Expert Systems with Applications 39 (2012) 6738–6753

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Improved response modeling based on clustering, under-sampling, and ensemble

Pilsung Kang [a], Sungzoon Cho [b,*], Douglas L. MacLachlan [c]

[a] IT Management Programme, International Fusion School, Seoul National University of Science and Technology (Seoultech), 232 Gongneoung ro, Nowon-gu, 139-743 Seoul, South Korea
[b] Department of Industrial Engineering, Seoul National University, 599 Gwanak-ro, Gwanak-gu, 151-744 Seoul, South Korea
[c] Department of Marketing and International Business, Foster School of Business, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Keywords:
Direct marketing
Response modeling
Class imbalance
Data balancing
CRM
Clustering
Ensemble

ABSTRACT

The purpose of response modeling for direct marketing is to identify those customers who are likely to purchase a campaigned product, based upon customers' behavioral history and other information available. Contrary to mass marketing strategy, well-developed response models used for targeting specific customers can contribute profits to firms by not only increasing revenues, but also lowering marketing costs. Endemic in customer data used for response modeling is a class imbalance problem: the proportion of respondents is small relative to non-respondents. In this paper, we propose a novel data balancing method based on clustering, under-sampling, and ensemble to deal with the class imbalance problem, and thus improve response models. Using publicly available response modeling data sets, we compared the proposed method with other data balancing methods in terms of prediction accuracy and profitability. To investigate the usability of the proposed algorithm, we also employed various prediction algorithms when building the response models. Based on the response rate and profit analysis, we found that our proposed method (1) improved the response model by increasing response rate as well as reducing performance variation, and (2) increased total profit by significantly boosting revenue.

1. Introduction

Response modeling has become one of the most effective tools for firms seeking to sustain long-term relations with their customers (Berry & Linoff, 2004; Gönül & Hofstede, 2006; Sun, Li, & Zhou, 2006). The goal of response modeling is to identify customers who are likely to purchase a product, based on customers' purchase history and other information. Based on model predictions, firms attempt to induce higher potential buyers to purchase the campaigned product using their communication channels, e.g., phone, mailed catalog, or e-mail. A well-developed response model can contribute to business in two ways. First, it increases total revenue. The customers during the marketing campaign are typically divided into two groups: one group who would buy the product anyway whether or not they are targeted, and the other group who would not buy the product had they not been targeted. By timely reminding the latter group of what they might need, they may be persuaded to open their wallets. Thus, the additional sales made to those customers are the obvious contribution of the response model. Second, it lowers total marketing cost. Generally, mass advertising is extremely expensive, since a customer's average likelihood of purchase is very low. Contrary to mass marketing, the response model suggests attempting to attract only customers with a relatively high purchase likelihood. Therefore, it saves the money that would have been spent to expose customers to promotional messages who have little interest in buying the product. With increased revenue and lowered cost, firms' net profit increases (Berry & Linoff, 2004; Elsner, Krafft, & Huchzermeier, 2004; Gönül & Hofstede, 2006; Zhang & Krishnamurthi, 2004).

Past studies have shown that while increasing response rate is not an easy task, its impact is quite incredible. For instance, Coenen, Swinnen, Vanhoof, and Wets (2000) pointed out that even a small improvement of response rate can change the total result of a direct mailing campaign from failure to success. Baesens, Viaene, Van den Poel, Vanthienen, and Dedene (2002) illustrated how a small improvement of response rate could result in huge additional profit. In their example, only 1% of increased response rate for an actual mail-order company yielded an additional 500,000 Euro. Knott, Hayes, and Neslin (2002) reported that for a retail bank, only 0.7% of increased response rate tripled total revenue and raised revenue per respondent by 20%. Sun et al. (2006) noted that improvement of the response rate can not only increase profit but also strengthen customer loyalty because properly targeted customers are more likely to be satisfied and stay with the firm over the long run.

Encouraged by its noticeable positive effect when successful, a large number of studies have been conducted with the objective of increasing response rate through improving the prediction

* Corresponding author.
E-mail addresses: pskang@seoultech.ac.kr (P. Kang), zoon@snu.ac.kr (S. Cho), macl@uw.edu (D.L. MacLachlan).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.12.028

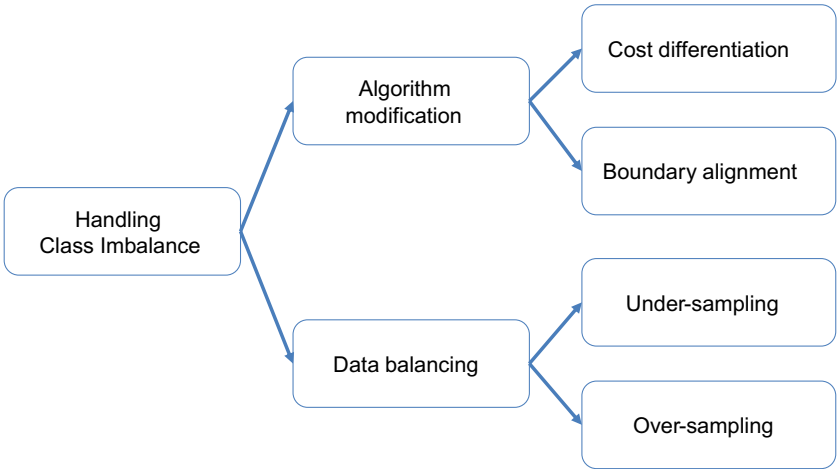

Fig. 1. Approaches to handling class imbalance.

algorithms used in response modeling. Logistic regression has been widely employed as a base model due to its simplicity and availability (Aaker, Kumar, & Day, 2001; Hosmer & Lemeshow, 1989). Besides logistic regression, stochastic RFM models (Colombo & Jiang, 1999) and hazard function models (Gönül, Kim, & Shi, 2000) were proposed from statistics traditions, while artificial neural networks (Baesens et al., 2002; Kaefer, Heilman, & Ramenofsky, 2005), bagging artificial neural networks (Ha, Cho, & MacLachlan, 2005), Bayesian neural networks (Baesens et al., 2002) support vector machines (Shin & Cho, 2006), and decision trees (Coenen et al., 2000) were proposed from pattern recognition and data mining researchers.

The most prevalent difficulty of response modeling is the class imbalance problem. In classification tasks, class imbalance occurs when the incidence of one class extremely outnumbers that of other classes. Class imbalance usually degrades the performance of classification algorithms. Most classification algorithms require sufficient instances from all classes to yield stable models that provide unbiased classification. If one class greatly outnumbers other classes, classification results tend to be biased toward the majority class. For customer databases used for response modeling, it is common that non-respondents overwhelmingly outnumber respondents. For example, less than 10% of customers are respondents in the DMEF4 data set used in Ha et al. (2005) and Shin and Cho (2006), while only about 6% of customers are respondents in the CoIL Challenge 2000 data set (van der Putten, de Ruiter, & van Someren, 2000). To make matters worse, response rates in general direct marketing situations are often much lower. If an appropriate remedy for class imbalance is not taken, classification algorithms employed by response modeling are likely to judge most customers as not to respond, which leads to a high opportunity cost. For that reason, handling the class imbalance of customer data has been recognized as a critical factor for the success of direct marketing (Blaszczyński, Dembczyński, Kotlowski, & Pawlowski, 2006; Hill, Provost, & Volinsky, 2006; Lai, Wang, Ling, Shi, & Zhang, 2006; Ling & Li, 1998).

Setting response modeling aside, class imbalance is a common symptom of classification tasks in many subject area, such as image processing (Kubat, Holte, & Matwin, 1998; Yan, Liu, Jin, & Hauptmann, 2003), remote sensing (Bruzzone & Serpico, 1997), and medical diagnosis (Lee, Cho, & Shin, 2008; Pizzi, Vivanco, & Somorjai, 2001). Therefore, a number of methods to overcome class imbalance have been proposed, which can be grouped into two categories, algorithm modification and data balancing as shown in Fig. 1. Methods based on algorithm modification insert an additional specialized mechanism into the original algorithm. There are two ways to do this: (1) giving different misclassification costs to each class, or (2) shifting the decision threshold toward the minority class. For example, Wu and Chang (2003) proposed giving a larger misclassification cost to the minority class than to the majority class and modifying the kernel matrix when training support vector machines.[1] Bruzzone and Serpico (1997) divided the training process of neural networks into two phases. In the first phase, neural networks were trained with misclassification costs that were inversely proportional to the number of patterns in each class.[2] In the second phase, using the obtained weights in the first phase as the initial weights, networks were trained again to minimize mean squared error (MSE). Huang, Yang, King, and Lyu (2004) tried to tackle the class imbalance by training a biased "*Minimax*" machine. In the *Minimax* machine, the objective function was formulated in order to maximize the accuracy of the minority class classification given a lower bound of majority class accuracy.

Data balancing methods build a new training data set in which all classes are well-balanced, using different sampling strategies for each class. They have an advantage over algorithm modification methods in that they are universal. Because data balancing methods work independently from classification algorithms, they can be combined with any classifiers while algorithm modifications work well only with the particular classifiers for which they are designed.[3] Under-sampling and over-sampling are two major recipes for data balancing. Under-sampling reduces the number of majority class instances while keeping all the minority class instances. The portion of minority class entities in the training data increases as a consequence. Under-sampling is effective in reducing training time, but it often distorts the class distribution because a large number of majority class instances are removed. Random sampling is the simplest way to implement under-sampling. In random under-sampling, a set of majority class instances is selected at random and combined with the minority class patterns. SHRINK (Kubat, Holte, & Matwin, 1997) and one-sided selection (OSS) (Kubat & Matwin, 1997) are other well-known under-sampling methods.

Fig. 2 shows the OSS algorithm. OSS removes majority class instances identified as either noise, redundant, or borderline. A "*noise*" instance is surrounded by the other class's instances while a "*redundant*" instance is surrounded by the same class's instances.

[1] What the data mining literature denotes as "training", the statistics literature calls fitting or estimating.

[2] "Patterns" in the data mining terminology corresponds to vectors of observations in statistics.

[3] The term "classifier" is used in data mining to denote the model or rule used to classify entities.