Expert Systems With Applications 42 (2015) 8403–8412

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Improving direct mail targeting through customer response modeling

Kristof Coussement[a,*], Paul Harrigan[b,1], Dries F. Benoit[c,2]

[a] IESEG School of Management – Université Catholique de Lille (LEM, UMR CNRS 9221), Department of Marketing, 3 Rue de la Digue, F-59000 Lille, France
[b] The University of Western Australia – UWA Business School, M263, 35 Stirling Highway, Crawley, 6009, Australia
[c] Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, B-9000 Ghent, Belgium

ARTICLE INFO

Keywords:
Direct marketing
Direct mail
Response modeling
Database marketing

ABSTRACT

Direct marketing is an important tool in the promotion mix of companies, amongst which direct mailing is crucial. One approach to improve direct mail targeting is response modeling, i.e. a predictive modeling approach that assigns future response probabilities to customers based on their history with the company. The contributions to the response modeling literature are three-fold. First, we introduce well-known statistical and data-mining classification techniques (logistic regression, linear and quadratic discriminant analysis, naïve Bayes, neural networks, decision trees, including CHAID, CART and C4.5, and the k-NN algorithm) to the direct marketing community. Second, we run a predictive benchmarking study using the above classifiers on four real-life direct marketing datasets. The 10-fold cross-validated area under the receiver operating characteristics curve is used as evaluation metric. Third, we give managerial insights that facilitate the classifier choice based on the trade-off between interpretability and predictive performance of the classifier. The findings of the benchmark study show that data-mining algorithms (CHAID, CART and neural networks) perform well on this test bed, followed by simplistic statistical classifiers like logistic regression and linear discriminant analysis. It is shown that quadratic discriminant analysis, naïve Bayes, C4.5 and the k-NN algorithm yield poor performance.

1. Introduction

The move from mass-marketing to mass-customization is no better reflected than in the area of direct marketing, and in particular direct mail. Marketers no longer distribute their messages to a mass market, nor do they distribute based on basic demographic characteristics; rather they distribute and optimize different messages to different segments that are developed based on past behavior (Jonker, Piersma, & Van den Poel, 2004; Rowe, 1989; Wierich & Zielke, 2014). Still, the need to improve the effectiveness of direct mail campaigns is a persistent issue in many industries (Guido, Prete, Miraglia, & De Mare, 2011; Mahdiloo, Noorizadeh, & FarzipoorSaen, 2014).

Before sending direct mail, a key dilemma for marketers is which customers to target. In an effort to answer this question, marketers tend to use response modeling. Response modeling identifies customers that are likely to respond better to the marketing campaign based on their past response behavior.

The above perfectly fits in the philosophy underpinning one-to-one marketing communications seen in the customer relationship management (CRM) domain (Mahdiloo et al., 2014). CRM is a strategic approach to marketing underpinned by relationship marketing theory (Morgan & Hunt, 1994), which has been defined as “a comprehensive strategy and process that enables an organization to identify, acquire, retain and nurture profitable customers by building and maintaining long-term relationships with them” (Sin, Tse, & Yim, 2005, p. 1266). At the heart of CRM is data on customers. The increasing power of CRM technologies enables more and more sophisticated data collection, storage and analysis techniques. The ability to draw powerful analyses from customer data makes CRM – and thus response modeling – a critical success factor in today’s rapidly changing environment (Danaher & Rossiter, 2011; Kumar, 2008; Ngai, Xiu, & Chau, 2009).

The focus of this paper is on customer response modeling. The contributions of our research study are three-fold. First, we will introduce the most popular response modeling methods to the direct marketing community. In particular, we review a range of popular classification algorithms borrowed from the statistical and data-mining community (logistic regression, linear and quadratic discriminant analysis, naïve Bayes, neural networks, decision trees (CHAID, CART and C4.5) and the k-NN algorithm). Second, we complement the existing response modeling literature by integrating and

* Corresponding author. Tel.: +33 320545892.
E-mail addresses: k.coussement@ieseg.fr (K. Coussement), paul.harrigan@uwa.edu.au (P. Harrigan), dries.benoit@ugent.be (D.F. Benoit).
[1] Tel.: +61 8 6488 1979.
[2] Tel.: +32 9 264 3552.

http://dx.doi.org/10.1016/j.eswa.2015.06.054
0957-4174/

contrasting all classification algorithms into a framework that aims to benchmark their predictive capabilities in discriminating responders from non-responders in four real-life direct mail companies. Third, managerial insights on classifier choice are given to the direct marketing community taken into account the comprehensibility and predictive performance of the response model.

This paper is structured as follows. The next section introduces the direct marketing field and its links with customer response modeling. Following that the range of classification algorithms are introduced and explained. We then describe the evaluation metric used and further explain the characteristics of the datasets and the experimental setting. Finally, we present the results and their implications.

2. Direct marketing and response modeling

Direct marketing is defined as the 'interactive system of marketing which uses one or more advertising media to affect a measurable response and/or transaction at any location' (Direct Marketing Association 2009). Direct marketing is big business. It is projected that direct marketing expenditures in the US will grow to $196 billion in 2016, with direct mail forming part of this growth. Direct mail is targeted at customers that are most likely to be enticed by particular offers, as opposed to a traditional mass marketing approach whose promotional activities are addressed to customers and prospects indistinctly (Guido et al., 2011; Mahdiloo et al., 2014; Risselada et al. 2014). Direct mail is not being killed off by the Internet; rather it is being used as a complementary channel (Danaher & Rossiter, 2011). Winterberry Group confirms that direct mail is still on the rise (Conlon, 2015). In 2014, direct mail spending grew with 2.7% in the United States compared to the projected 1.1% growth. Moreover, the market analysts project a 1% growth increase in direct mail spending for 2015, equivalent to $45.7 billion of the $156.8 billion representing the total direct and digital spending projection for 2015. The reason by Winterberry group is that direct mail costs will stay steady, and thus they expected that the projected 1% growth to come from volume increases.

Continued growth will be predicated upon the levels of return on investment of direct mail campaigns, which significantly depends on marketers being able to use specialized targeting techniques to come up with the right set of customers to contact (Lamb, Hair, & McDaniel, 1994).

Thus, the importance in knowing which customers are more likely to respond to a certain mailing is of paramount importance to marketers. Determining or predicting those customers who have a high probability to respond to a specific mailing based on their past behavior is called the customer response modeling (Bose & Chen, 2009; Mahdiloo et al., 2014). Response modeling is part of the classification literature stream. Classification is the procedure where customers are predicted to belong to predefined groups or target classes based on their historical customer information (Blattberg, Kim, & Neslin, 2008). Typically, a response model is estimated on a training set in which both the independent variables, describing and profiling a particular customer, and the dependent (response) variable, whether the customer responded on a certain mailing, are observed. Then, the estimated model on the training data is applied to a new set of customers that are not used during training (the test set). The result is a response probability for each customer in the test set, dependent on his or her past behavior. Managerially speaking, depending on the direct mail campaign budget, the company is able to target the top *x%* of customers with the highest response probability given by the response model.

The next section of this paper will introduce and describe the range of statistical and data-mining algorithms that can be used in customer response modeling.

3. Classification algorithms

The essence of one-to-one marketing communication is providing the right customers with marketing messages that they can easily act on (Ryals, 2005) This means that 'prediction and targeting are both key to decision making underlying direct marketing campaigns' (Zahavi & Levin, 1997, p.35). Therefore, understanding which techniques yield the best predictive capabilities is vital for direct marketers (Bose & Chen, 2009; Rada, 2005). With increased efficiencies and effectiveness, marketers could reduce mailing costs (Barwise & Farley, 2005), increase conversion rates (Kaefer, Heilman, & Ramenofsky, 2005), and increase customer retention (Watjatrakul & Drennan, 2005).

Our literature review reveals that existing literature utilizes several statistical and data-mining classification algorithms in various research setups to separate responders from non-responders. However, we complement the academic literature by presenting and integrating the most popular classifiers into one predictive benchmark study over multiple response datasets, while summarizing the managerial implications for managers. Several statistical classification methods to predict customer responses have been proposed and utilized, such as logistic regression, discriminant analysis and naïve Bayes (Baesens, Viaene, Van den Poel, Vanthienen, & Dedene, 2002; Berger & Magliozzi, 1992; Coussement, Van den Bossche, & De Bock, 2014; Cui, Wong, & Zhang, 2010; Deichmann, Eshghi, Haughton, Sayek, & Teebagy, 2002; Kang, Cho, & MacLachlan, 2012; Lee, Shin, Hwang, Cho, & MacLachlan, 2010). These techniques can be very powerful, but each algorithm also makes several stringent, but different, assumptions on the underlying distribution between the independent variables and the dependent variable. To counter this, more advanced data-mining algorithms have been proposed for discriminating between responders and non-responders, such as artificial neural networks (Baesens et al., 2002; Chen, Hsu, & Hsu, 2011; Curry & Moutinho, 1993; Zahavi & Levin, 1997), decision tree-generating techniques (Buckinx, Moons, Van den Poel, & Wets, 2004; Chen, Hsu, & Chu, 2012; Haughton & Oulabi, 1997; McCarty & Hastak, 2007; Rada, 2005) and k-NN learners (Govindarajan & Chandrasekaran, 2010; Kang et al., 2012).

The following sections review the most popular response models by describing their functioning, and by discussing their merits and drawbacks.

3.1. Logistic regression

Logistic regression (LOG) is a well-known and industry-standard classification technique for predicting a dichotomous dependent variable such as respond/do not respond to a mailing (Coussement et al., 2014; Suh, Noh, & Suh, 1999). Besides applications in direct marketing, it is an often used technique in a variety of predictive business settings like customer segmentation (McCarty & Hastak, 2007), churn prediction (Neslin, Gupta, Kamakura, Lu, & Mason, 2006), customer choice modeling (West, Brockett, & Golden, 1997) and many others. Moreover, logistic regression has several advantages (Hosmer & Lemeshow, 2000).

For a given training set with N labeled training examples $(x_i, y_i)\}$ with $i = 1, 2, \ldots, N$ with input data $xi \in Rn$ and corresponding binary target labels $y_i \in \{0, 1\}$, the logistic regression tries to estimate the probability $P(y = 1|x)$ given by

$$P(y = 1|x) = \frac{1}{1 + exp(-(w_0 + \boldsymbol{wx}))} \tag{1}$$

with $x \in R^n$ being equal to an n-dimensional input vector, $\mathbf{w}$ to the parameter vector and w_0 to the intercept. The parameters w_0 and $\mathbf{w}$ are usually estimated using a maximum likelihood procedure (Hosmer & Lemeshow, 2000).