Improving customer retention management through cost-sensitive learning

Kristof Coussement
Department of Marketing, IESEG School of Management, Université Catholique de Lille, Lille, France

Received 27 March 2012
Revised 28 September 2012
17 December 2012
Accepted 18 January 2013

Abstract

Purpose – Retailers realize that customer churn detection is a critical success factor. However, no research study has taken into consideration that misclassifying a customer as a non-churner (i.e. predicting that (s)he will not leave the company, while in reality (s)he does) results in higher costs than predicting that a staying customer will churn. The aim of this paper is to examine the prediction performance of various cost-sensitive methodologies (direct minimum expected cost (DMECC), metacost, thresholding and weighting) that incorporate these different costs of misclassifying customers in predicting churn.

loyal

Design/methodology/approach – Cost-sensitive methodologies are benchmarked on six real-life churn datasets from the retail industry.

Findings – This article argues that total misclassification cost, as a churn prediction evaluation measure, is crucial as input for optimizing consumer decision making. The practical classification threshold of 0.5 for churn probabilities (i.e. when the churn probability is greater than 0.5, the customer is predicted as a churner, and otherwise as a non-churner) offers the worst performance. The provided managerial guidelines suggest when to use each cost-sensitive method, depending on churn levels and the cost level discrepancy between misclassifying churners versus non-churners.

Practical implications – This research emphasizes the importance of cost-sensitive learning to improve customer retention management in the retail context.

Originality/value – This article is the first to use the concept of misclassification costs in a churn prediction setting, and to offer recommendations about the circumstances in which marketing managers should use specific cost-sensitive methodologies.

Keywords Retailing, Relationship marketing, Customer retention, Database marketing

Paper type Research paper

1. Introduction

Intense competition and increased choices available to customers create new pressures on retailers to manage their customers in long-term relationships (Shaw *et al.*, 2001; Demoulin and Zidda, 2009). More and more retailers are pushed from a product-centric focus into the practice of a customer-centric strategy. Indeed, many companies are turning to customer relationship management (CRM) to better serve and facilitate closer relationships with customers (Noble and Phillips, 2004; Nagar and Rajan, 2005;

The authors thank the editors, Prof. Dr Nick Lee and Prof. Dr Malcolm Wright, and the anonymous reviewers for their suggestions and remarks, which led to improvements over previous versions of this work; the anonymous companies for their collaboration in the various academic projects; the PhD students who prepared these research datasets on which this study is based; and friendly reviewers for their fruitful comments on prior versions of this research.

European Journal of Marketing
Vol. 48 No. 3/4, 2014
pp. 477-495

0309-0566
DOI 10.1108/EJM-03-2012-0180

Torkzadeh *et al.*, 2006; Leenheer and Bijmolt, 2008; Sen and Sinha, 2011; Hung *et al.*, 2010; Maklan and Knox, 2009). Retailers acknowledge that one of the cornerstones of a solid CRM strategy is the existence of a loyalty (card) or reward program that collects customer data. In recent years, the availability of large volumes of data on customers has created new opportunities to leverage customer knowledge and gain competitive advantage. The growing popularity of customer retention policies results in increased interest in academia as well as in business (Magi, 2003; Keh and Lee, 2006; Corti⟩as *et al.*, 2008; Dowling and Uncles, 1997; Taylor and Neslin, 2005; Allaway *et al.*, 2006). A critical domain for improving companies' retention strategies involves customer churn prediction, which can have serious implications for companies' cumulative profit. For example, Clark (1997) reports that a 1 percent point decrease in churn rate of a major UK retail bank increased its total earnings from €392.2 million to €419.7 million – that is, by €27.5 million – over a time period of 25 years and a 6 percent discount rate.

This study presents a benchmark framework of different cost-sensitive algorithms that take into account different misclassification costs for customer churn predictions in a retail environment. Benchmarking is an important learning tool for organizations for a variety of applications (Vorhies and Morgan, 2005; Homburg *et al.*, 2012), including customer defection detection (Wright and Riebe, 2010; Verbeke *et al.*, 2011). This research study seeks to clarify which factors influence the performance of cost-sensitive learners in defection detection analysis.

The next section introduces customer retention literature and churn prediction methodologies, while highlighting the importance and relevance of incorporating different misclassification costs. Next a summary of frequently used performance measures in customer churn classification also invokes a discussion of their weaknesses related to cost asymmetries. Consequently, the classification technique and the non-sampling based cost-sensitive learners, namely, relabeling (with direct minimum expected cost classification and metacost), threshold adjusting and weighting, are explained. After the empirical setting is presented (i.e. datasets and practical threshold), the next section reports the findings, followed by the managerial implications and suggestions for further research in the final section.

2. Customer retention and churn prediction

Retailers acknowledge that customer retention management becomes increasingly important when markets become saturated and highly competitive (Keaveney and Parthasarathy, 2001; Ang and Buttle, 2006). For instance, the defection rate appears more often as a key marketing metric (Rust *et al.*, 2004). The drivers of successful customer retention management have been studied widely, mainly because of their impact on the customer lifetime value and thus on companies' bottom line profits (Bolton, 1998; Rust and Chung, 2006; Verhoef, 2003; Gupta *et al.*, 2004; Reichheld, 1996; Reichheld and Sasser, 1990; Wright and Riebe, 2010). Many studies seek to examine customer retention drivers, including inter-customer dynamics through social influence (Nitzan and Libai, 2011), multichannel relational communication (Godfrey *et al.*, 2011), satisfaction (Bolton, 1998; Oliver, 2010; Gustafsson *et al.*, 2005; Mittal and Kamakura, 2001; Bolton *et al.*, 2006), loyalty programs (Leenheer *et al.*, 2007; Lewis, 2004; Meyer-Waarden, 2007), customer usage patterns (Verhoef, 2003; Bolton and Lemon, 1999), brand equity (Stahl *et al.*, 2012) and customer characteristics (Cooil *et al.*, 2007).

Exploring the literature results in the fact that customer retention could have various managerial and financial implications. First, through customer retention, organizations can better serve their current customer base by building and maintaining relationships, rather than expending efforts to recruit new customers, which are often characterized by a high attrition rate (Reichheld and Sasser, 1990; Richards and Jones, 2008). Second, long-lasting customers spend more with companies, provide new referrals through positive word of mouth if satisfied, and are less costly to serve, because the company has more knowledge about them (Ganesh *et al.*, 2000). Third, long-term customers are less prone to competitive marketing actions. Fourth, losing customers leads to opportunity costs and missed possibilities to cross- or up-sell products, while also it increases the necessary costs to attract new customers.

The use of information technology and decreasing computer performance-to-cost ratios shed light on customer churn prediction as a popular instrument in the customer retention management toolbox (Neslin *et al.*, 2006; Lemmens and Croux, 2006). In this research, we define churn prediction as the identification of customers who have a high likelihood of completely ending the relationship with the firm, by means of data mining techniques. The idea is to improve customer retention rates by classifying customers as churners or non-churners, based on their predicted churn probability, and by targeting incentives to those customers at risk to persuade them to stay with the firm. Data mining techniques are able to explore and analyze huge amounts of available customer data to assist in the selection of customers at risk (Hung *et al.*, 2006; Dierkes *et al.*, 2011). Developments in database processing (Ziarko, 1991), data warehousing (Kelly, 1996), machine learning (Holsapple *et al.*, 1993; Shaw, 1993) and knowledge management (Amidon, 1998) have made data mining an attractive and useful tool to uncover hidden patterns in customer data.

From a methodological perspective, customer churn prediction is the process by which customers are predicted to belong to the churner or non-churner class, based on their historical profile information. Practically, a classification model is built on a set of customers for which historical profile information and target classes, churner or non-churner, are available (the training set). This classification model tries to link the customer profile information with the behavioral churning outcome. After the classification model is trained, it is evaluated on a set of customers that were not used during the model building phase (the test set). The result is a churn probability assigned to every customer in the test set, indicating the chance that he or she will end the relationship with the company (Au *et al.*, 2003; Coussement and Van den Poel, 2008b). Classifying these customers into churners or non-churners based on their probability to leave the company makes it possible to come up with a tailor-made retention program that focuses on the customers at risk.

However, prior research focused on churn prediction rarely considers the existence of different misclassification costs (Neslin *et al.*, 2006; Lemmens and Croux, 2006; Risselada *et al.*, 2010, Coussement and Van den Poel, 2008b; Verbeke *et al.*, 2011). It assumes that classifying a churner as a non-churner has the same impact on retention profitability as categorizing a non-churner as a churner. This assumption has several complications, because in real-world decision-making situations, the assumption of equal misclassification costs, the default operating mode for many classifiers, is most likely violated (Viaene and Dedene, 2005). Medical diagnosis is a classic case. In this case, failing to detect a disease may have fatal consequences, whereas diagnosing a disease for a patient that does not actually have it may be less serious. A comparable situation arises for customer churn detection. Here, failing to identify a churner (false negative prediction) has larger managerial and financial consequences than categorizing a non-churner as a churner (false positive prediction). In the case of false negative predictions, the company does not even have the opportunity to induce customers to stay by sending them marketing incentives, and therefore, it loses the customers and forgoes profits, opportunities to cross-sell and up-sell and so forth. For false positive classifications, there is only a redundant pampering cost. In summary, many classifiers are designed to optimise classification performance without taking into account the different costs of misclassification (Domingos, 1999).

3. Misclassification cost as a performance measure

Although the error rate, or the area under the receiver operating characteristics curve (AUC), and the top-decile lift are often used to evaluate the performance of churn prediction models (Lemmens and Croux, 2006; Coussement and Van den Poel, 2008a; Risselada *et al.*, 2010), they ignore the presence of asymmetry in misclassification costs. The error rate, or one minus the accuracy, is the most common performance measure used in churn prediction applications (Coussement and Van den Poel, 2008a). Practically, all customers are assigned a churn probability by the classification algorithm, and they are ranked from most likely to churn to least likely to churn. All customers with a churn probability above a certain threshold are classified as churners, while all others are seen as stayers. The error rate computes the ratio of incorrectly classified customers to the total number of customers to be classified. Despite its popularity, using the error rate as a performance measure is subject to criticism (Provost and Fawcett, 2001), because it assumes equal misclassification costs and relatively balanced distributions between churners and non-churners. In a churn prediction context, there are always more non-churners than churners present in the dataset. Hence, wrong predictions for the churners are very costly; however, they do not enormously influence the error rate. A churn prediction model minimizing the error rate usually results in a useless model, because it predicts customers as non-churners. For example, suppose one encounters a churn prediction setting with 5 percent churners. A classification algorithm that always predicts customers as being non-churners – and thus fails to identify the real churners – provides an error rate of only 5 percent. This seems an excellent prediction performance. However, such a model neglects to true purpose of the churn prediction model, that is, the identification of churning customers.

Several authors (e.g. Bradley, 1997) have come up with the concept of AUC to evaluate the performance of binary classification models. Intuitively, the AUC is the estimated probability that a randomly chosen churner has a higher predicted churn probability than a randomly selected non-churner. For instance, suppose that a churn model has an AUC of 0.70. It means that if we were to randomly pick a real churner and a non-churner from the churn dataset, 70 percent of the time, the churner will have a higher predicted churn probability assigned by the classifier than the non-churner. The AUC is a ranking measure that evaluates the churn prediction model based on the ranking of the customers using the predicted churn probability and real churning behavior. The values of the AUC measure range from 0.5, the purely random ranking, to 1, the perfect ranking. However, the AUC measure suffers from two main drawbacks from a decision-making perspective, that is, for trying to classify customers into churning or staying based on the predicted churn probabilities. First, it does not show the marketing manager how to set the classification threshold on the predicted churn probabilities to optimize decision making (Margineantu, 2000; Margineantu, 2001). Second, the accurate ranking of customers that underlies the AUC measure semantics is not equivalent to the accurate predicted churn probability estimation task (Viaene and Dedene, 2005). For making optimal decisions about whether a customer is considered as a churner or a non-churner, a model that perfectly ranks the customers based on the predicted churn probabilities but fails to estimate those churn probabilities accurately is insufficient. For more technical details about the AUC measure, see Bradley (1997).

Another prediction performance measure that is popular within the marketing community is the top-decile lift (Lemmens and Croux, 2006; Neslin *et al.*, 2006; Risselada *et al.*, 2010). It measures the increase in density of the real churners within the 10 percent of customers predicted as most likely to churn. Practically, all customers are sorted from most likely to churn to least likely to churn using the predicted churn probabilities. Among the top 10 percent of customers most likely to churn, the proportion of real churners is compared with the proportion of real churners in the total database. This increase in density is called the top-decile lift. From an optimal decision-making perspective, several problems arise. First, the threshold (i.e. 10 percent) for classifying a customer as a churner is arbitrarily chosen and neglects the minimization of the total misclassification cost. Second, by focusing on maximizing the relative percentage of churners in the top 10 percent, the top-decile lift does not take into account the misclassification of customers.

In summary, the error rate, the AUC and the top-decile lift do not lead to an optimal evaluation of the classification algorithm when different misclassification costs occur. The extra element needed for optimal decision making is an evaluation in terms of total misclassification cost. Improving marketing decision making by minimizing the overall misclassification cost in a churn prediction context is recommended and supported in many practical situations. For instance, suppose that customer x enters a bank office. Based on the customer's historical information, the company's churn model indicates that client x has a probability of 70 percent of churning within the next six months. To take the appropriate churn prevention action, the company needs an indication whether customer x will be classified as leaving or staying. Marketing decision makers must make an optimal decision in correspondence with the minimization of the total misclassification cost.

Typically, the confusion/cost matrix C has the following structure in a churn prediction context (see Table I).

Customers are predicted by the churn prediction model to be churners or non-churners, such that customers with a higher (lower) churn probability than the threshold are classified as (non-) churners. Depending on their real behavior of staying or leaving the firm, four possibilities could occur: TP (true positives) are the number of churners that are correctly identified, FP (false positives) are the number of non-churners

Table I. Confusion/cost matrix for customer churn classification

		Actual behavioral outcome	
		Churner	Non-churner
Predicted behavioral outcome	Churner	$TP\ c_{11}$	$FP\ c_{10}$
	Non-churner	$FN\ c_{01}$	$TN\ c_{00}$

that are classified as churners, FN (false negatives) are the number of churners that are identified as non-churners and TN (true negatives) are the number of non-churners that are classified as non-churners. By following the convention of recent papers on cost sensitivity, the cost matrix rows correspond to alternative predicted classes, while the columns correspond to actual classes (i.e. row/column = t/j = predicted/actual class). The cost for a TP is denoted as c_{11}, the cost for a TN is denoted as c_{00}, the cost for a FP is denoted as c_{10} and the cost for a FN is denoted as c_{01}. Note that if *c..* is larger than 0, it represents an actual cost, whereas if *c..* is smaller than 0, it represents a benefit.

Cost matrices often implicitly meet two reasonableness conditions formulated by Elkan (2001). The first reasonableness condition implies that neither row in the matrix dominates the other. A simple, intuitive criterion exists for when this happens: Say that row 1 dominates row 2 in a cost matrix *C* in Table I if for all $j \in \{0, 1\}, c_{1j} \geq c_{0j}$. In this case, the cost of predicting non-churner is no greater than the cost of predicting churner, regardless of what the true class j is. So, it is optimal never to predict churner. The second reasonableness condition implies that the cost of labeling an instance correctly is always lower than the cost of labeling an instance incorrectly. Finally, the total misclassification cost is calculated as follows:

$$\text{Total misclassification cost} =_{c11} \text{TP} +_{c10} \text{FP} +_{c00} \text{TN} +_{c01} \text{FN} \tag{1}$$

From an optimal decision making perspective, one wants to set the threshold on the predicted churn probabilities so that the total misclassification cost is minimized.

4. Classification technique

In cost-sensitive learning, Domingos (1999) suggests using a classifier that is well-suited to the domain. As such, logistic regression is used as the classifier throughout this research paper, because Neslin *et al.* (2006) state that it is a popular classification technique in traditional marketing applications like customer churn prediction, and it is widely used in different business settings (Akinci *et al.*, 2007). Finally, it is a simple technique (Bucklin and Gupta, 1992), providing quick and robust results.

Logistic regression is able to estimate the churn probability $p(j = 1|x)$ via:

$$\ln\left(\frac{p(x|j = 1)}{p(x|j = 0)}\right) = b + w^t x \tag{2}$$

where $w \in \Re^n$ is the coefficient vector, and $b \in \Re$ is the intercept. As such, the churn probability $p(j = 1|x)$ is obtained from Equation (2) by:

$$p(j = 1|x) = \frac{\exp(b' + w^t x)}{1 + \exp(b' + w^t x)} \tag{3}$$

where:

$$b' = b + \ln\left(\frac{p(j = 1)}{p(j = 0)}\right),$$

$p(j = 1)$ and $p(j = 0)$ as the class priors. The class membership probabilities are used to obtain the maximum likelihood estimates for w and b'.

5. Non-sampling based cost-sensitive learning techniques

In this section, the most popular cost-sensitive learning techniques are revisited. The idea is to optimize decision making in a customer churn prediction context by taking into account that misclassifying a churner has greater consequences than misclassifying a non-churner. Instead of building a churn prediction model that delivers a good fit to the data, cost-sensitive learners have as their objective the minimization of the total misclassification cost. Two categories of cost-sensitive learners exist in literature: non-sampling and sampling based (Sheng and Ling, 2006). In contrast with non-sampling based cost-sensitive learners, sampling-based cost-sensitive learners increase the proportion of churners in the training dataset in an attempt to improve churn prediction performance by duplicating churning customers (over-sampling) or randomly deleting non-churners (under-sampling). Sampling-based cost-sensitive learners have several drawbacks though. First, they distort the distribution of customers, which may affect the performance of some classification algorithms. Second, they reduce the data available for training, if sampling is carried out by under-sampling. Third, over-sampling drastically increases the learning time of classifiers. Therefore, we focus on the non-sampling based cost-sensitive learning techniques, which do not touch the original training dataset. For clarity, the term cost-sensitive refers to non-sampling based cost-sensitive in the remainder of the text.

Cost-sensitive learners are divided into three categories: relabeling, threshold adjusting and weighting (Sheng and Ling, 2006). The first two categories are categorized as wrappers, or techniques that could be wrapped around any existing classification technique to make it cost sensitive. These methods make any classifier minimize total misclassification cost rather than minimize the error rate. The latter category, weighting, adjusts the internal classifier's unique fitting function so that the weights assigned to customers, which depend on the class label, are incorporated in the training process.

Relabeling is based on reassigning the target class labels, churner or non-churner, of the customers by applying the direct minimum expected cost criterion that assigns a customer to the target class with the lowest misclassification cost. Relabeling can be done in the post-training phase, using a method called direct minimum expected cost classification (Duda and Hart, 2001), or in the pre-training phase, as proposed by Domingos (1999), using the metacost algorithm. Threshold adjusting searches for the best churn probability on the training set as a threshold for optimal future decision making in terms of misclassification cost on the test set (Sheng and Ling, 2006). Weighting assigns different weights to customers depending on the target class. The purpose is to favor the classifier to the class with the highest weight/cost.

5.1 Relabeling: direct minimum expected cost classification (DMECC)

Many classification techniques produce predicted churn probabilities that could be used to classify customers into churners or non-churners. Viaene and Dedene (2005) revisited DMECC as a valuable tool to improve decision making when misclassification costs are different. In a churn classification context, the optimal Bayes decision-making criterion dictates that a customer with a predictor vector $x \in \Re^n$ should be assigned to the target class $t \varepsilon \{0,1\}$ (with 0 non-churner and 1 churner) associated with the minimum expected cost (Duda and Hart, 2001). The optimal Bayes' decision criterion for a customer x is the class t where:

$$\underset{t \in \{0,1\}}{\arg\min} \sum_{j=0}^{1} p(j|x) c_{tj} \tag{4}$$

where $p(j|x)$ is the conditional probability of class j given predictor vector x, and c_{tj} is the cost of classifying a customer with predictor vector x and actual class j as class t. If $t = j$, then the prediction is correct, while if $t \neq j$, the prediction is incorrect.

Given the reasonableness conditions of the cost matrix C and for a churn classification setting, the prediction for the churner class is optimal if the expected cost of this prediction is less than the expected cost of predicting the non-churner class, that is, if:

$$p(j = 0|x)c_{10} + p(j = 1|x)c_{11} < p(j = 1|x)c_{01} + p(j = 0|x)c_{00}, \tag{5}$$

which is equivalent to:

$$(1 - p)c_{10} + pc_{11} < pc_{01} + (1 - p)c_{00}, \tag{6}$$

with $p = \mathrm{p}(j = 1|x)$. If this inequality is an equality, then predicting a customer as either a *churner* or a *non-churner* is optimal. The threshold for making optimal decisions is p^*, such that:

$$(1 - p^*)c_{10} + p^*c_{11} = p^*c_{01} + (1 - p^*)c_{00}. \tag{7}$$

Rearranging the equation for p^* leads to the following solution:

$$p^* = \frac{c_{10} - c_{00}}{c_{10} - c_{11} + c_{01} - c_{11}} \tag{8}$$

In other words, the classification rule that assigns the churner class is:

$$p(j = 1|x) > \frac{c_{10} - c_{00}}{c_{10} - c_{11} + c_{01} - c_{11}}, \tag{9}$$

and the instance is assigned to the non-churner class otherwise.

The DMECC costing method categorizes customers into churners or non-churners by applying the direct minimum expected cost criterion on the predicted churn probabilities $p(j = 1|x)$ using Equation (9). The advantages of using this method are that any classification technique that produces churn probabilities $p(j = 1|x)$ can make use of Equation (9) to determine whether the optimal class is churner or non-churner; and the classification model does not need to be retrained when the structure of cost matrix C changes, because relabeling is only introduced in the post-training phase, after the classification model is trained and the churn probabilities are obtained.

5.2 Relabeling: metacost

Metacost is aimed at making a churn prediction classifier cost-sensitive by manipulating the churn labels on the training data before the classifier is built (Domingos, 1999; Viaene and Dedene, 2005). It is based on wrapping a 'meta-training stage around the classifier that optimally relabels customers as churner or non-churner, in such a way that the classifier effectively minimizes the total misclassification cost. In other words, metacost treats the underlying classifier as a black box, requiring no knowledge of its functioning or changes to it.

Methodologically speaking, Domingos (1999) estimates the churn probability $p(j = 1|x)$ by using a variant of Breiman's (1996) bagging (or bootstrap aggregation). Multiple bootstrap replicates of the training set are formed and a classifier is trained on each of the bootstrap samples. The final $p(j = 1|x)$ for a customer is obtained by taking the unweighted average of its churn probabilities given by the bootstrap models. Using the optimal Bayes' decision criterion as formulated by Equation (9), each customer in the training set is relabeled with its optimal target class, churner or non-churner. In a final stage, the classifier is built on the relabeled training set to obtain the final churn prediction model, which is then applied to the test set. More technical information is available in Domingos (1999).

5.3 Threshold adjusting

Threshold adjusting (Sheng and Ling, 2006) finds the optimal predicted churn probability threshold in the training set by minimizing the total misclassification cost. It uses this churn probability as the optimal threshold to categorize customers in the test set. A test instance with predicted churn probability above or equal to this optimal cut-off point is predicted as a churner; otherwise it is considered a non-churner. Practically, the total misclassification cost for the training data is calculated for each possible value on the predicted churn probabilities. The churn probability that yields the lowest total misclassification cost is retained and used as the optimal threshold to classify the customers in the test set into churners or non-churners.

5.4 Weighting

Weighting assigns different weights to customers in the training set according to their target class (Ting, 2002). As such, the classification algorithm could take into account these weights during training, and it thus favors the class with the highest weight. Because logistic regression is used as the classification technique within this research paper, a weighted logistic regression is applied. Elkan (2001) proposes to weight the majority class or the non-churner group by $\frac{c_{10}}{c_{01}}$. In other words, the weights assigned to the training data resemble the cost ratios between the false positive and the false negative classifications. The weights are linearly incorporated into the calculation of the likelihood function of the logistic regression. The likelihood function ℓ for a given beta vector β and n training points is given by:

$$\ell(\beta) = \prod_{i=1}^{n} w_i \{\pi(x_i)\}^{y_i} \{1 - \pi(x_i)\}^{1-y_i} \qquad (10)$$

where w_i is the weight for data point i, $\pi(x_i)$ is the conditional probability that y is equal to 1 for a given x_i or $p(y = 1|x_i)$, and $y_i = 1$ for churners and $y_i = 0$ otherwise.

6. Empirical setting

6.1 Datasets

In this study, six real-life datasets from large retail companies are used to analyze the performance of the cost-sensitive learners in predicting complete defection. All datasets contain variations of traditionally-used customer information like transactional, socio-demographical and client/company interaction information, to differentiate churners from non-churners. Table II gives a description of the different datasets.

Moreover, Table III gives an overview of the different datasets in terms of industry, number of customers, and percentage of churners (churn incidence).

The datasets are collected from retailers in a rich variety of business settings: two information media providers, two financial institutions and two traditional store retailers (one do-it-yourself store and one food supermarket). These contexts are well studied in retail literature (Koukova *et al.*, 2008; Demoulin and Zidda, 2009; Farquhar and Panther, 2008), and consequently, these datasets represent an ideal test bed to benchmark the different algorithms in this research paper.

In this study, experiments are conducted using the five cost models with $c_{00} = c_{11} = 0$, $c_{10} = 1$ and $c_{01} = r$ where r is set alternately to 2, 3, 5, 10 and 20. Note that the absolute difference in the values for c_{10} and c_{01} is irrelevant for algorithm comparison purposes, because only their ratio 1:r is significant.

To create a reliable and valid testing environment, the datasets are randomly divided into a training set and a test set following the setup of other large-scale benchmark studies in the data mining field (Lessmann *et al.*, 2008; Baesens *et al.*, 2003), including churn prediction (Verbeke *et al.*, 2011). The training set (70 percent of the data) is used to learn the classification model, while the test set (the remaining 30 percent of the data) is used to validate the model (Lemmens and Croux, 2006). Measuring the performance on separate test data is necessary because that avoids (statistical) relations that predict idiosyncratic characteristics of the training data that would not hold up in the real world (Blattberg *et al.*, 2008). Both datasets contain a proportion of churners that is representative for the true population, to approximate a real-life situation. The total misclassification cost on the test set is used as performance measure to evaluate the different decision-making alternatives (see Equation 1).

6.2 Practical threshold

The threshold on the predicted churn probabilities in many practical decision-making situations is based on the arbitrary probability of 0.5. Many marketing managers

Table II.
Description of the churn prediction datasets

Dataset	Description
Information media provider	Defection of customers after subscription period
Information media provider	Defection of customers after subscription period
Food supermarket	Defection of customers on a five-month period
Financial institution	Defection of current account holders on a three-month period
Do-it-yourself company	Defection of customers on a four-month period
Financial institution	Defection of current account holders on a 12-month period

Table III.
Descriptive statistics of the churn datasets

Dataset	Number of customers	Churn incidence (%)
Information media provider	134,084	11.95
Information media provider	143,198	13.07
Food supermarket	100,000	30.86
Financial institution	117,808	6.29
Do-it-yourself company	32,371	25.14
Financial institution	102,279	5.98

implicitly decide to classify a customer as a churner when the churn probability exceeds 0.5. Moreover, standard classification algorithms are often designed to minimize error-rate, while their classification decisions are based on the probability threshold of 0.5 (Elkan, 2001). This practically oriented threshold, hereafter abbreviated as P50, is benchmarked to other, more sophisticated cost-sensitive learners in terms of total misclassification cost.

7. Findings

This section reports the findings of the benchmark study for the different cost-sensitive learners.

7.1 General results

The empirical results are summarized in Table IV. An entry *w/l* means that the approach at the corresponding row wins *w* runs and loses *l* runs compared to the approach at the corresponding column. Thirty runs of five cost ratios (that is 1:2, 1:3, 1:5, 1:10 and 1:20) on each of the six datasets are available to compare the different algorithms. For instance, the first cell in the matrix compares P50 with metacost and shows a figure '1/29, meaning that P50 beats metacost once, while P50 loses 29 times to metacost.

Table IV clearly shows that P50 always performs worse than the other cost-sensitive approaches. Incorporating unequal misclassification costs and classifying the customers based on the arbitrary threshold of 0.50 results in a poor classification performance in terms of total misclassification cost. Moreover, incorporating costs/weights into the classification algorithm itself (weighting) always performs better than the relabeling techniques (DMECC and metacost) and threshold adjusting. In contrast, thresholding performs worse than the relabeling approaches (DMECC and metacost). Between the relabeling approaches, there is a slight advantage for metacost over DMECC. A translation of the mutual challenges to a ranking, based on the wins/losses matrix, is found in Figure 1.

Figure 1 shows the rank of the different approaches based on the wins/losses; the lower the rank, the better the approach.

7.2 Sensitivity to cost ratio and churn incidence

To evaluate the sensitivity of the different approaches in terms of the cost ratio, two categories are created: the low cost ratio category contains cost ratios 1:2, 1:3 and 1:5, while the cost ratios 1:10 and 1:20 are categorized under the high cost ratio category. To evaluate how the cost-sensitive learners perform in circumstances with different

	Metacost	Thresholding	Weighting	DMECC
P50	1/29	1/29	1/29	1/29
Metacost		18/12	13/17	16/14
Thresholding			11/19	12/18
Weighting				17/13

Table IV. Summary of the empirical findings in terms of wins/losses

Low	Weighting ↔ Metacost ↔ DMECC ↔ Thresholding ↔ P50	High

Figure 1. Ranking of the approaches based on wins/losses

churn levels, the datasets are divided based on their churn incidence into low, medium or high. A dataset belongs to the low churn level when the churn incidence is less than 10 percent (dataset 4 and 6), the medium churn level when the churn incidence lies between 10 and 20 percent (dataset 1 and 2) and the high churn level when the churn incidence is higher than 20 percent (dataset 3 and 5).

Table V represents the results using the single best criterion. In other words, the figures in Table V represent the number of times that an approach is the winner, because it has the lowest misclassification cost in a particular setting. For instance, a figure '8 in the matrix means that the cost-sensitive learner has eight times the lowest misclassification cost. The evaluation is done by aggregating the results Overall and splitting by Cost ratio (low-high), Churn level (low-medium-high) and Cost ratio (low-high) x Churn level (low-medium-high).

Table V indicates that overall weighting is the winner (12 out of 30), followed by metacost and thresholding (8 wins each out of 30), whereas P50 only wins once over all the runs. These results are in line with the ranks given in Figure 1. The performance of DMECC seems contradictory with the ranking in Figure 1, in that it is only once the winner; this discussion is postponed to the end of this section.

Evaluating the results also should take into account the different cost ratios (see Table V). Weighting thus outperforms the other techniques in the high cost ratio group. It performs best in nine out of 12 runs. However, metacost and thresholding are the preferred techniques for the low cost ratios: Metacost wins eight and thresholding wins six out of the eighteen runs in this category.

The evaluation taking into account the impact of the different churn levels shows that metacost performs better when the churn level ranges from medium to high, while thresholding performs well in cases with a low to medium churn incidence. Weighting performs similarly over the different churn levels, and competitively with metacost and thresholding.

If we cross-tabulate the cost ratios with the different churn levels, several interesting conclusions emerge. In the case of high cost ratios, weighting is the absolute winner. It is the best in three out of four runs for each churn level. When cost ratios are low, there is a dominance of thresholding and metacost. Thresholding achieves good performance at low and medium churn levels, while metacost offers good performance when the churn level is medium or high.

The DMECC is only the absolute winner once across all 30 runs according to the results in Table V. This result seems contradictory with Figure 1, which ranks DMECC third on the basis of mutual comparisons of the wins/losses. Figure 2 helps clarify the results, by depicting the absolute ranking of the cost-sensitive learners for the different datasets. For every cost ratio (1:2, 1:3, 1:5, 1:10 and 1:20) of a given dataset, a ranking of the different techniques can be based on the misclassification cost. The cost ratios shape the corners of a pentagonal web; the ranks are represented by the threads of the web. A rank 1 means that for the given cost ratio, this cost-sensitive learner has the lowest misclassification cost and is the best approach, whereas a rank 5 indicates that the approach has the highest misclassification cost compared with the other techniques.

Figure 2 shows that DMECC is the most robust technique in terms of cost ratio and churn incidence. The ranking of DMECC (orange line) in all the datasets is very close to the best technique, but it is the best approach only once (in dataset 5, churn level high, cost ratio 1:10). The DMECC gives robust performance over the various churn levels

	Overall	Cost ratio Low	Cost ratio High	Churn level Low	Churn level Medium	Churn level High	Cost ratio Low, Churn level Low	Cost ratio Low, Churn level Medium	Cost ratio Low, Churn level High	Cost ratio High, Churn level Low	Cost ratio High, Churn level Medium	Cost ratio High, Churn level High
Metacost	8	8	0	1	3	4	1	3	4	0	0	0
Weighting	12	3	9	5	3	4	2	0	1	3	3	3
Thresholding	8	6	2	3	4	1	2	3	1	1	1	0
DMECC	1	0	1	0	0	1	0	0	0	0	0	1
P50	1	1	0	1	0	0	1	0	0	0	0	0
Total	30	18	12	10	10	10	6	6	6	4	4	4

Table V.
Overview of the single best approach

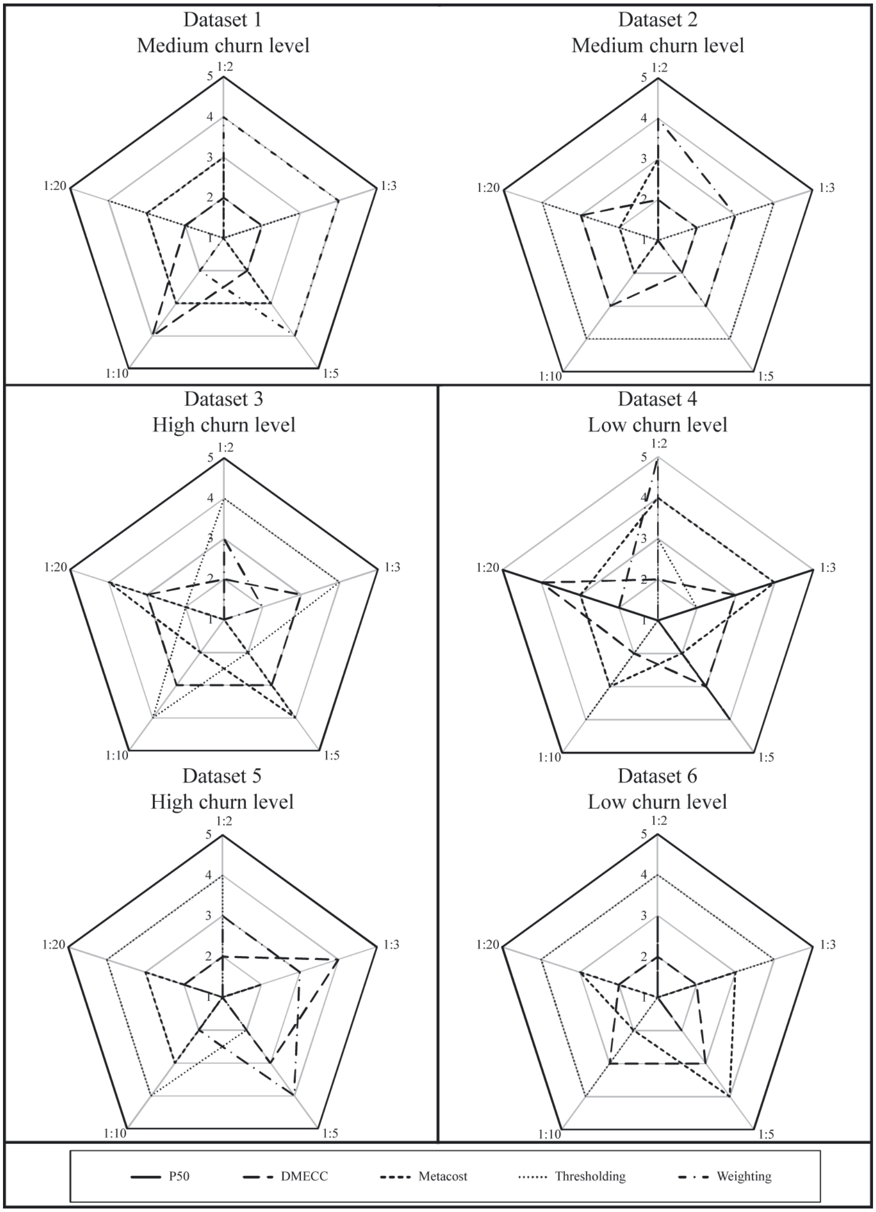

Figure 2.
Absolute ranking of the cost-sensitive learners per dataset (with rank 1 = lowest to rank 5 = highest misclassification cost)

and cost ratios. In the end, Figure 2 offers the same conclusions for weighting, metacost, thresholding and P50 as provided at the beginning of this section. For instance, P50 (blue line) performs worse than the other techniques and ranks fifth nearly every time. Weighting (pink line) dominates when cost ratios are high, and the pink line is then situated closer to the center.

8. Conclusions

This study focuses on comparing various cost-sensitive learners that incorporate different misclassification costs for churn prediction. It thus delivers insights to churn prediction literature. First, it motivates marketing managers to use the total misclassification cost as an alternative measure for evaluating churn prediction frameworks. Compared with performance measures like error rate, the AUC or top-decile lift, as are often used in churn prediction papers (Neslin *et al.*, 2006; Lemmens and Croux, 2006; Risselada *et al.*, 2010), the total misclassification cost does accommodate different misclassification costs. Yet such an accommodation is inevitable and of crucial importance for optimal churn detection, because classifying a churner as a non-churner has a significantly more negative impact on the company's retention policy than classifying a non-churner as a churner. In the former case, the company loses direct contact with the client, such that it cannot sell additional products or engage its customer contact strategy, whereas in the latter case, only the 'pampering cost is lost. Second, this article demonstrates that marketing managers in the retail sector should acknowledge that the often used, practically oriented classification threshold of 0.5 (P50) for predicted churn probabilities is not a viable strategy when misclassification cost asymmetries are considered. They should shift instead toward DMECC as the most robust, cost-sensitive approach in any given situation. Third, this study suggests managerial recommendations for optimizing the decision-making process in a customer churn prediction context, considering the cost ratio and churn incidence (see Table VI).

When cost ratios are high, weighting is the most appropriate technique for all churn levels. When cost ratios are low, thresholding tends to perform well in situations with low to medium churn levels, whereas metacost performs well when the churn level ranges from medium to high. Moreover, using the guidelines presented in Table VI, it is possible to optimize the decision-making process in a customer churn context.

This research contributes to churn literature, but it also leaves several pathways for further research. The current study uses logistic regression as the classification technique. However, other classification techniques might be used and benchmarked against logistic regression. Moreover, the focus of this research paper is on cost-sensitive learning techniques. Further research should add and compare other sampling-based techniques, including over-sampling and under-sampling. This research paper takes only single cost models into consideration. Additional experiments might contrast the performance of these techniques with Adacost, a variant of Adaboost that captures different misclassification costs in an ensemble classification environment (Fan *et al.*, 1999). Finally, the concept of this research paper could be extended to other classification contexts in marketing, such as response modeling, next-product-to-buy modeling and so forth.

<table>
<tr><th colspan="2"></th><th colspan="3"></th><th colspan="3">Cost ratio
Low</th><th colspan="3">Cost ratio
High</th></tr>
<tr><th colspan="2">Cost ratio</th><th colspan="3">Churn level</th><th colspan="3">Churn level</th><th colspan="3">Churn level</th></tr>
<tr><th>Low</th><th>High</th><th>Low</th><th>Medium</th><th>High</th><th>Low</th><th>Medium</th><th>High</th><th>Low</th><th>Medium</th><th>High</th></tr>
<tr><td>Metacost</td><td>Weighting</td><td></td><td colspan="2">Metacost</td><td></td><td colspan="2">Metacost</td><td colspan="3">Weighting</td></tr>
<tr><td>Thresholding</td><td></td><td></td><td>Weighting</td><td></td><td colspan="2">Thresholding</td><td></td><td colspan="3"></td></tr>
<tr><td></td><td></td><td colspan="2">Thresholding</td><td></td><td></td><td></td><td></td><td colspan="3"></td></tr>
</table>

Table VI. Managerial recommendations

References

Akinci, S., Kaynak, E., Atilgan, E. and Aksoy, A. (2007), "Where does the logistic regression analysis stand in marketing literature? A comparison of the market positioning of prominent marketing journals", *European Journal of Marketing*, Vol. 41 Nos 5-6, pp. 537-567.

Allaway, A.W., Gooner, R.M., Berkowitz, D. and Davis, L. (2006), "Deriving and exploring behavior segments within a retail loyalty card program", *European Journal of Marketing*, Vol. 40 Nos 11-12, pp. 1317-1339.

Amidon, D.M. (1998), "Blueprint for 21st century innovation management", *Journal of Knowledge Management*, Vol. 2 No. 1, pp. 23-31.

Ang, L. and Buttle, F. (2006), "Customer retention management processes: a quantitative study", *European Journal of Marketing*, Vol. 40 Nos 1-2, pp. 83-99.

Au, W.H., Chan, K.C.C. and Yao, X. (2003), "A novel evolutionary data mining algorithm with applications to churn prediction", *IEEE Transactions on Evolutionary Computation*, Vol. 7 No. 6, pp. 532-545.

Baesens, B., Gestel, T.V., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen, J. (2003), "Benchmarking state-of-the-art classification algorithms for credit scoring", *The Journal of the Operational Research Society*, Vol. 54 No. 6, pp. 627-635.

Blattberg, R.C., Kim, B.-D. and Neslin, S.A. (2008), *Database Marketing: Analyzing and Managing Customers*, Springer, New York, NY.

Bolton, R.N. (1998), "A dynamic model of the duration of the customer's relationship with a continuous service provider: the role of satisfaction", *Marketing Science*, Vol. 17 No. 1, pp. 45-65.

Bolton, R.N. and Lemon, K.N. (1999), "A dynamic model of customers usage of services: usage as an antecedent and consequence of satisfaction", *Journal of Marketing Research*, Vol. 36 No. 2, pp. 171-186.

Bolton, R.N., Lemon, K.N. and Bramlett, M.D. (2006), "The effect of service experiences over time on a supplier's retention of business customers", *Management Science*, Vol. 52 No. 12, pp. 1811-1823.

Bradley, A.P. (1997), "The use of the area under the roc curve in the evaluation of machine learning algorithms", *Pattern Recognition*, Vol. 30 No. 7, pp. 1145-1159.

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

Bucklin, R.E. and Gupta, S. (1992), "Brand choice, purchase incidence, and segmentation – an integrated modeling approach", *Journal of Marketing Research*, Vol. 29 No. 2, pp. 201-215.

Clark, M. (1997), "Modelling the impact of customer-employee relationships on customer retention rates in a major UK retail bank", *Management Decisions*, Vol. 35 Nos 3/4, pp. 293-301.

Cooil, B., Keiningham, T.L., Aksoy, L. and Hsu, M. (2007), "A longitudinal analysis of customer satisfaction and share of wallet: investigating the moderating effect of customer characteristics", *Journal of Marketing*, Vol. 71 No. 1, pp. 67-83.

Corti⟩as, M., Margarita, E. and Mugica, J.M. (2008), "The use of loyalty cards databases: differences in regular price and discount sensitivity in the brand choice decision between card and non-card holders", *Journal of Retailing and Consumer Services*, Vol. 15, pp. 52-62.

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

Coussement, K. and Van den Poel, D. (2008b), "Integrating the voice of customers through call center emails into a decision support system for chum prediction", *Information & Management*, Vol. 45 No. 3, pp. 164-174.

Demoulin, N.T.M. and Zidda, P. (2009), "Drivers of customers adoption and adoption timing of a new loyalty card in the grocery retail market", *Journal of Retailing*, Vol. 85 No. 3, pp. 391-405.

Dierkes, T., Bichler, M. and Krishnan, R. (2011), "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks", *Decision Support Systems*, Vol. 51 No. 3, pp. 361-371.

Domingos, P. (1999), *Metacost: A General Method for Making Classifiers Cost-Sensitive*, 5th ed., translated by San Diego, CA.

Dowling, G.R. and Uncles, M. (1997), "Do customer loyalty programs really work?", *Sloan Management Review*, Vol. 38 No. 4, pp. 71-82.

Duda, R.O. and Hart, P.E. (2001), *Pattern Classification*, 2nd ed., John Wiley & Sons, New York, NY.

Elkan, C. (2001), "The foundations of cost-sensitive learning", translated by Seattle, WA, pp. 973-978.

Fan, W., Stolfo, S.J., Zhang, J. and Chan, P.K. (1999), *Adacost: Misclassification Cost-Sensitive Boosting*, 16th ed, Morgan Kaufmann Publishers, San Francisco, CA, pp. 97-105.

Farquhar, J.D. and Panther, T. (2008), "Acquiring and retaining customers in UK banks: an exploratory study", *Journal of Retailing and Consumer Services*, Vol. 15 No. 1, pp. 9-21.

Ganesh, J., Arnold, M.J. and Reynolds, K.E. (2000), "Understanding the customer base of service providers: an examination of the differences between switchers and stayers", *Journal of Marketing*, Vol. 64 No. 3, pp. 65-87.

Godfrey, A., Seiders, K. and Voss, G.B. (2011), "Enough is enough! The fine line in executing multichannel relational communication", *Journal of Marketing*, Vol. 75 No. 4, pp. 94-109.

Gupta, S., Lehmann, D.R. and Stuart, J.A. (2004), "Valuing customers", *Journal of Marketing Research*, Vol. 41 No. 1, pp. 7-18.

Gustafsson, A., Johnson, M.D. and Roos, I. (2005), "The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention", *Journal of Marketing*, Vol. 69 No. 4, pp. 210-218.

Holsapple, C.W., Pakath, R., Jacob, V.S. and Zaveri, J.S. (1993), "Learning by problem processors – adaptive decision-support systems", *Decision Support Systems*, Vol. 10 No. 2, pp. 85-108.

Homburg, C., Artz, M. and Wieseke, J. (2012), "Marketing performance measurement systems: does comprehensiveness really improve performance?", *Journal of Marketing*, Vol. 76 No. 3, pp. 56-77.

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

Hung, S.Y., Hung, W.H., Tsai, C.A. and Jiang, S.C. (2010), "Critical factors of hospital adoption on CRM system: organizational and information system perspectives", *Decision Support Systems*, Vol. 48 No. 4, pp. 592-603.

Keaveney, S.M. and Parthasarathy, M. (2001), "Customer switching behavior in online services: an exploratory study of the role of selected attitudinal, behavioral, and demographic factors", *Journal of the Academy of Marketing Science*, Vol. 29 No. 4, pp. 374-390.

Keh, H.T. and Lee, Y.H. (2006), "Do reward programs build loyalty for services? The moderating effect of satisfaction on type and timing of rewards", *Journal of Retailing*, Vol. 82 No. 2, pp. 127-136.

Kelly, S. (1996), *Data Warehousing: The Route to Mass Customization*, Wiley, New York, NY.

Koukova, N.T., Kannan, P.K. and Ratchford, B.T. (2008), "Product form bundling: implications for marketing digital products", *Journal of Retailing*, Vol. 84 No. 2, pp. 181-194.

Leenheer, J. and Bijmolt, T.H.A. (2008), "Which retailers adopt a loyalty program? An empirical study", *Journal of Retailing and Consumer Services*, Vol. 15 No. 5, pp. 429-442.

Leenheer, J., van Heerde, H.J., Bijmolt, T.H.A. and Smidts, A. (2007), "Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members", *International Journal of Research in Marketing*, Vol. 24 No. 1, pp. 31-47.

Lemmens, A. and Croux, C. (2006), "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, Vol. 43 No. 2, pp. 276-286.

Lessmann, S., Baesens, B., Mues, C. and Pietsch, S. (2008), "Benchmarking classification models for software defect prediction: a proposed framework and novel findings", *IEEE Transactions on Software Engineering*, Vol. 34 No. 4, pp. 485-496.

Lewis, M. (2004), "The influence of loyalty programs and short-term promotions on customer retention", *Journal of Marketing Research*, Vol. 41 No. 3, pp. 281-292.

Magi, A.W. (2003), "Share of wallet in retailing: the effects of customer satisfaction, loyalty cards and shopper characteristics", *Journal of Retailing*, Vol. 79 No. 2, pp. 97-106.

Maklan, S. and Knox, S. (2009), "Dynamic capabilities: the missing link in CRM investments", *European Journal of Marketing*, Vol. 43 Nos 11-12, pp. 1392-1410.

Margineantu, D.D. (2000), "On class probability estimates and cost-sensitive evaluation of classifiers", Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning (WCSL at ICML2000), Stanford University, June 29-July 2, Stanford, CA.

Margineantu, D.D. (2001), "Methods for cost-sensitive learning", unpublished thesis, Oregon State University, Corvallis, OR.

Meyer-Waarden, L. (2007), "The effects of loyalty programs on customer lifetime duration and share of wallet", *Journal of Retailing*, Vol. 83 No. 2, pp. 223-236.

Mittal, V. and Kamakura, W.A. (2001), "Satisfaction, repurchase intent, and repurchase behavior: investigating the moderating effect of customer characteristics", *Journal of Marketing Research*, Vol. 38 No. 1, pp. 131-142.

Nagar, V. and Rajan, M.V. (2005), "Measuring customer relationships: the case of the retail banking industry", *Management Science*, Vol. 51 No. 6, pp. 904-919.

Neslin, S.A., Gupta, S., Kamakura, W., Lu, J.X. and Mason, C.H. (2006), "Defection detection: measuring and understanding the predictive accuracy of customer churn models", *Journal of Marketing Research*, Vol. 43 No. 2, pp. 204-211.

Nitzan, I. and Libai, B. (2011), "Social effects on customer retention", *Journal of Marketing*, Vol. 75 No. 6, pp. 24-38.

Noble, S.M. and Phillips, J. (2004), "Relationship hindrance: why would consumers not want a relationship with a retailer?", *Journal of Retailing*, Vol. 80 No. 4, pp. 289-303.

Oliver, R.L. (2010), *Behavioral Perspective on the Consumer*, M.E. Sharpe, Armonk, NY.

Provost, F. and Fawcett, T. (2001), "Robust classification for imprecise environments", *Machine Learning*, Vol. 42 No. 3, pp. 203-231.

Reichheld, F.F. (1996), "Learning from customer defections", *Harvard Business Review*, Vol. 74 No. 2, pp. 56-70.

Reichheld, F.F. and Sasser, W.E. (1990), "Zero defections – quality comes to services", *Harvard Business Review*, Vol. 68 No. 5, pp. 105-111.

Richards, K.A. and Jones, E. (2008), "Customer relationship management: finding value drivers", *Industrial Marketing Management*, Vol. 37 No. 2, pp. 120-130.

Risselada, H., Verhoef, P.C. and Bijmolt, T.H.A. (2010), "Staying power of churn prediction models", *Journal of Interactive Marketing*, Vol. 24 No. 3, pp. 198-208.

Rust, R.T. and Chung, T.S. (2006), "Marketing models of service and relationships", *Marketing Science*, Vol. 25 No. 6, pp. 560-580.

Rust, R.T., Ambler, T., Carpenter, G.S., Kumar, V. and Srivastava, R.K. (2004), "Measuring marketing productivity: current knowledge and future directions", *Journal of Marketing*, Vol. 68 No. 4, pp. 76-89.

Sen, A. and Sinha, A.P. (2011), "IT alignment strategies for customer relationship management", *Decision Support Systems*, Vol. 51 No. 3, pp. 609-619.

Shaw, M.J. (1993), "Machine learning: methods for intelligence decision support – an introduction", *Decision Support Systems*, Vol. 10 No. 2, pp. 79-83.

Shaw, M.J., Subramaniam, C., Tan, G.W. and Welge, M.E. (2001), "Knowledge management and data mining for marketing", *Decision Support Systems*, Vol. 31 No. 1, pp. 127-137.

Sheng, V.S. and Ling, C.X. (2006), *Thresholding for Making Classifiers Cost-sensitive*, AAAI Press, Menlo Park, CA, pp. 476-481.

Stahl, F., Heitmann, M., Lehmann, D.R. and Neslin, S.A. (2012), "The impact of brand equity on customer acquisition, retention, and profit margin", *Journal of Marketing*, Vol. 76 No. 4, pp. 44-63.

Taylor, G.A. and Neslin, S.A. (2005), "The current and future sales impact of a retail frequency reward program", *Journal of Retailing*, Vol. 81 No. 4, pp. 293-305.

Ting, K.M. (2002), "An instance-weighting method to induce cost-sensitive trees", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 14 No. 3, pp. 659-665.

Torkzadeh, G., Chang, J.C.J. and Hansen, G. (2006), "Identifying issues in customer relationship management at Merck-Medco", *Decision Support Systems*, Vol. 42 No. 2, pp. 1116-1130.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B. (2011), "New insights into churn prediction in the telecommunication sector: a profit driven data mining approach", *European Journal of Operational Research*, Vol. 218 No. 1, pp. 211-229.

Verhoef, P.C. (2003), "Understanding the effect of customer relationship management efforts on customer retention and customer share development", *Journal of Marketing*, Vol. 67 No. 4, pp. 30-45.

Viaene, S. and Dedene, G. (2005), "Cost-sensitive learning and decision making revisited", *European Journal of Operational Research*, Vol. 166 No. 1, pp. 212-220.

Vorhies, D.W. and Morgan, N.A. (2005), "Benchmarking marketing capabilities for sustainable competitive advantage", *Journal of Marketing*, Vol. 69 No. 1, pp. 80-94.

Wright, M. and Riebe, E. (2010), "Double jeopardy in brand defection", *European Journal of Marketing*, Vol. 44 No. 6, pp. 860-873.

Ziarko, W. (1991), "The discovery, analysis and representation of data dependencies in databases", *Knowledge Discovery in Databases*, MIT Press, Cambridge, MA.

About the author

Kristof Coussement is Associate Professor and Chair of the Excellence Center in Database Marketing at IÉSEG School of Management (Lille/Paris) in France. His main research interests are all aspects in customer intelligence, B-2-B Intelligence, direct marketing and CRM. Improving his "practical" experience over the years by doing several real-life research projects in a different number of industries, as research associate of a Brussels-based company. He mainly focuses on profound academic research with a high-added value to business. He has published in *Decision Support Systems*, *Information & Management*, *Journal of Business Research*, *Expert Systems with Applications*, *European Journal of Operational Research*, *Computational Statistics and Data Analysis*, among others. Moreover, he is founder and committee member of BAQMaR, the largest online European Association for Quantitative & Qualitative Marketing Research.