to normal patterns and far away from novel ones, leading to a learning rule different from the conventional LVQ algorithm. Given an input pattern $\mathbf{x}_i$, $\mathbf{w}_k$ is updated as follows:

$$\mathbf{w}_k \leftarrow \begin{cases} \mathbf{w}_k, & \text{if } \mathbf{x}_i \notin \mathbf{S}_k, \\ \mathbf{w}_k + \eta(\mathbf{x}_i - \mathbf{w}_k), & \text{if } \mathbf{x}_i \in \mathbf{S}_{R_k}, \\ \mathbf{w}_k - \eta(\mathbf{x}_i - \mathbf{w}_k), & \text{if } \mathbf{x}_i \in \mathbf{S}_{NR_k}. \end{cases}$$

According to this rule, if a pattern does not belong to the Voronoi region $\mathbf{S}_k$ that $\mathbf{w}_k$ represents, $\mathbf{w}_k$ remains unchanged. If $\mathbf{x}_i$ does belong to $\mathbf{S}_k$, $\mathbf{w}_k$ moves toward $\mathbf{x}_i$ if $\mathbf{x}_i$ is normal, or moves away from $\mathbf{x}_i$ if $\mathbf{x}_i$ is novel. That is, normal patterns "pull" their codebooks while novel patterns "push" theirs.

A pattern is accepted or rejected according to its quantization error, the distance to its codebook, i.e. $e(\mathbf{x})=\|\mathbf{x}-\mathbf{m}(\mathbf{x})\|^2$, for which a threshold should be explicitly determined. While some codebooks lie inside dense lumps of input patterns, others lie in regions where patterns are sparsely scattered. For that reason, it is desirable to set different thresholds for different codebooks. A hypersphere centered at $\mathbf{w}_k$ is found for each Voronoi region. It is desirable that the hypersphere includes as many normal patterns and excludes as many novel patterns as possible while having a small radius. This can be formulated as an optimization problem:

$$\begin{aligned} \min_{r_k} E_k(r_k^2) = r_k^2 &+ C_1 \sum_{\substack{\mathbf{x}_i \in \mathbf{S}_{R_k} \\ e(\mathbf{x}_i) > r_k^2}} \Big(e(\mathbf{x}_i) - r_k^2\Big) \\ &+ C_2 \sum_{\substack{\mathbf{x}_i \in \mathbf{S}_{NR_k} \\ e(\mathbf{x}_i) < r_k^2}} \Big(r_k^2 - e(\mathbf{x}_i)\Big). \end{aligned}$$

Here the radius acts as a threshold. On one hand, a hypersphere with a large radius can surround many normal patterns, but may increase false acceptance. On the other hand, a hypersphere with a small radius can exclude many novel patterns, but may increase false rejection. The trade-off between false acceptance and false rejection is controlled by two positive constants, C_1 and C_2. The optimal radius can be found by replacing r_k by $\mathrm{e}(\mathbf{x}_i)$; $\forall \mathbf{x}_i \in \mathbf{S}_{Rk}$, which is simply an exhaustive search with $|\mathbf{S}_{Rk}|$ computations for each codebook. Ultimately, a small threshold is found for a codebook in a dense region while a large threshold for one in a sparse region. Given a customer's pattern $\mathbf{x}$, its codebook $\mathbf{m}(\mathbf{x})=\mathbf{w}_q$ is found. Then, the decision function is written as follows,

$$f(\mathbf{x}) = \mathrm{sign}\Big[r_q^2 - \|\mathbf{x} - \mathbf{w}_q\|^2\Big].$$

The customer will be classified as a non-respondent if $f(\mathbf{x})=+1$, or as a respondent otherwise.

3. Dataset and Experimental Settings

3.1 DMEF4 Dataset

A catalogue mailing task involving DMEF4 dataset was analyzed. It is concerned with an up-scale gift business that mails general and specialized catalogs to its customers several times each year. The original problem is to estimate how much each customer will spend during the test period, from September 1992 to December 1992, based on the base time period, from December 1971 to June 1992. From the original problem, a classification problem is formulated where the target class labels are +1 for respondents who spent a non-zero amount and -1 for non-respondents who did not spend at all. The dataset contains 101,532 customers each of whom is described by 91 input variables. The response rate is 9.4% with 9,571 respondents and 91,961 non-respondents, which means that the class distribution is moderately imbalanced. While selecting or extracting relevant variables is very important, it is not our main concern. Malthouse (2001) extracted 17 out of the 91 input variables for this dataset, and Ha et al. (2005) used 15 among them, removing two variables whose variations are negligible. In this paper, these 15 variables were used as input variables.

The dataset was partitioned into training and test sets for performance evaluation. A half of customers were randomly assigned to the training set while the other half to the test set. Since performance of a model shows a large variation with regard to a specific data split (Malthouse, 2001), ten different training/test splits were generated. All experimental results were averaged over the ten test sets. The effecct of different response rates was also investigated. Although the response rate is 9.4% in DMEF4 dataset, it is lower than that in typical response modeling tasks (Gönül et al., 2000). For each split, six additional training sets were generated by randomly sampling respondent patterns so that the response rates were 0.1, 0.5, 1, 3, 5, and 7%, respectively.

3.2 Response Models

Two classification models and two novelty detection models were constructed. The two classifiers are LR and SVM. For an LR model, a new training set was constructed by randomly sampling a number of the non-respondents so that there were equal numbers of patterns in both classes. For an SVM model the dataset was not modified, but a modified optimization problem was solved where the modified cost coefficients were used. For novelty detectors, since we designated the non-respondents as normal, the class labels were reversed with +1 for the non-respondents and -1 for the respondents. 1-SVM was trained only with the non-respondent class while LVQ-ND was trained with not only the non-respondents and but also the small number of respondents.

In order to implement each model, a particular set of parameters should be selected in advance. LR has no parameter to be pre-specified. For SVM and 1-SVM, the kernel width and the trade-off parameter have to be specified in advance. For LVQ-ND, one should predetermine the number of codebooks and the cost coeefficients. Five-fold cross validation was conducted on the training sets for each model and the best parameter set was selected which resulted in the best model selection criterion. We employed the "ROC distance" in Eq.(15) as the criterion similarly to He et al. (2004) and Yu and Cho (2006):

$$\text{ROC distance}=FPR^2+FNR^2.$$

FPs correspond to wasted marketing costs while FNs to opportunities lost. An ROC distance indicates how distant the result of a model is from the perfect classification in an ROC chart. To achieve a small ROC distance, both FP and FN should have low values. The more correct a model is, the smaller the ROC distance becomes.

3.3 Performance Measures

We measured model performances in terms of both goodness-of-fit and profit. In response modeling, goodness-of-fit and profit are not necessarily equivalent (Malthouse, 2002). A more accurate model can yield a lower profit, or vice versa. The simple accuracy was not considered since it is inadequate for an imbalanced class problem (Weiss, 2004). For example, with 9.4% response rate, 90.6% accuracy can be achieved by simply classifying every customer as a non-respondent. Measures that give balanced assessments on the two classes have to be adopted such as balanced classification rate (BCR) (Shin and Cho, 2006; Yu and Cho, 2006) which incorporates TPR and TNR in the following way:

$$BCR=\sqrt{TPR\cdot TNR}.$$

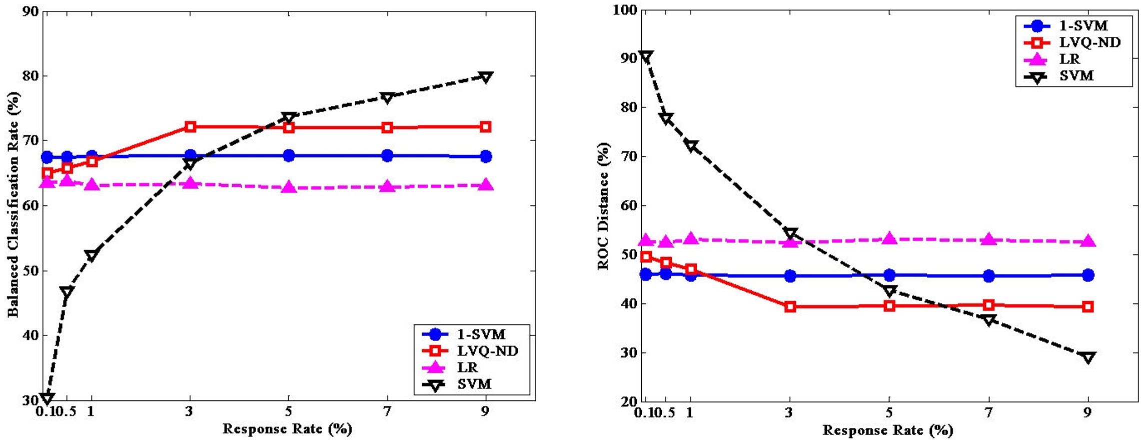
Model profit should be evaluated as to how much money the response model would make. Since DMEF does not provide information on the mailing cost, the cost per mail was assumed to be $1, 3, 5, 7, and 9. Model profits were evaluated in a sensitivity analysis. The total revenue of a model can be computed as the sum of the amounts spent by customers who are predicted to respond. The model profit is simply the total revenue subtracted by the total mailing cost. The total mailing cost grows with a high FP while the total revenue shrinks with a high FN.

4. Experimental Results

4.1 Goodness-of-Fit

Fig. 2 shows the average BCRs and ROC distances of the four models against the response rates. The goodness-of-fit of 1-SVM and LR, hardly changed as the response rates increased, while LVQ-ND and SVM improved accordingly. 1-SVM was not affected by the response rate, because it was trained only with the non-respondent patterns anyway. LR is not recommendable for this problem since it obviously fails to capture the non-linear relationships. When the response rate was very low at 1% or lower, 1-SVM was the most accurate model. When the response rate increased to 3%, LVQ-ND was the best. With the response rate of 5%, LVQ-ND and SVM performed the best, resulting in the comparable goodness-of-fits. Then, SVM became the most accurate with the higher response rates. Under an extreme imbalance, i.e. a response rate of 1% or lower, one should employ a novelty detector only with the majority class. It is not desirable to utilize the underrespresented minority class during training. With a less severe imbalance, i.e. a response rate of 3 to 5%, a novelty detector trained with both classes seems the most appropriate. The minority patterns of this proportion is not sufficient to train a binary classifier, but sufficient to refine a novelty detector. On the other hand, with a moderate imbalance, i.e. a response rate higher than 5%, a balanced binary classifier is the most appropriate. Since the minority patterns are relatively abundant to represent their own class reasonably well, it is desirable to adopt a binary classifier.

The ROC graphs of the four models are depicted in Fig. 3. An accurate response model will be in the lower left corner of the charts. First, SVM was strongly affected by response rates while others were not. SVM produced FPRs near zero but very high



(a) BCRs (b) ROC distances

Fig. 2. BCRs and ROC distances of LR, SVM, 1-SVM and LVQ-ND against the response rates.

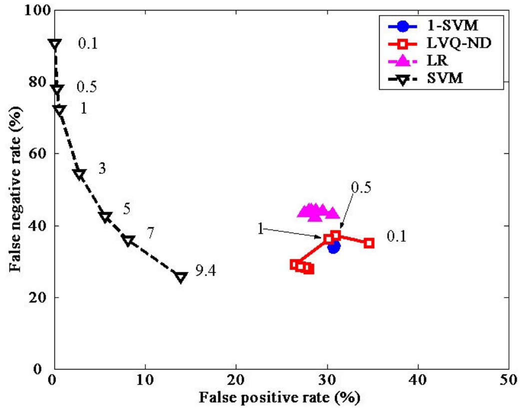


Fig. 3. ROC graphs of LR, SVM, 1-SVM and LVQ-ND with respect to the response rates: False positive (FP) and false negative (FN) rates are plotted for each model. The figures beside the markers indicate the response rates.

FNRs when the response rate was 0.1%. As the response rates increased, however, FPRs slightly increased but FNRs rapidly decreased, thus the overall performances significantly improved. SVM is the most conservative model, suitable when the mailing cost is high. Both 1-SVM and LVQ-ND are superior to LR. 1-SVM was slightly better than LVQ-ND when the response rate was very low with 1% or lower, but LVQ-ND caught up with 1-SVM as the response rate increased. LVQ-ND produced comparable values of FPRs and FNRs and each of them decreased as the response rates increased, although there was no improvement after the response rate exceeded 3%. LVQ-ND is suitable when the mailing cost is relatively low.

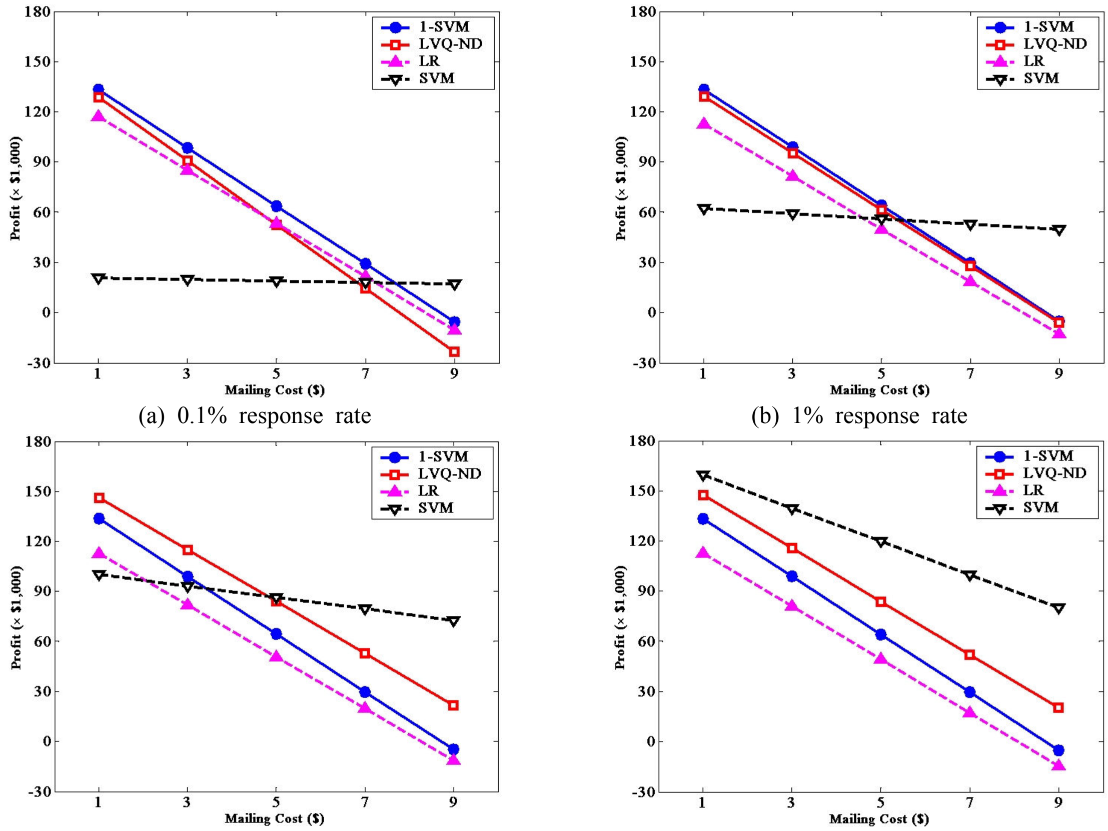
4.2 Profit

The average profits of the models are plotted against the mailing costs for various response rates of 0.1, 1, 3, and 9.4% in Fig. 4. The profits of SVM decreased little as the mailing cost increased, since it was the most conservative model and would mail to a very limited number of customers anyway. However, the profits of the other models rapidly decreased since they would send mails to relatively larger numbers of customers. The novelty detection models were better than the binary classifiers when the response rate was low and the mailing cost was low. 1-SVM was the most profitable with the response rate of 0.1% except when the mailing cost was $9. LVQ-ND was comparable to 1-SVM with the response rate of 1% but was the most profitable when the response rate was 3% with low mailing costs. On the other hand, SVM did best as the response rates increased and the mailing costs got higher. In particular, when the response rate was 9.4%, it yielded the most profits regardless of the response rates. Although LR was the most robust yielding virtually the same profits regardless of the response rates, its profits were not on a par with other models in most cases.

We can conclude that a novelty detector should be used as the response model if both the response rate and the mailing cost are low. However, if the response rate or the mailing cost is relatively high, a balanced binary classification model is more suitable, because with more respondents present in training data, a more accurate binary classifier can be obtained. In addition, a binary classifier is rather conservative, predicting a smaller number of respondents. Thus, when the mailing cost is high, a higher profit can be made.

4. Conclusions and Discussion

In response modeling, the class imbalance due to a low response rate is one of the most prevalent and noticeable difficulties. To alleviate the class imbalance, novelty detection approaches were proposed. We considered two novelty detectors, 1-SVM and LVQ-ND. Experiments were conducted on DMEF4 dataset. When the response rate was 5% or lower, the novelty detectors were more accurate than the binary classifiers in terms of BCR and ROC distance. In particular, with the response rate of 1%



(a) 0.1% response rate (b) 1% response rate

(c) 3% response rate (d) 9.4% response rate

Fig. 4. Profits of LR, SVM, 1-SVM and LVQ-ND against the mailing costs when the response rate is 0.1, 1, 3, or 9.4%. The results from the response rates of 0.5, 5 and 7% are not shown because they are not much different from those from the response rates of 0.1, 3, and 9.4%, respectively.

or lower, 1-SVM was the best model and with the response rate of 3 or 5%, LVQ-ND was the best. When the response rate exceeded 5%, on the other hand, SVM, a balanced binary classification model, came ahead of them. The sensitivity analysis on the mailing cost was also conducted. With a moderate response rate or a high mailing cost, a balanced binary classifier should be employed as the response model. On the other hand, with a very low response rate and a low mailing cost, a novelty detector should be employed. Since the response rates are often very low, novelty detection approaches can be good alternatives to binary classification models.

There are some limitations and future works to be done. First, we have considered a response modeling task as a pure classification problem. However, more generally, it can be formulated as a scoring or ranking problem, that is, scoring the customers by their likelihood to respond. LR, and SVM in a less straightforward manner, can produce a set of scores of the customers, while the novelty detectors cannot. If we can obtain scores of the customers from the novelty detectors, they can be compared using more comprehensive assessment tools such as the ROC analysis and the lift analysis. Second, the 15 variables used in the experiments were the ones extracted based on linear regression models (Malthouse, 2001). Since different methods may require different variables, variable selection schemes can improve the performance of the novelty detectors. Finally, the objective of novelty detection, in general, is to accept many normal patterns and to reject many novel patterns, given a specific bias in its pre-determined parameters. However, in response modeling, recognizing profitable customers may be more desirable even if it leads to missing some respondents or accepting some non-respondents. We may need to incorporate the concept of profit into the training process of novelty detectors.

References

Bishop, C. (1994), Novelty detection and neural network validation. *Proceedings of IEE Conference on Vision, Image and Signal Processing*, **141**(4), 217-222.

Cristianini, N., and Shawe-Taylor, J. (2000), *An introduction to support vector machines and other kernel-based learning methods*. Cambridge: Cambridge University Press.

Domingos, P. (1999), MetaCost : A general method for making classifiers cost-sensitive. *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining* (pp. 155-164). San Diego, US.

Frosini, A., Gori, M., and Priami, P. (1996), A neural network-based model for paper currency recognition and verification. *IEEE Transactions on Neural Networks*, **7**(6), 1482-1490.

Gönül, F.F., Kim, B.D., and Shi, M. (2000), Mailing smarter to catalog customers. *Journal of Interactive Marketing*, **14**(2), 2-16.

Gori, M., Lastrucci, L., and Soda, G. (1996), Autoassociator-based Models for Speaker Verification. *Pattern Recognition Letters*, **17**(3), 241-250.

Ha, K., Cho, S., and MacLachlan, D. (2005), Response models based on bagging neural networks. *Journal of Interactive Marketing*, **19**(1), 17-30.

He, C., Girolami, M., and Ross, G. (2004), Employing optimized combinations of one-class classifiers for automated currency validation. *Pattern Recognition*, **37**(6), 1085-1096.

He, Z., Xu, X., Huang, J.Z., and Deng, S. (2004), Mining class outliers: concepts, algorithms and applications in CRM. *Expert Systems with Applications*, **27**(4), 681-697.

Japkowicz, N. (2001), Supervised versus unsupervised binary-learning by feed-forward neural networks. *Machine Learning*, **42**(1-2), 97-122.

Kubat, M., Holte, R., and Matwin, S. (1997), Learning when negative examples abound. *Proceedings of the 9th European Conference on Machine Learning (ECML 97), Lecture Notes in Artificial Intelligence (LNAI 1224)* (pp. 146-153). Prague, The Czech Republic.

Lee, H., and Cho, S. (2005), SOM-based novelty detection using novel data. *Proceedings of Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL), Lecture Notes in Computer Science (LNCS 3578)* (pp. 359-366). Brisbane, Australia.

Malthouse, E.C. (2001), Assessing the performance of direct marketing scoring models. *Journal of Interactive Marketing*, **15**(1), 49-62.

Malthouse, E.C. (2002), Performance-based variable selection for scoring models. *Journal of Interactive Marketing*, **16**(4), 37-50.

Markou, M., and Singh, S. (2003a), Novelty detection: a review - part 1: statistical approaches. *Signal Processing*, **83**(12), 2481-2497.

Markou, M., and Singh, S. (2003b), Novelty detection: a review - part 2: neural network based approaches. *Signal Processing*, **83**(12), 2499-2521.

Marsland, S. (2003), Novelty detection in learning systems. *Neural Computing Surveys*, **3**, 157-195.

McLachlan, G.J. (1992), *Discriminant analysis and statistical pattern recognition*. New York: John Wiley and Sons.

Raskutti, B., and Kowalczyk, A. (2004), Extreme re-balancing for SVMs: a case study. *SIGKDD Explorations*, **6**(1), 60-69.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R.C. (2001), Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**(7), 1443-1471.

Shin, H.J., and Cho, S. (2006), Response modeling with support vector machines. *Expert Systems with Applications*, **30**(4), 746-760.

Tax, D.M.J. (2001), *One-class classification*. PhD Thesis. Delft: Delft University of Technology, The Netherlands.

Tax, D.M.J., and Duin, R.P.W. (2004), Support vector data description. *Machine Learning*, **54**(1), 45-66.

Weiss, G.M. (2004), Mining with rarity: a unifying framework. *SIGKDD Explorations*, **6**(1), 7-19.

Yu, E., and Cho, S. (2006), Constructing response model using ensemble based on feature subset selection. *Expert Systems with Applications*, **30**(2), 352-360.

Zahavi, J., and Levin, N. (1997), Issues and problems in applying neural computing to target marketing. *Journal of Direct Marketing*, **11**(4), 63-75.