Solving Imbalanced Classification Problems with Support Vector Machines.

1 author:

Stefan Lessmann
Humboldt-Universität zu Berlin
139 PUBLICATIONS   **2,628** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Reject Inference in Credit Scoring View project

Project   E-Mail and Web Tracking – Empirical Analysis, Privacy Impact, Countermeasures View project

SOLVING IMBALANCED CLASSIFICATION PROBLEMS WITH SUPPORT VECTOR MACHINES

Stefan Lessmann
Inst. of Business Information Systems
University of Hamburg, Germany
E-mail: Lessmann@econ.uni-hamburg.de

Abstract—**The Support Vector Machine (SVM) is a powerful learning mechanism and promising results have been obtained in the field of medical diagnostics and text-categorization. However, successful applications to business oriented classification problems are still limited. Most real world data sets exhibit vast class imbalances and an accurate identification of the economical relevant minority class is a major challenge within this domain. Based upon an empirical experiment, we evaluate the adequacy of SVMs to identify the respondents of a mailing campaign, massively underrepresented in our data set finding SVM to be capable of handling class imbalances in an internal manner providing robust and competitive results when compared to re-sampling methods which are commonly used to account for class imbalances. Consequently, the overall process of data pre-processing is simplified when applying a SVM classifier leading to less time consuming and more cost-efficient analysis.**

Keywords: support vector machine, sampling, imbalanced classification, data mining

I. INTRODUCTION

As technical innovations like the internet have led to a higher market transparency and increasing competition in the last years we observe a shift from the classical, mainly transaction oriented marketing towards a more customer relationship oriented one [1]. In a competitive consumer market customers are regarded as precious business resources and as a result the concept of Customer Relationship Management (CRM) [2], providing companies with techniques and tools to retain and exploit customer relationships, has found increasing consideration in management science.

At the core of CRM we have an analytical component, recording customer centered data in Data Warehouses and analyzing these data stores with mathematical methods originating from various scientific domains like statistics, artificial intelligence and machine learning. A common task within this domain is the prediction of a customer group membership, formally known as classification.[1]

Lately, SVMs [3] have found increasing consideration in the CRM community, providing effective and efficient solutions for managerial problems in similar domains. However, as most CRM related classification tasks involve the prediction of an – often heavily – underrepresented class of interest we evaluate the sensibility of SVM towards class imbalances, striving to exemplify their adequacy for the task of response optimization based upon an empirical, numerical experiment from an ongoing project with a large publishing house.

Following a brief introduction to analytical CRM (aCRM) and the relevance of classification in this domain section 3 assesses the problem of imbalanced class distributions and introduces techniques to overcome this issue. The principles of support vector classification are given in chapter 4 where we focus on SVM's inbuilt capabilities for modeling asymmetric misclassification costs. The suitability of different approaches to deal with imbalanced class distributions for the SVM classifier is evaluated within a numerical experiment in section 5. Conclusions are given in section 6.

II. CLASSIFICATION FOR ANALYTICAL CUSTOMER RELATIONSHIP MANAGEMENT

Whereas the CRM front-office includes techniques and tools for campaign management, sales force automation and service management the analytical back-end consist mainly of a data warehouse to record customer centered transaction and analytical components like online analytical processing and data mining, aiming at the detection of economically relevant information within the data masses in order to achieve operational, tactical and strategic competitive advantages [4].

Among the typical tasks in the field of data mining for aCRM, including regression, classification, segmentation and association, classification analysis is of primary importance with logistic regression and decision trees most widely used in practical applications. Typically, we have a specific class of interest, e.g. a group of customers with high probability of responding to direct mail, and strive to accurately identify these customers among all of our customers using a classifier which has learned a mapping between e.g. demographic and transaction-oriented customer data and the corresponding group membership, previously. Common aCRM tasks like

---

[1] Throughout this paper, we focus on concept-learning problems in which one class represents the concept at hand (positive class denoted by *B*) while the other represent counter-examples of the concept (negative class denoted by *A*).

response optimization, fraud detection, churn prediction and cross-selling [5] can be cast in this framework. Major challenges within this field are:

- a large number of attributes,
- asymmetric misclassification costs and
- highly imbalanced class distributions.

Having huge amounts of data at hand the problem of large attribute numbers is not that serious, especially as efficient algorithms for subset selection are available and modern classification techniques like SVM are able to provide stable results in high dimensional feature spaces [3] .

However, in order to obtain useful and profitable classification results the fact that the economically relevant group of customers is usually underrepresented in the data has to be considered. In our real world data set, further discussed in section 5, the proportion of relevant customers was for example below two percent, which is absolutely not untypical for this kind of problems. Such imbalances hinder classification and need to be addressed in an appropriate manner.

III. CLASSIFICATION WHEN CLASS DISTRIBUTIONS ARE IMBALANCED

A. Handling class imbalances

When class distributions are imbalanced traditional classification algorithms can be biased towards the majority class due to its over-prevalence [6]. This problem has been observed in various applications, as different as the identification of fraudulent telephone calls [7] and the detection of oil spills in satellite radar images [8]. We can categorize the proposed approaches to deal with imbalanced class distributions in internal ones, which modify existing algorithms to take the class imbalance into consideration, e.g. [8], and external ones that use unmodified existing algorithms, but resample the data in order to diminish the negative effect of the imbalance [9, 10].

Since the sensibility of a classifier for a specific class can be increased by assigning a higher cost of misclassification to this class [11], approaches for cost-sensitive classification are most prominent among the former category. As will be shown in the following section SVM supports this methodology by nature.

On the other hand, resampling aims at the elimination of the over-prevalence of one class, presenting only a selected subset of the available data to the classifier. Basically, this is accomplished either by randomly removing instances of the majority class population (under-sampling), or by randomly duplicating instances from the minority class (over-sampling), until some specified ratio between majority and minority examples is reached. A possible drawback of over-sampling is the fact that the decision region of the minority class becomes very specific, which can lead to over-fitting problems. Consequently, more advanced resampling procedures have been introduced that create synthetic examples of the minority class in the decision space [10] or form a hybrid resampling system by combining over- and under-sampling [12].

B. Measuring classifier performance in imbalanced environments

A major problem caused by class imbalances and ubiquitous in the aCRM area is to find an appropriate indicator to measure classifier performance. It is common practice to visualize the results of a classification analysis by means of a confusion matrix, as show in Table 1.

TABLE 1
Confusion matrix for binary classification problem
with output domain $\{A,B\}$

		predicted		
		A	B	Σ
actual	A	h_{00}	h_{01}	$h_{0.}$
	B	h_{10}	h_{11}	$h_{1.}$
	Σ	$h_{.0}$	$h_{.1}$	L

Ordinary performance measures can be derived directly or indirectly from the confusion matrix with accuracy, given as $\frac{h_{00} + h_{11}}{L}$ being the indicator most widely used in practical applications. However, accuracy is known to be inappropriate whenever class and/or cost distributions are highly imbalanced, as it is trivial to obtain a low error rate by simply ignoring the minority class completely, thereby achieving classification accuracy as high as the prior probability of the majority class [13].

Receiver operating characteristics (ROC) analysis is a powerful tool to compare induction algorithms in such imprecise environments and offers the possibility to determine a cost-optimal classifier [14]. A major drawback of ROC analysis is the fact, that it does not deliver a single, easy to use performance measure like accuracy directly. An alternative is to use the area under the ROC curve (AUC) for single number evaluation [15] or the geometric mean of sensitivity and specificity [16] instead. This leads to the measure $G = \sqrt{sensitivity * specificity} = \sqrt{\frac{h_{11}}{h_{1.}} * \frac{h_{00}}{h_{0.}}}$, which strives to maximize the accuracies of each individual class while keeping them balanced and is directly related to a point in ROC-space. [9]

Other prominent performance metrics include:

Precision $P = \frac{h_{11}}{h_{.1}}$; corresponding to the proportion of examples classified as positive that are truly positive and Recall $R = \frac{h_{11}}{h_{1.}}$; giving the proportion of truly positives examples that were correctly classified and being identical to sen-

sitivity. P and R are both desirable and typically trade off against each other so that it is convenient to combine them in a single measure called the F-measure (F) [17] which is calculated as $F = \dfrac{2PR}{(P+R)}$.[2] Being closely related F and G are both not influenced by imbalanced class distributions and thus generally applicable in our domain. However, we expect F to be the more important indicator, as it focuses more directly on one particular class, e.g. class B, which is consistent with typical demands in the field of aCRM related classification.

IV. SUPPORT VECTOR MACHINES FOR CLASSIFICATION

The original SVM can be characterized as a supervised learning algorithm capable of solving linear and non-linear classification problems. The main building blocks of SVM are structural risk minimization, non-linear optimization and duality as well as kernel induced features spaces, underlining the technique with an exact mathematical framework [18].

The main idea of support vector classification is to separate examples with a linear decision surface and maximize the margin of separation between the two different classes.
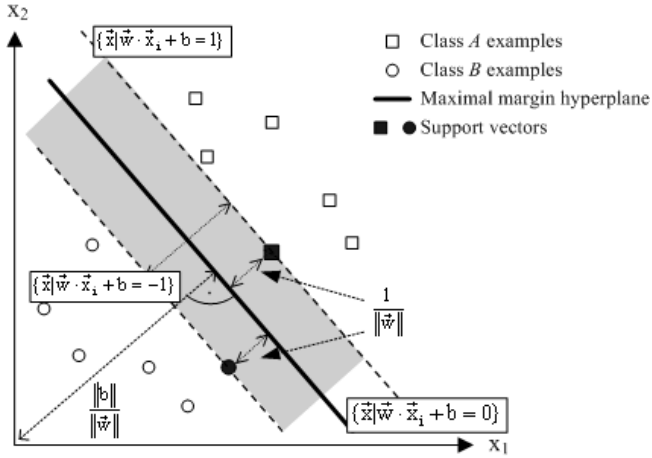

Fig. 1: Maximal margin hyperplane for discriminating between two classes [19]

The idea to construct a separating hyperplane with maximal margin leads to the well known soft-margin optimization problem [18].

$$min_{\vec{w},\xi,b} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{L}\xi_i$$

s.t.
$$k_i(\vec{w}\cdot\vec{x}_i + b) \geq 1 - \xi_i \quad \forall i = 1,...,L$$
$$\xi_i \geq 0 \quad \forall i = 1,...,L \tag{1}$$

where L denotes the number of training examples, $\vec{x}_i$ represents the attribute vector of example i, $k_i \in \{0,1\}$ is the class label of example i and C is a constant cost parameter, enabling the user to control the trade-off between learning error and model complexity, given by the margin of the separating hyperplane [3]. The slack variables ξ_i accounts for the fact, that the training data is not necessarily linearly separable, such that some examples will be misclassified by a linear discriminant function.

Data points closest to the maximal margin hyperplane, that is points satisfying $k_i(\vec{w}\cdot\vec{x}_i + b) = 0$, are called (bounded) support vectors as they define the position of the separating plane; see Fig. 1. Consequently, the solution of a support vector classifier depends only on a (possibly very) small number of training examples, the support vectors, and removing all other instances from the training set would leave the solution unchanged. From this understanding of a support vector we could expect SVM to be insensitive to imbalanced class distributions since there should always be a sufficient number of examples from each class to form a reasonable support vector set [11]. However, our experiment reveals that this assumption is not true.

Problem (1) forms the basis for SVM classification and an internal modification to account for imbalanced class distributions by means of asymmetric misclassification cost is straightforward. A simple revision of the objective function gives

$$min_{\vec{w},\xi,b} \frac{1}{2}\|\vec{w}\|^2 + C_+\sum_{k_i=1}\xi_i + C_-\sum_{k_i=0}\xi_i$$

s.t.
$$k_i(\vec{w}\cdot\vec{x}_i + b) \geq 1 - \xi_i \quad \forall i = 1,...,L$$
$$\xi_i \geq 0 \quad \forall i = 1,...,L \tag{2}$$

providing two independent cost parameters while leaving the overall algorithm almost unchanged. Formulation (2) is the one incorporated in the SVM solver LIBSVM [20] which we used for our study.

For constructing more general non-linear decision surfaces than hyperplanes, SVMs implement the idea to map the input vectors into a high-dimensional feature space Ψ via an a priori chosen non-linear mapping function $\Phi : X \to \Psi$. The construction of a separating hyperplane in this features space leads to a non-linear decision boundary in the original space; see Fig. 2. Expensive calculation of dot products $\Phi(\vec{x}_i)\cdot\Phi(\vec{x}_j)$ in a high-dimensional space can be avoided by introducing a kernel function $K(\vec{x}_i,\vec{x}_j) = \Phi(\vec{x}_i)\cdot\Phi(\vec{x}_j)$ [3].

Therewith, SVM enable a considerably easier parameterization when compared to other learning machines like for example multi-layer perceptron neural networks [21]. The only degrees of freedom are the selection of a kernel function together with corresponding kernel parameters and the choice of the cost parameter C or C_+ and C_-, respectively.

---

[2] Although not widely used, the geometric mean of P and R is a suitable performance metric as well.

input space
(two dimensions)

feature space
(three dimensions)

□ Class A examples
○ Class B examples

----- class boundary (nonlinear)

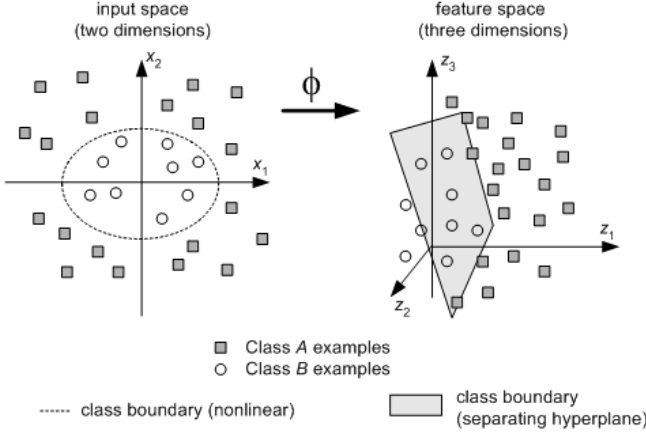▨ class boundary
(separating hyperplane)

Fig. 2. Non-linear -mapping from two-dimensional input space with non-linear class boundaries into a linear separable feature space

V. SIMULATION EXPERIMENT OF SUPPORT VECTOR SENSIBILITY TO CLASS IMBALANCES

A. Objective

A broader adoption of SVM in the field of aCRM related problems is just beginning and in order to become a major classification technique within this particularly difficult domain SVM has to proof empirically its capability of handling highly imbalanced data sets. For the SVM classifier the question if imbalances have to be adjusted and which method, e.g. internal or external approaches, is preferable has to our best knowledge not been answered by now as most research in this field is based on decision trees or artificial neural networks [6, 9, 11, 12]. Thus, we evaluate SVM's capabilities to address class imbalances internally in comparison to external balancing by resampling within an empirical, numerical study.

In our experiment, we consider the case of response optimization as a representative example for aCRM related classification. The goal of response optimization is to identify a subset of customers who exhibit a substantially higher probability of reacting to a certain offer than the average customer, based on experiences from past campaigns. Here, the cost of making an offer to a person who does not respond is typically small compared to the cost of not contacting a customer, who would otherwise have ordered an item. The imbalance is introduced as usually only a very small group of people who were contacted purchase a product.

B. Experimental setup

Our data is based on a mailing campaign which included 300,000 addresses and aims at selling an additional magazine subscription to customers who have already subscribed to at least one periodical. The response rate of this campaign was 1.3%, meaning that only 4019 customers showed a positive reaction.

In order to discriminate these economically relevant customers from all others, the data contains 50 numerical as well as categorical attributes, which provide demographic and transactional information about each customer. While numerical attribute value were scaled to the interval [-1;1] using a linear transformation we applied one-of-N remapping to account for discrete attributes [22].

Following, we randomly selected 100,000 records as a hold-out set to enable out of sample validation on unseen data. The remaining 200,000 customer records formed the training data and were used in five different training scenarios, as is described in Table 2 where the class label B denotes the group of customers, who responded to a previous campaign.

Experiment 1 consists of a randomly selected sub-sample of 10,000 records of the available training data. Here, it is left to the SVM to adjust the imbalance internally. Under-sampling leads to experiment 2 and 3 where all class B records of the training data base together with some randomly selected class A records were used, so that we obtain a class B to A ratio of 1:2 and 1:1, respectively.[3] Within the remaining two experiments over-sampling was used to achieve the same class ratios of 1:2 and 1:1. That is, the 2,693 class B records within the available training data where randomly duplicated until the respective target ratios between class B and class A records were reached.

TABLE 2
Setup for numerical evaluation of SVM's sensibility towards imbalanced class distributions

Experiment No.	training data partitioning				
	records class A		records class B		records total
	number	percent	number	percent	
1	9885	98,85	115	1,15	10000
2	5368	66,67	2963	33,33	8079
3	2693	50,00	2693	50,00	5368
4	9885	66,67	4942	33,33	14827
5	9885	50,00	9885	50,00	19770

Concerning SVM parameterization, we refused to use polynomial kernel functions as several pre-test revealed their computational inefficiency and incorporated linear and gaussian kernels instead [18]. It is common practice to use the same value for C_+ and C_- when class distributions are balanced and we will denote this as symmetric costing (SC). A correspondingly parameterized SVM will be called symmetric costing support vector machine (SC-SVM). We expect SC to provide competitive performance when class imbalances

---

[3] To ensure comparability, the class A records were fixed throughout all experiments. That is, experiment 2 and 3 used a randomly selected sub-sample of the class A records that were used in experiment 1, 4 and 5.

are externally adjusted through resampling and consequently included according classifiers in our study varying *log(C)* stepwise from -3 to 4. If on the other hand imbalances are not externally adjusted, as is done in experiment 1, we can hardly expect SC-SVM to deliver reasonable classification results. Therefore, we incorporated SVM with asymmetric costing (ASC-SVM) as well and evaluated 20 parameter settings for C_- in the range of 0.001 to 0.02 while leaving C_+ fixed at 1.

Since the kernel width σ can have a crucial impact on the classification ability of the gaussian SVM [23] we evaluated six different settings ($\sigma = \{0.05; 0.075; 0.1; 0.125; 0.3; 0.5\}$) for any cost parameterization.

Combining all parameter settings for the linear and the gaussian SVM, we obtain a total number of more than 130 classifiers which were evaluated for every experiment.

C. Results

Our study revealed that differences in classifier performance between individual experiments are not severe. However, we have to keep in mind that even a small difference can have a noticeable monetary impact in economical environments. The maximal observed performance within each experiment by means of F and G is given in Table 3.

TABLE 3
Maximal observed performance by means of F and G

Exp. No.	best observed results			
	F	rank	G	rank
1	**0,0306**	**1**	0,5126	3
2	0,0294	2	0,5207	2
3	0,0281	4	0,4631	5
4	0,0286	3	**0,5213**	**1**
5	0,0279	5	0,4966	4

The unadjusted experiment 1 delivered quite competitive results indicating that SVM is indeed capable of handling imbalanced data sets by assigning different cost parameters to each class. This is a promising result as re-sampling complicates the overall data mining process and is therefore time- and cost-consuming. Detailed results for experiment 1 are given in Fig. 3

Surprisingly, the linear SVM seems to dominate non-linear classifiers with gaussian kernel when G is used as performance indicator but this result is not confirmed by the probably more important F-measure. Regarding F we find that only a small range of C_- between 0.012 and 0.015 delivers utilizable classification results indicating the particular challenge of analyzing complex real world data sets.
For the adjusted data sets (experiment 2-5) ASC leads to naive classifiers where all instances were classified as belonging to class B. This is probably due to the considered ratio between C_+ and C_-. However, the idea of re-sampling is to

avoid internal balancing so that we compare different re-sampling techniques for the SC-SVM.[4]
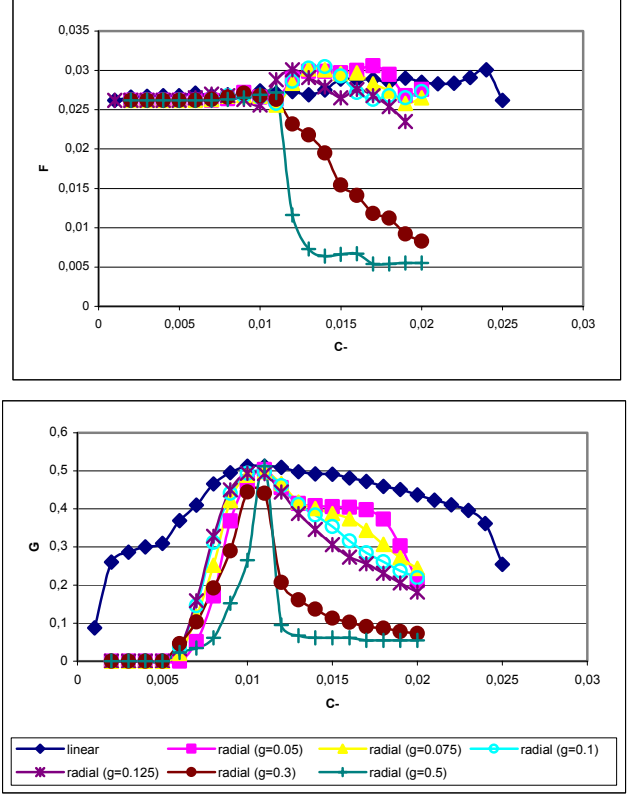

Fig. 3: Results for different linear and radial ASC-SVM in experiment 1 by means of F and G

Regarding the poor performance of experiment 1 SC is obviously inappropriate when class imbalances are not adjusted. This proofs that SVM is indeed sensitive to imbalanced class distributions and contrasts the results of [11]. Surly, this is due their univariate experiment design which is not representative for real world aCRM problems.

While for experiment 3 and 5 with completely balanced class distributions we can select very small values for the cost parameter *C* this leads to naïve classification in all remaining experiments with imbalanced class distributions. Considering the SVM objective function (1) a low value for *C* results in a classifier which focuses primary on margin maximization instead of accuracy. Hence, if data similarity and the risk of over-fitting is increased, e.g. through over-sampling, SVM naturally compensates this by enabling lower settings for *C* leading to robust classifiers with large margin of separation and improved generalization ability.

---

[4] SVM with linear kernel showed the same trend over all experiments on a slightly lower performance level and were therefore excluded for clarity. We report results for the radial SVM with $\sigma = 0.05$ as it consistently delivered superior performance.
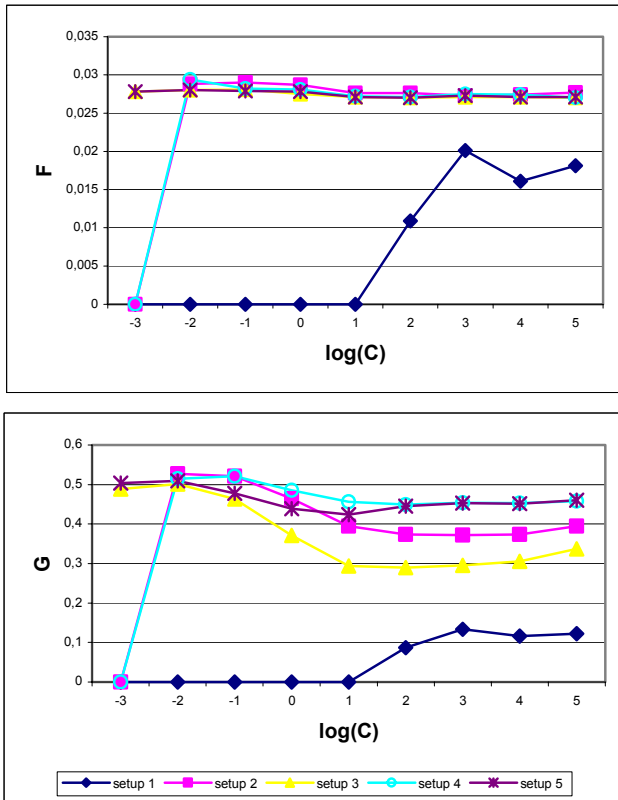
Fig. 4: Performance of radial SVM by means of F and G for all experiments

VI. CONCLUSION

We analyzed the problem of imbalanced class distributions in the field of aCRM related classification exemplifying the need to regard this issue during classification analysis by means of internal or external data adjustments and suitable performance metrics.

Our experiment revealed that the SVM classifier is able to account for class imbalances in an internal manner through according parameterization within the model selection stage. Consequently, the data pre-processing phase, preceding any data mining analysis, can be simplified significantly when the SVM classifier is used.

On the other hand, internal modifications are usually not reusable among different classification algorithms [12] and therefore complicate the comparison of different methods. If such a comparison is desirable, e.g. to determine a superior algorithm for a specific problem, it is wiser to account for imbalances externally through re-sampling. Our experiment revealed that SVM is robust towards re-sampling methods, working with under- and over-sampling alike. However, when applied in conjunction with under-sampling SVMs provide competitive results while using considerably less records leading to an increased computational efficiency.

We restricted our analysis to basic re-sampling techniques randomly downsizing the majority class and randomly upsizing the minority class, respectively. More elaborate ap-

proaches are proposed in [9, 10, 12] and the question if ASC-SVM is still competitive to re-sampling when such techniques are applied and if the potential gain in classification performance would justify the additional sampling effort under economical considerations needs further research.

We used F and G to measure classification performance in imprecise environments which is consistent with other research conduced in this field [6, 11, 12, 24]. Though both measures are not influenced by class distributions it is questionable if they are ideal for the field of aCRM where the minority class is generally of primary importance. Hence, it can be economically sensible to sacrifice precision in order to achieve a higher recall and the F-measure will give poor advice on classifier selection. The same argumentation holds for G so that it seems worthwhile to investigate the question of an economical performance metric for classification analysis in future research.

REFERENCES

[1]   M. Bruhn, *Relationship marketing : management of customer relationships*. Harlow [u.a.]: Financial Times Prentice Hall, 2002.

[2]   S. Lessmann, "Customer Relationship Management," *WISU - das Wirtschaftsstudium*, vol. 32, pp. 190-192, 2003.

[3]   V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.

[4]   A. Berson, S. Smith, and K. Thearling, *Building Data Mining Applications for CRM*. New York: McGraw Hill, 1999.

[5]   H. Hippner and K. D. Wilde, "Data Mining im CRM," in *Effektives Customer Relationship Management*, S. Helmke, M. Uebel, and W. Dangelmaier, Eds., 2 ed. Wiesbaden: Gabler, 2002, pp. 211-232.

[6]   N. V. Chawla, "C4.5 and Imbalanced Datasets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure," presented at ICML Workshop on Learning from Imbalanced Datasets II, Washington DC, 2003.

[7]   T. Fawcett and F. J. Provost, "Adaptive Fraud Detection," *Data Mining and Knowledge Discovery*, vol. 1, pp. 291-316, 1997.

[8]   M. Kubat, R. C. Holte, and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Radar Images," *Machine Learning*, vol. 30, pp. 195-215, 1998.

[9]   M. Kubat and S. Matwin, "Addressing the Curse of Imbalanced Training Sets: One-Sided Selection.," presented at Proceedings of the 14th International Conference on Machine Learning, ICML'97, Nashville, TN, U.S.A., 1997.

[10]   N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.

[11]   N. Japkowicz and S. Stephen, "The class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis*, vol. 6, pp. 429-450, 2002.

[12]   A. Estabrooks, T. Jo, and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalanced Data

Sets," *Computational Intelligence*, vol. 20, pp. 18-36, 2004.

[13]   S. Lawrence, I. Burns, A. D. Back, A. C. Tsoi, and C. L. Giles, "Neural Network Classification and Prior Class Probabilities," in *Neural Networks: Tricks of the Trade*, vol. 1524, *Lecture Notes in Computer Science*, G. B. Orr and K.-R. Müller, Eds. Heidelberg: Springer, 1998, pp. 299-313.

[14]   F. Provost and T. Fawcett, "Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions," presented at Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, 1997.

[15]   A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, pp. 1145-1159, 1997.

[16]   M. Kubat, R. C. Holte, and S. Matwin, "Learning when Negative Examples Abound," presented at Proceedings of the 9th European Conference on Machine Learning ECML'97, Prague, Czech Republic, 1997.

[17]   C. J. Van Rijsbergen, *Information retrieval*, 2d ed. London ; Boston: Butterworths, 1979.

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

[19]   C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, 1998.

[20]   C.-C. Chang and C.-J. Lin, "LIBSVM - A Library for Support Vector Machines," 2.6 ed: Software available at http://www.csie.ntu.edu.tw/\verb"~"cjlin/libsvm, 2001.

[21]   S. F. Crone, S. Lessmann, and R. Stahlbock, "Empirical Comparison and Evaluation of Classifier Performance for Data Mining in Customer Relationship Management," presented at Proceedings of the IEEE 2004 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2004.

[22]   D. Pyle, *Data preparation for data mining*. San Francisco, Calif.: Morgan Kaufmann Publishers, 1999.

[23]   S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Computation*, vol. 15, pp. 1667-1689, 2003.

[24]   F. Provost, T. Fawcett, and R. Kohavi, "The Case Against Accuracy Estimation for Comparing Induction Algorithms," presented at Proceedings of the Fifteenth International Conference on Machine Learning, San Francisco, 1998.