

Direct Marketing When There Are Voluntary Buyers

Yi-Ting Lai and Ke Wang
Simon Fraser University
{llai2, wangk}@cs.sfu.ca

Daymond Ling, Hua Shi, and Jason Zhang
Canadian Imperial Bank of Commerce
{Daymond.Ling, Hua.Shi, Jason.Zhang}@cibc.com

Abstract

In traditional direct marketing, the implicit assumption is that customers will only purchase the product if they are contacted. In real business environments, however, there are “voluntary buyers,” who will still make the purchase in the absence of a contact. While no direct promotion is needed for voluntary buyers, the traditional response-driven paradigm tends to target such customers. This paper presents “influential marketing,” targeting only those whose purchase decisions can be positively influenced, i.e. buyers who are non-voluntary. Our novel, practical solution to this problem gives promising results.

1. Introduction

Direct marketing campaigns promote products to potential customers by contacting them via a direct channel of communication, such as telephone or mail. The traditional practice for such a campaign typically involves the following steps. First, collect historical data from a previous campaign. Each customer record is associated with a number of individual characteristics (e.g. age, income), and a response variable indicating whether a customer responded after receiving the direct promotion. Second, perform data mining on the historical data to construct a model, with the goal of estimating the probability that a customer will respond to the promotion. Third, deploy the model to rank all potential customers in the current campaign by their estimated probability of responding. Contact some top ranked customers.

The effectiveness of a campaign is then determined by the response rate of contacted customers. This objective, which targets the *most likely responders*, has been widely adopted by both academia and industries [8, 6, 10]. An implicit assumption is that all purchases are generated by a direct contact. How a customer would behave without the direct promotion is of no concern. The following example shows that if certain customers have decided to buy the product *before* receiving the direct promotion, then the traditional objective does not address the right problem.

Example 1. John has recently got married and the young couple has a joint account at bank X. Planning to buy a house, John has decided that he will apply for a mortgage at bank X. The decision was made because John learnt about the mortgage program when visiting the bank.

Applying traditional direct marketing strategies, bank X discovered that young newlyweds are more likely to respond to the direct promotion on the bank’s mortgage program. Hence, the bank sent John a brochure about its mortgage program. Though it is true that John will respond positively, he would have done so even without the promotion. In other words, contacting John added no new value to the campaign. ■

The example reveals two interesting points. First, certain customers base their purchase decisions on factors other than a direct marketing campaign, e.g. John. We call such customers “voluntary buyers.” Customers may become voluntary buyers due to word-of-mouth or viral marketing [2] or prior knowledge about the product. Second, the traditional response-driven paradigm tends to target voluntary buyers because such customers generate the highest possible responses, though no direct marketing is required for them. Clearly, a campaign that has a 5.4% response rate among contacted customers is not very successful if, from a similar group of customers who are not contacted, we observe that 5% purchase the product.

The traditional strategy of maximizing the response rate aims to avoid non-buyers. In this paper, we advocate another largely ignored strategy: avoid voluntary buyers. When both strategies are in place, the campaign will focus on the customers who will buy if only if they are contacted. We believe that this focus is the right objective of a direct marketing campaign.

One question is — how significant in practice is the portion of voluntary buyers? If not significant, it may be acceptable to “push voluntary buyers through” to close the deal while focusing on avoiding non-buyers. To answer this question, a real campaign was carried out (see details in Section 5). Instead of contacting all selected customers, we withheld the contact action for a random subset of selected customers. It turns out that while the contacted group has a response rate of 5.4%, the not-contacted group has a response rate of 4.3%. In other words, 80% of the responders contacted would have responded even without the contact! Aside from cost considerations, unnecessary promotions can potentially

annoy customers and, in the worst case, lead customers to switch to a competing product or company. This consideration alone is a big enough concern to the company.

The contribution of this paper is as follows. First, we introduce a new taxonomy of customers based on their purchase behaviors. All customers are classified into three classes: *decided*, *undecided*, and *non*. While decided and non customers have made up their minds on whether to buy the product, undecided customers will buy if and only if they are contacted. Since only undecided customers can be *positively influenced* by a campaign, direct marketing should target undecided customers. *Influential marketing* refers to this objective. The major challenge is that undecided customers are not explicitly labeled; thus, standard supervised learning is not directly applicable. Our novel solution addresses this challenge while requiring no major changes to the standard campaign practice. The evaluation is conducted with real campaign data.

In Section 2, we formally define the new taxonomy and influential marketing. In Section 3, we give our solution to influential marketing. We discuss related work in Section 4. Section 5 reports the experimental results.

2. Influential Marketing

Consider a pool P of potential customers. Ultimately, a marketing campaign aims to maximize the net profit over P . As is the case of many direct marketing campaigns, we assume a fixed revenue R for each purchase and a fixed cost C for each direct contact; $R > C$. The net profit for each purchase due to a direct contact is $R - C$. The *set of buyers over P* consists of all the customers who eventually buy the product, either voluntarily or due to a direct contact. From now on, we express the budget of a campaign as a fixed number of contacts. To maximize the total net profit, we need to maximize the number of buyers over P for a given number of contacts. We now introduce the three classes of customers:

- **Decided:** the customers who voluntarily buy the product, regardless of a direct promotion.
- **Undecided:** the customers who buy the product if and only if the product is directly promoted to them. These customers can be positively influenced by a direct promotion.
- **Non:** the customers who will not buy the product, regardless of a direct promotion.

Each customer belongs to exactly one of these classes. Our discussion will refer to customers by this classification; however, our solution does not depend on determining the class of each customer a priori. The set of buyers now consists of (i) all decided customers and (ii) the undecided customers contacted by the campaign.

While (i) is fixed across all campaigns, (ii) depends on whom are selected to contact.

Let S be the set of customers contacted by a campaign. Let D be the set of decided customers in S , and U be the set of undecided customers in S . $|X|$ denote the number of customers in X .

Definition 1 (RR, DBR, UBR) For S , the *response rate*, RR , is $(|D|+|U|)/|S|$. The *decided buyer rate*, DBR , is $|D|/|S|$. The *undecided buyer rate*, UBR , is $|U|/|S|$. ■

Example 2. Two direct marketing campaigns, M1 and M2, each contact the same number of customers. We illustrate this in Figures 1 and 2. The largest rectangle represents the customer pool P which consists of a fixed number of decided, undecided, and non customers. The group of contacted customers, S , is represented by a dashed rectangle.

Figure 1 highlights the set of buyers over S while Figure 2 highlights the set of buyers over P . M1 has a higher RR than M2, as illustrated in Figure 1. If we were to follow the traditional objective of maximizing RR , then M1 is preferable to M2. Yet, M1 actually produces a smaller set of buyers over P , as shown in Figure 2. The reason is that M1 targets mostly voluntary buyers, but contacting such customers does not increase the total number of buyers. Judging from the net profit generated, M2 should actually be preferable to M1. ■

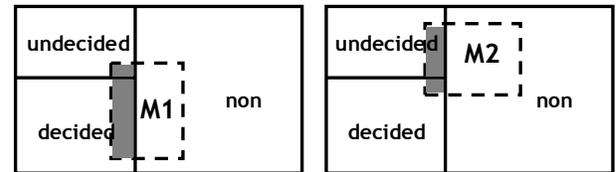


Figure 1. M1 has a larger RR than M2.

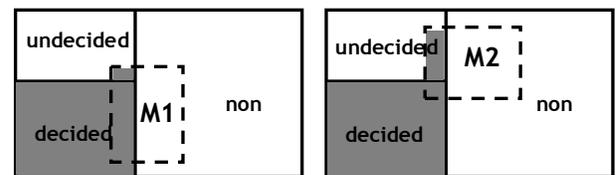


Figure 2. M1 has a smaller set of buyers over P than M2.

Clearly, for a fixed number of contacts, the only way to generate a higher net profit is to contact undecided customers. Therefore, instead of indiscriminately targeting all likely responders, we should target only undecided customers who can contribute to the increase of total buyers. This is the core concept of *influential marketing*.

Definition 2 (Influential Marketing) For a given number of contacts, *influential marketing* aims to maximize UBR by targeting undecided customers. ■

Influential marketing increases the total number of buyers by generating “new” buyers, i.e. undecided customers. While the traditional paradigm may have a larger RR , many of the responders are in fact decided buyers who are not “new” buyers. To increase the number of new buyers, the right objective is not to maximize RR but to maximize UBR .

The main challenge of influential marketing is that undecided customers were not explicitly labeled in the normal campaign data. When a contacted customer responds, she could be either a decided or undecided customer; we do not know which. Therefore, standard supervised learning where the positive and negative classes are explicitly labeled is not directly applicable. Another challenge is that given their existing investments, campaign managers generally do not wish to switch to a different set of marketing methodologies. This consideration is particularly important in the business environment where deliveries have a priority over publications. A solution that addresses these requirements must be innovative in that it not only identifies undecided buyers, but also requires little or no changes to standard methods. We present a solution with this goal in mind.

3. Proposed Solution

3.1. Data Collection

The novelty of our method lies in collecting the training data in such a way that standard supervised learning can be used to solve the influential marketing problem. Let S be the set of customers contacted. Recall that $RR = UBR + DBR$, where RR can be directly computed from S . If we can estimate DBR , then we will be able to compute UBR . Suppose we have a set of customers *similar* to S , denoted S_{sim} ; while customers of S are contacted, customers of S_{sim} are not contacted. Then DBR can be estimated by observing the response rate of S_{sim} . Based on this observation, we collect two disjoint sets of data, Treatment and Control, from the previous campaign.

- **Treatment:** a set of customers who were contacted.
- **Control:** a set of customers who were not contacted. The purchase behaviors of Control are used to approximate those of Treatment under the alternative marketing decision of not contacting them.

For Control to approximate Treatment, the two groups need to share similar characteristics, i.e. come from *the same underlying population*. Suppose the training data came from a previous campaign. In recent years, it has become a standard industry practice to withhold a random subset of selected customers for comparative purposes. In particular, for each selected customer c , toss a $\rho/(1-\rho)$ -sided coin to determine whether c was added to Treatment

or Control; ρ specifies the ratio of Treatment and Control. Customers withheld will not be contacted, but their purchase behaviors will still be recorded.

With the same split ratio, we randomly split Treatment into two sets, $T1$ and $T2$, and randomly split Control into two sets, $C1$ and $C2$. $\langle T1, C1 \rangle$ is for model training and $\langle T2, C2 \rangle$ is for model evaluation. Subsequently, $\langle T1, C1 \rangle$ is called the *training set* and $\langle T2, C2 \rangle$ is called the *validation set*.

3.2. Model Construction

We partition $\langle T1, C1 \rangle$ into four groups, as shown in the *learning matrix* in Table 1:

- Group (1): The customers in $T1$ who were contacted and responded. They can be either decided or undecided customers; we do not know which.
- Group (2): The customers in $T1$ who were contacted but did not respond. They must be non customers.
- Group (3): The customers in $C1$ who purchased the product without the contact. They must be decided customers.
- Group (4): The customers in $C1$ who received no contact and made no purchase. They can be either non or undecided customers; we do not know which.

Table 1. The learning matrix.

		Response	
		Yes	No
Treatment ($T1$)	decided + undecided (1)	non (2)	
	Control ($C1$)	decided (3)	non + undecided (4)

Observe that undecided customers are present only in groups (1) and (4). This motivates us to combine (1) and (4) to form the positive class, PC , which covers all undecided customers. However, PC also includes some decided and non customers. To “remove” decided and non customers from PC , we combine (2) and (3) to form the negative class, NC . Note that the only class of customer appearing in PC but not NC is the undecided. Therefore, if we apply a supervised learning algorithm Alg with PC as the positive class and NC as the negative class, the algorithm *will tend to pick out the characteristics exclusive of PC* — such characteristics are likely those of undecided customers. We assume that Alg is able to rank records by the probability of belonging to PC . Most supervised learning algorithms have this feature.

Alg is expected to rank a customer c higher if c has a higher probability of belonging to PC than to NC . From the above analysis, we know that the only class of

customer belonging to PC but not NC is the “undecided.” Therefore, a higher rank given by Alg indicates a higher probability of being undecided.

Claim. Assume that Alg is capable of ranking customers according to their probability of belonging to the positive class PC . Then for the model returned by Figure 3, if customer $c1$ is ranked higher than customer $c2$, $c1$ has a higher probability of being undecided than $c2$. ■

Our proposed solution also alleviates the notorious class imbalance problem suffered by most supervised learning algorithms [5]. Since the response rate in a direct marketing campaign is typically very low, the size of (1) tends to be much smaller than (2), and the size of (3) tends to be much smaller than (4). By combining the small (1) with the large (4) to classify against the combination of another small (3) and large (2), we effectively lessen the degree of imbalance of classes.

Further, the learning matrix provides a flexible way for the campaign to focus on the different class of customers. Typically, the size of (3) is much smaller than the size of (4) due to the vast majority of non customers. When targeting undecided customers as in influential marketing, we may want to over-sample (3) to emphasize the characteristics of decided customers in NC so to “remove” the decided customers in PC . On the other hand, when the focus is on all responders (either undecided or decided) as in the traditional paradigm, we may want to under-sample (4) to limit the number of non customers in PC .

Model Construction

Input: The training data $\langle T1, C1 \rangle$.

Parameter: A supervised learning algorithm Alg .

Output: A model that ranks records by their probability of belonging to the undecided class.

1. Combine (1) and (4) into the positive class PC ;
2. Combine (2) and (3) into the negative class NC ;
3. Feed PC and NC into Alg ;
4. Return the model learnt by Alg .

Figure 3. Model construction.

3.3. Model Evaluation

A direct marketing campaign contacts the top $x\%$ customers as ranked by the model; x is the *marketing percentile*. The choice of x should maximize the net profit within the budget constraint. To evaluate a model, M , we apply M to rank the validation data $\langle T2, C2 \rangle$. For a specified x , let $T2x$ and $C2x$ denote the top x -percentile of the ranked list of $T2$ and $C2$, respectively. Let

- MT denote RR in $T2x$,
- MC denote RR in $C2x$.

From Definition 1, $MT = UBR + DBR$; UBR and DBR

are the percentages of undecided and decided customers in $T2x$, respectively. For an estimated DBR , we can compute UBR by $MT - DBR$. To estimate DBR , as discussed in Section 3.1, we can use RR of $C2x$, i.e. MC , to approximate DBR of $T2x$ because $T2x$ and $C2x$ are ranked by applying the same model to a similar population, and because all the responses of $C2x$ are from decided customers. This gives the following computation of UBR for $T2x$.

Theorem 1. For $T2x$, UBR is given by $MT - MC$. ■

An effective direct marketing campaign should perform better than random marketing where customers are randomly selected for contact. Let Random denote the selection model for random marketing. Note that the expected RR of Random is the same across all marketing percentiles x . Hence, we take MT and MC of Random to be those at $x = 100\%$, denoted RT and RC , respectively. UBR of Random is given by $RT - RC$.

The model evaluation is summarized using the *positive influence curve (PIC)* in Figure 4. The x -axis represents the marketing percentile x . The y -axis represents $MT - MC$, or UBR , which is the *marketing value* relative to the purchases made only by decided customers. $(MT - MC) - (RT - RC)$ gives the *model value* relative to the marketing value of Random. An effective model should have both a large marketing value and a large model value. In PIC , this means the curve $MT - MC$ is well above the curve $RT - RC$ and has a decreasing trend from left to right, as undecided customers concentrate in the top percentiles.

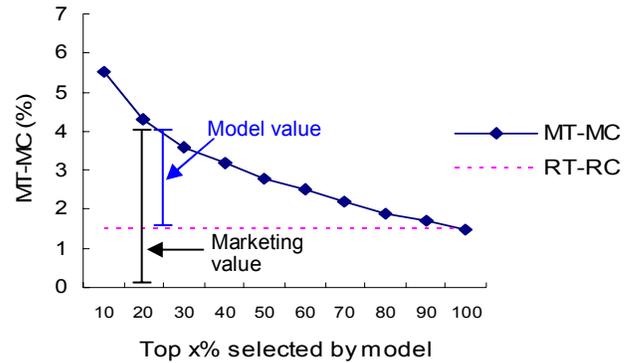


Figure 4. The positive influence curve (PIC).

4. Related Work

The response rate of contacted customers is a common objective for direct marketing. Some works use the accumulative lift curve (ALC) as the performance criterion [8]. Others use the receiver operating characteristic (ROC) curves [3, 4] to measure the relative trade-offs between true positives and false positives. Alternatively, cost-sensitive learning [1] recognizes that

identifying a buyer as a non buyer incurs a higher cost than identifying a non buyer as a buyer. Essentially, with any of these evaluation criteria, a model tends to target decided customers instead of undecided customers.

To our knowledge, the only published work addressing the same problem as ours is Lo’s work [9]. For every customer i , Lo predicts the probability of i responding if contacted, and the probability of i responding if not contacted. A treatment variable T is added to the training data; $T_i = 1$ when i was contacted and $T_i = 0$ when i was not contacted. A supervised learning algorithm is then used to train a model for the response variable. As in the traditional approach, the positive and negative classes represent whether a customer responds. Given a customer i , the model computes the two probabilities assuming $T_i = 1$ and $T_i = 0$. Customers are then ranked by the difference between the two probabilities.

However, for the treatment variable T to be useful, Lo’s approach implicitly requires that T is discriminative of the classes; that is, $T = 1$ is more associated with the positive class than the negative class. This implies that the campaign from which such data was collected has successfully targeted undecided customers. In other words, a successful model has been deployed in the previous campaign, which defeats the purpose of constructing another such model from the training data. In addition, we observed that Lo [9] has only experimented with a set of simulated data that has an unusually high response rate, e.g. $MT = 90\%$ at $x = 10\%$. Essentially, the fundamental limitation of Lo’s approach is the lack of attempt in isolating undecided customers. In particular, all undecided customers in Control are labeled as negative.

5. Experimental Evaluation

We experimented on real campaign data, provided by the Canadian Imperial Bank of Commerce (CIBC). This data set contains 24,506 records collected from a loan product promotion conducted by the bank. Each record is described by 608 independent attributes, in which about 1/6 are categorical and the remaining numerical. A detailed breakdown is given in Table 2. Observe that the response rate of Treatment, or RT , is 5.4%, and the response rate of Control, or RC , is 4.3%. This indicates that the marketing value of Random is $RT - RC = 1.1\%$.

Note that the two groups, “Response = Yes” and “Response = No,” have a similar distribution: 92% and 90% of the records received the treatment (i.e. $T = 1$), respectively. This suggests that Lo’s approach of using the treatment variable T will not be effective on this campaign data set.

All results reported below were obtained with 3-fold cross validation, i.e. $k = 3$. If k is too large, $C2$ will consist of only $1/k$ of (3), introducing instability to the results. The association-rule based classifier (ARC) [11] is chosen

for its ability to handle imbalanced classes; numerical attributes are discretized using MLC++ [7]. The decision tree in SAS Enterprise Miner [10] (SAS EM Tree) is also used. SAS is the leading software for business analytics across many industries. Both algorithms are capable of ranking records.

Table 2: Breakdown of the campaign data.

	Response	
	Yes	No
Treatment	(1) 1,182	(2) 20,816
Control	(3) 108	(4) 2,400

The results are summarized in Figures 5 and 6. We experimentally evaluate how effective each of the following approaches is in terms of targeting undecided customers. “Traditional” refers to the traditional objective of maximizing RR . “Lo’s” refers to the approach in [9] with the use of the treatment variable T . “Influential” refers to our approach as presented in this paper; “Influential-10” also refers to our approach but with group (3) over-sampled at 10 times. In both figures, the horizontal axis represents the percentile x , and the vertical axis represents the difference between MT and MC , or $MT - MC$. Recall that $MT - MC$ gives UBR .

5.1. Traditional Approach

The traditional paradigm only uses records of contacted customers, i.e. TI , as it does not consider how one would behave without the contact. Using ARC, we observe that MT and MC are highly correlated, leading to a small $MT - MC$ as shown by “Traditional” in Figure 5. Indeed, MT is higher in the top percentiles, having a decreasing trend along the x -axis; however, MC behaves similarly. The high correlation suggests most responders in $T2x$ are also likely to respond given no contact, i.e. the majority of responders in $T2x$ are voluntary buyers.

SAS EM Tree performed poorly with the highly imbalanced class distribution (in TI only 5.4% is positive). [8, 5] suggested under-sampling the majority class may help, so we under-sample at different rates and construct a model for each rate. The model with the best lift index [8] is selected as the final model. The result is similar to ARC, as shown in Figure 6.

The experiments clearly show that models built by the traditional paradigm favor decided customers. An effective model should target undecided so that more buyers can be generated with fewer contacts. Evidently, MT by itself is not an effective evaluation criterion. Even if MT is high, $MT - MC$ may actually be small.

5.2. Lo’s Approach

Following Lo’s approach with ARC, we observe $MT -$

MC has increased slightly compared to the traditional approach, as show by “Lo’s” in Figure 5. Note that the treatment variable T is not a good indicator to differentiate between classes as $T = 1$ is similarly associated with both classes. This means that the model is likely to ignore T . When this occurs, Lo’s approach becomes similar to the traditional approach. Quite often, the predicted probability given $T_i = 1$ and the predicted probability given $T_i = 0$ is the same, making it ineffective to rank customers based on the difference of the two probabilities. SAS EM Tree produces worse result as it is even more similar to the result of Section 5.1, as in Figure 6.

5.3. Influential Marketing

By combining groups (1) and (4) to form PC and groups (2) and (3) to form NC , the percentage of positive records in the training data to 14.6% from the original 5.4%. With a more balanced class distribution, we did not need to under-sample when using SAS EM Tree. The results produced by SAS EM Tree and ARC are similar.

Using our approach, we observe MT and MC have a reverse trend up to the top 50% percentile: MT decreases while MC increases. The larger $MT - MC$ (given by “Influential”), demonstrates the model’s effectiveness at targeting undecided customers. Note that the behaviors of decided and non customers are the same in $T2x$ and $C2x$. While the undecided customers in $T2x$ contribute to MT (because they were contacted), the undecided customers in $C2x$ do not contribute to MC (because they were not contacted). Thus, the undecided customers solely accounts for the difference $MT - MC$. Additionally, we observe that both MT and MC have decreased slightly compared to the results in Section 5.1 and 5.2. It is our intention to not contact decided customers. As a model focuses on undecided customers, fewer decided customers are selected, which explains the decrease in both MT and MC .

Due to the small number of (3), the effect of having (3) in NC to “remove” decided customers from PC may be limited. Therefore, (3) is over-sampled 5, 10, and 20 times in the training data. The models built from these over-sampled training sets have all shown a further increase in $MT - MC$ (see “Influential-10”). The increase is mainly due to the further decline in MC , suggesting even fewer decided customers were selected. This is exactly what we want to achieve by over-sampling (3) in NC .

The decreasing trend of both “Influential” and “Influential-10” indicates that our approach does in fact focus on undecided customers. On the other hand, the nearly flat trend of the traditional approach suggests that it fails badly at targeting undecided customers.

Our approach is significantly better in the top 20% for ARC and the top 40% for SAS EM Tree, considering the

generally very low response rates. As a campaign typically has a small marketing percentile, e.g. 10%–30%, a campaign shall benefit from our approach.

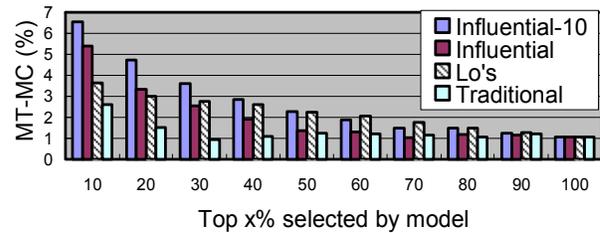


Figure 5. PIC comparison, ARC.

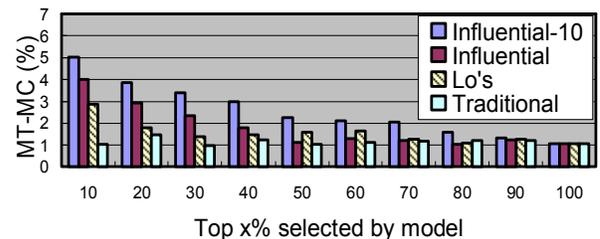


Figure 6. PIC comparison, SAS EM Tree.

References

- [1] P. Domingos, Metacost: A general method for making classifiers cost sensitive, In *SIGKDD*, 1999.
- [2] P. Domingos and M. Richardson, Mining the network value of customers, In *SIGKDD*, 2001.
- [3] P. Giudici, *Applied Data Mining*, John Wiley & Sons, 2003.
- [4] J.A. Hanley and B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143(1), 1982, pp. 29-36.
- [5] N. Japkowicz, The class imbalance problem: Significance and strategies, In *ICAI*, 2000.
- [6] KDD-Cup-98. <http://www.kdnuggets.com/meetings/kdd98/kdd-cup-98.html>.
- [7] R. Kohavi, D. Sommerfield, and J. Dougherty. Data mining using MLC++: A machine learning library in C++. (<http://www.sgi.com/tech/mlc>)
- [8] C.X. Ling and C. Li, Data mining for direct marketing: Problems and solutions. In *SIGKDD*, 1998.
- [9] V.S.Y. Lo, The true lift model – A novel data mining approach to response modeling in database marketing, *SIGKDD Explorations*, 4(2), 2002, pp. 78-86.
- [10] SAS Institute. The SAS system. (<http://www.sas.com>)
- [11] K. Wang, S. Zhou, Q. Yang and J.M.S. Yeung, Mining customer value: From association rules to direct marketing, *Journal of Data Mining and Knowledge Discovery*, 11(1), 2005, pp. 57-80.