ELSEVIER

Predicting customer loyalty using the internal transactional database

Wouter Buckinx, Geert Verstraeten *, Dirk Van den Poel

Department of Marketing, Ghent University, Hoveniersberg 24, 9000 Ghent, Belgium

Abstract

Loyalty and targeting are central topics in Customer Relationship Management. Yet, the information that resides in customer databases only records transactions at a single company, whereby customer loyalty is generally unavailable. In this study, we enrich the customer database with a prediction of a customer's behavioral loyalty such that it can be deployed for targeted marketing actions without the necessity to measure the loyalty of every single customer. To this end, we compare multiple linear regression with two state-of-the-art machine learning techniques (random forests and automatic relevance determination neural networks), and we show that (i) a customer's behavioral loyalty can be predicted to a reasonable degree using the transactional database, (ii) given that overfitting is controlled for by the variable-selection procedure we propose in this study, a multiple linear regression model significantly outperforms the other models, (iii) the proposed variable-selection procedure has a beneficial impact on the reduction of multicollinearity, and (iv) the most important indicator of behavioral loyalty consists of the variety of products previously purchased.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Predictive modeling; Customer relationship management; Behavioral loyalty; Overfitting; Multicollinearity; Data enrichment

1. Introduction

In the two latest decades, Customer Relationship Management (CRM) has grown to be one of the major trends in marketing, both in academia and in practice. This evolution took form in a dramatic shift in the domain, evolving from transaction-oriented marketing to relationship-oriented marketing (Grönroos, 1997), and builds strongly on the belief that it is several times less demanding – i.e. expensive – to sell an additional product to an existing customer than to sell the product to a new customer (Rosenberg & Czepiel, 1984). Hence, it has been argued that it is particularly beneficial to build solid and fruitful customer relationships, and in this discourse, customer loyalty has been introduced as one of the most important concepts in marketing (Reichheld, 1996).

From an analytical point of view, several tools have emerged in recent years that enable companies to strengthen their relationships with customers. Moreover, the rise of new media such as the World Wide Web, and the continuous technological improvements have further increased the opportunities to communicate in a more direct, one-to-one manner with customers (Van den Poel & Buckinx, 2005). Response modeling – i.e. predicting whether a customer will reply to a specific offer, leaflet or product catalog – represents the most central application in this domain, and serves as a tool to manage customer relationships. Indeed, it would be beneficial for the company–customer relationship that the latter party would receive only information that is relevant to him/her, hence allowing the company to present only those offers for which the individual customer shows a high response probability (Baesens, Viaene, Van den Poel, Vanthienen, & Dedene, 2002). Related to this, cross-selling analysis is involved with finding the optimal product to offer to a given customer (Chintagunta, 1992; Larivière & Van den Poel, 2004). Additionally, upselling analysis is focused on selling more – or a more expensive version – of the products that are currently purchased by the customer. Both techniques share a similar goal, i.e. to intensify the customer relationship by raising the share of products that is

---

* Corresponding author. Tel.: +32 9 264 35 24; fax: +32 9 264 42 79.
  E-mail address: Geert.Verstraeten@UGent.be (G. Verstraeten).

purchased at the focal company, and to prevent that these products would be purchased at competitive vendors. The fear of losing sales to competitors also features in churn analysis, which is focused on detecting customers exhibiting a large potential to abandon the existing relationship. Churn analysis has received great attention in the domain ever since it has been proven that even a small improvement in customer defection can greatly affect a company's future profitability (Reichheld & Sasser, 1990; Van den Poel & Larivière, 2004). Finally, lifetime value (LTV) analysis is a widely used technique to predict the future potential of customers, in order to target only the most promising customers (Hwang, Jung, & Suh, 2004). While these techniques can each serve individually to enhance customer relationships, it should be clear that additional advantages reside in the combination of these analytic techniques. Two recent attempts to integrate such techniques can be found in Baesens et al. (2004) and Jonker, Piersma, and Van den Poel (2004).

2. The need for predicting customer loyalty

In sum, we could state that both the focus on customer loyalty and the analytic tools described above have emerged from the CRM discourse. However, it is very unusual that actual customer loyalty is used to either devise or evaluate a company's targeted marketing strategies. The major cause of this deficiency lies most likely in the unavailability of information. Currently, while companies are maintaining transactional databases that store all details on any of a given customer's contacts with the focal company, these databases cannot capture the amount of products that this customer purchases at competing stores. Indeed, a study by Verhoef, Spring, Hoekstra, and Leeflang (2002) showed that only 7.5% of companies involved in database marketing activities collect such purchase behavior. Hence, the real behavioral loyalty of a certain customer is generally unavailable in the company's records,

whereby the full potential of the customer (i.e., the total needs of the customer for products in the relevant category) is unknown to any specific company. However, this information could prove to be extremely valuable in different applications.

First, the knowledge of a customer's loyalty would be useful for improving CRM. We illustrate this with an example from a banking context. It would most likely be more lucrative to offer an additional savings product to a customer who has a high balance at the focal bank and at the same time has large amounts invested at other banking institutions, than to offer the savings product to a customer that has an equally high balance, but where all his/her money is invested at the focal bank. Secondly, a notion of a customer's loyalty could be used for adapting the usefulness of the model-building process. For example, currently, cross-selling models are being built on the total customer database, whereby the users will estimate the probability of purchasing this product *at the focal company*, whereas from a cross-sales point of view, it would be more interesting to estimate whether they are interested in the product category *in general*. To overcome this, it could be interesting to build a cross-selling model on loyal customers only, because only for these customers, their total product needs are known. In this context, when attempting to model the real – and total – product needs of customers, it might seem suboptimal to include nonloyal customers into the analysis. Thirdly, the knowledge of a customer's loyalty and the evolution therein could be useful for evaluating the results of CRM-related investments, and monitoring whether certain actions lead to the desired results in the relevant customer segments.

While such loyalty information can be obtained through a questionnaire, it would prove to be financially infeasible to obtain this information for each individual customer, especially when customers would have to be surveyed regularly in order to track changes in their loyalty profile. Consequently, in this paper, we will prove that it is
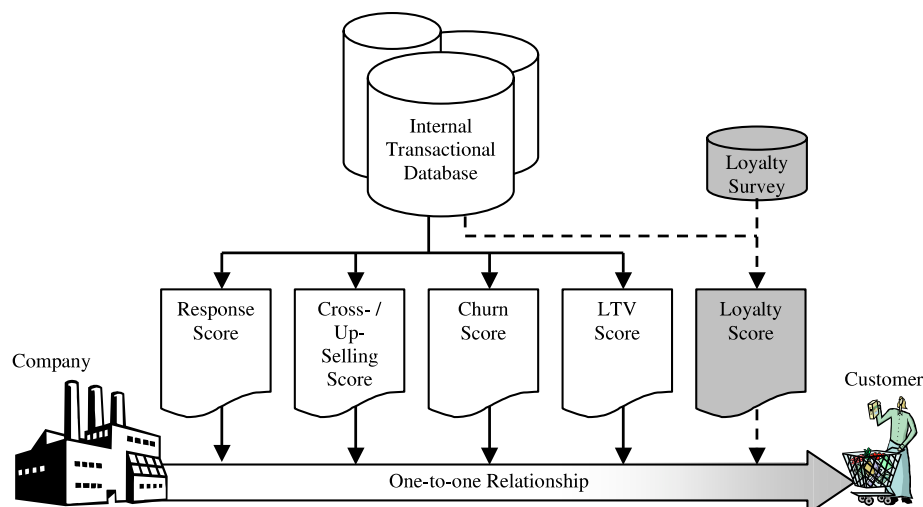

Fig. 1. Creating a loyalty score from transactional data and a loyalty survey.