Combining data mining and optimization for campaign management

C. Vercellis

Dipartimento di Ingegneria Gestionale, Politecnico di Milano, Italy.

Abstract

The process of marketing campaign optimization takes as input a set of offers, a set of customer segments and a set of communication channels, and determines the most profitable combinations by which offers should go to segments over channels, taking into account a set of constraints for the campaign. In this paper, we argue that the combination of data mining techniques with optimization models can lead to more effective approaches to campaign management, and to an overall improved support for marketing decision makers. Given a specific marketing task, such as customer retention or acquisition, a class of multivariate splitting rules, in which an optimization problem is solved at each node in the tree, is proposed in the first stage to derive a set of interesting segments, by scoring the customers or the prospects. Then, in the second stage of our procedure, a mixed integer optimization model is formulated and solved for the overall campaign optimization, taking as input the customer segmentation derived in the first stage, together with the set of offers defined by the marketing managers, and the constraints on the limited resources available for the whole campaign.

1 Introduction

Data mining techniques are emerging as a key tool for implementing advanced marketing approaches, such as one-to-one marketing [7] or customer relationship management (CRM) [2]. There are usually a wide variety of ways in which a company may interact with its customers and prospects. Whenever a new campaign is planned, marketing is faced with many possible choices along three main dimensions: campaign targeting, that is to whom the offer should be made;
campaign content, that is which offer should be made to each target; campaign broadcasting, that is through which communication channels the offer should be made. The process of marketing campaign optimization takes a set of offers, a set of customer segments and a set of communication channels, and determines the most profitable combinations by which offers should go to segments over channels, taking into account a set of constraints for the campaign.

In this paper, we argue that the combination of data mining techniques with optimization models can lead to more effective approaches to campaign management, and to an overall improved support for marketing decision makers. Given a specific marketing task, such as customer retention, acquisition, cross sell and up sell, classification trees [6] can be used in the first stage to derive a set of interesting segments, by scoring the customers or the prospects. Although in principle any algorithm for building classification trees could be used, we propose to use a class of multivariate splitting rules recently developed [5], in which an optimization problem is solved at each node in the tree, based on the structural minimization principle derived for support vector machines [4], [8]. Then, in the second stage of our procedure, a mixed integer optimization model is formulated and solved for overall campaign optimization, taking as input the customer segmentation derived in the first stage, together with the set of offers defined by the marketing managers, and the constraints on the limited resources available for the whole campaign. Although an extensive testing of our solution framework has not yet been performed, the two steps procedure has been used to model some campaign optimization real-world problems arising in the telecom and finance industries, showing encouraging results.

The paper is organized as follows: Section 2 describes the business scenario in which campaign optimization takes place; in Section 3 the segmentation procedure based on classification trees is presented; Section 4 is devoted to the formulation of a constrained integer programming capturing the logic of campaign optimization; finally, Section 5 pinpoints the main advantages of the outlined approach, also discussing some future extensions.

2 Business scenarios for campaign management

With the advent of Customer Relationship Management practices marketing activities have been profoundly redesigned. In particular, the traditional mass market approach based on interactions with a large number of customers simultaneously, using broadcast channels such as TV or magazine advertisements, has been replaced by more focused marketing activities aimed at addressing individual customer needs. The switch in attitude for the marketing department can be summarized by the new mantra of CRM oriented companies: select the right offer to the right customer at the right time.

This marketing approach involves identifying and understanding from past data the relevant customer patterns as well as designing customized offers for each customer, or more often for each segment of customers, that correspond to
those patterns. For example, a pattern might be that customers of a mobile telephone service provider who subscribed with the company by less than 15 months, spend less than 40 € monthly, and more than 30 € in calls directed toward a different mobile operator, have a very high likelihood of 0.89 to churn, that is to switch to a competitor in the next future. Thus a marketing manager could use this pattern to identify individuals to whom the next retention offer should be addressed, providing an opportunity to increase the customer lifetime value without wasting offers targeted to uninteresting people.

The reach and growing set of methodologies and tools denoted currently as CRM analytics – statistics, data mining, optimization, visualization – may help marketing managers to systematically analyze customer needs, preferences, and service requirements and learn how to gain new customers and increase customers loyalty with personalized marketing programs. These decision support tools have to access a marketing data mart, or simply a customer database, that integrates all customer-related activities over various communication channels such as Internet, call center, mailing, personal contact or fax, and can supply essential information for planning one-to-one marketing activities with customer specific product offers.

There are four main marketing tasks which have emerged within CRM, and which will be generally addressed in this paper. First is customer acquisition, in which the past history of offers and promotions addressed to prospects is available together with the information on which prospects accepted the offer and became actual customers, that is who bought the proposed product or utilized the offered service. In this case, the aim of the analysis is to identify the most promising segments to which new offers should be addressed to maximize the overall return on the marketing investment (ROMI). To each segment is associated a score representing the likelihood that any individual belonging to that segment will adhere to the proposed marketing offer.

The second marketing task is customer retention, which consists of identifying those actual customers who have a high risk of churn, that is of leaving the company to get products or services from a competitor, and discriminating them from loyal customers. In this case the predictive model will learn from the past history of transactions for each customer, together with the information of who left the company by churning. Again, the identification of segments with high likelihood of churn will help managers to best target their retention initiatives.

Finally, the third and forth CRM tasks we consider are aimed at cross selling – selling additional products or services to individuals who already are our customers – and up selling – trying to make a customer switch from a product or service she/he already buys from us to a different one of higher value and profitability. In this case too, it is self-evident how the marketing department could benefit from a segmentation and scoring analysis which pinpoints those customers who are more likely to accept specific offers for cross selling and up selling.
The described approach relies on predictive data mining models to target customers for offers. These models accurately estimate the probability that a customer will respond to a specific offer and can significantly increase the response rate to a product or service offering. However, simply knowing a customer likelihood of responding to a particular offer is not enough when a company has several products or services to promote, alternative offers for each segment of customers, multiple communication channels for broadcasting its offers, not to mention other business constraints to be considered in developing the marketing plans.

Here is when optimization models come into place. In a very broad perspective, optimization can be seen as a modeling paradigm particularly useful when dealing with a decision making problem in which a usually large set of alternative actions are competing against each other for a set of scarce resources available in a limited amount. Each of the alternative actions is characterized by different costs and revenues, and the model is required to determine the best combination of alternative actions which maximizes the total difference between revenues and costs, subject to the constraints on resource availability and other existing business constraints. It should be already clear from this short description how the campaign management problem would properly fit within the outlined framework. However, the next sections will provide the technical details about the proposed two steps approach.

3 Market segmentation by multivariate classification trees

We assume that the decision maker faces a specific marketing task, such as acquisition, retention, cross and up selling, and she/he has devised a set of \( I \) alternative actions to be applied to the customer base. To simplify our terminology, the universe of the individuals to whom the marketing effort is addressed will be generally referred as “customer base”, even if in the case of the marketing acquisition task actions are actually directed to prospects. We can always suppose that one of the actions is simply “do nothing” and corresponds to the absence of any marketing effort for the selected customers. Hence without loss of generality it can be stated that among the \( I \) available marketing actions one and only one should be selected for each customer. We also assume that for each customer or prospect a set of \( P \) attributes are known, representing potentially relevant factors such as demographic information, past transactions and interactions with the company, responses to previous promotions.

As a first step towards campaign optimization, \( I \) segmentations of the customer base can be derived, each corresponding to one of the marketing actions. For each action \( i, i = 1, 2, \ldots I \), segments would ideally aggregate those customers whose response to action \( i \) is likely to be the same and whose key attributes have similar values.
Now there are two alternative cases: if action $i$ has been already applied in the past, among the available attributes we can assume that one is a binary indicator which takes on value 1 if the customer response to action $i$ has been positive, and 0 otherwise. In this case the segmentation corresponding to action $i$ can be derived by solving a classification problem through data mining techniques, as described below.

Alternatively, if no action similar to $i$ has ever been applied in the past, we suppose that the corresponding segmentation is generated a priori by the marketing decision maker based on her/his experience and knowledge of the customer base, by means of traditional marketing segmentation methods.

Suppose in the sequel that data on past responses to action $i$ are available. A classification problem [1] is defined as follows: Given a set of $M$ labeled couples $(y_m, t_m)$, $m = 1, 2, \ldots M$, where $y_m$ represents a $V$-dimensional vector of attributes and $t_m$ the value of the target binary indicator, one is required to find a function $f(y_m) \mapsto t_m$ which predicts the target $t_m$ on the basis of the explanatory attributes $y_m$ either correctly, or minimizing some measure of error.

Thus we can apply a classification trees algorithm in order to derive a segmentation of the customer base with respect to the target variable represented by the response to action $i$. The reader is referred to [6] for a survey of methods based on classification trees. Although any classification trees algorithm could be adopted for deriving the segmentation and developing the first stage of the proposed approach to campaign optimization, we suggest to employ a new technique developed in [5]. It is a hybrid method, in which multivariate splitting rules are derived at each node within the tree, by solving a mixed integer programming problem representing a modified version of a Support Vector Machine (SVM) (the reader is referred to [4], [8] for the general concept of SVM). The advantage of the proposed method over competing techniques is twofold [5]: first, the empirical prediction accuracy recorded on extensive comparisons on benchmark tests appears consistently superior; further, the number of classification rules generated, and therefore the number of derived segments, tends to be smaller than for traditional univariate splitting rules.

Now, suppose that a segmentation $S_i$ has been generated for each marketing action $i$, either by classification trees algorithms, or by traditional a priori methods when previous data on past response to $i$ are unavailable, as described. At the end of the segmentation process, we end up with $I$ partitions of the customer base, where each partition is made by a number of exhaustive and mutually exclusive sets $S_i = \{E_{1i}, E_{2i}, \ldots E_{Pi}\}, i = 1, 2, \ldots I$.

As the next step in our procedure, we propose to derive a further $(I+1)$-th segmentation of the customer base which depends from the specific marketing task faced by the decision maker. For instance, if the analysis is directed towards
customer retention, there will be a binary target variable which equals 1 if the customer has churned and 0 otherwise. Or, if customer acquisition is the goal of the analysis, the target attribute equals 1 if the prospect became a customer and 0 otherwise. Consequently, the \((I+1)-th\) segmentation can be obtained again by solving a classification problem, using the same algorithm utilized to achieve the first \(I\) segmentations.

Finally, a further \((I+2)-th\) segmentation of the customer base should be derived from customer value or profitability. The reason is that marketing actions should be assigned to customers on the basis of three factors: the likelihood of responding to the \(i\)-th action, the likelihood that they are positive targets of the investigated marketing task (for instance, that they are churning customers or promising prospects) and the value of the customer. Even if a customer is likely to churn and is likely to respond to a retention offer, we don’t want to carry out any action if her/his monthly bill on products and services acquired by our company is close to zero. This latter \((I+2)-th\) segmentation can be generated by simply partitioning the customer base in disjoint classes with respect to the overall amount of money spent by the customer, or her/his profitability.

At the end of the whole outlined procedure, we have derived \(I+2\) segmentations of the customer base. Consider now, as the last step, the resulting segmentation which can be generated by intersecting all the subsets of customers appearing along the \(I+2\) collections. In this way we can get a final segmentation \(R = \{D_1, D_2, ..., D_K\}\), containing \(K\) subsets of customers, where the cardinality \(K\) is bounded above by the product of the cardinalities of the \(I+2\) previous segmentations. In fact, it should be usually much lower than this limit, because a segment elimination and aggregation procedure can be applied to avoid unnecessary negligible or irrelevant segments. The segmentation \(R\) has the following properties:

1. each segment is homogeneous with respect to the likelihood of responding to action \(i\), for every \(i=1, 2, ..., I\);
2. each segment is homogeneous with respect to the value or the profitability of the customers;
3. each segment is homogeneous with respect to the target attribute which corresponds to the investigated marketing task, such as customer acquisition, retention, cross sell or up sell.

4 An optimization model for campaign management

This section is devoted to the description of a mathematical programming model which captures the logic of campaign optimization. According to the outcome of the first stage of the proposed procedure, as described in Section 3, we assume
from now on that a final segmentation $R = \{D_1, D_2, \ldots, D_K\}$ is available, possessing properties 1-3 expressed above. The following notation is then introduced:

- $d_k$: number of customers belonging to segment $D_k$
- $C_i$: cost of the marketing action $i$
- $p_{ik}$: probability that a customer belonging to segment $D_k$ will positively respond to marketing action $i$
- $V_{ik}$: value of each customer belonging to segment $D_k$ accepting marketing action $i$.

We incidentally notice that the probability $p_{ik}$ of positively responding to action $i$ is given a posteriori by the likelihood estimated for segment $D_k$, when for action $i$ there exists a past data history; otherwise it must be estimated a priori by marketing analysts for newly conceived marketing actions.

If the marketing task represents a retention analysis, we need a further cost parameter:

- $L_k$: loss due to each churning customer belonging to segment $D_k$.

The return on the marketing investment (ROMI) can then be calculated, according to the following expressions. The first case refers to acquisition, cross sell and up sell analysis:

$$ROMI_{ik} = V_{ik} = -d_k C_i + d_k p_{ik} V_{ik}$$  \hspace{1cm} (1)

For a retention analysis the ROMI expression includes an additional cost term:

$$ROMI_{ik} = V_{ik} = -d_k C_i + d_k p_{ik} V_{ik} - d_k (1 - p_{ik}) L_k$$  \hspace{1cm} (2)

As discussed in Section 2, in the context of campaign optimization the decision maker is faced with the following problem: given a set of alternative decisions and a set of customer segments, determine the optimal allocation of actions to segments, in such a way to maximize an objective function expressing the ROMI, subject to a specified set of constraints assigned to the marketing plan. The most obvious constraints are concerned with the availability of limited resources required to implement the different marketing actions. More specifically, we assume that there are:
We can now define the capacity parameters of the optimization model:

- $W_r$: availability of each global resource $r$
- $F_{ks}$: availability of each resource $s$ dedicated to segment $D_k$
- $Q_{ij}$: availability of each resource $j$ dedicated to action $i$
- $w_{rk}$: absorption of global resource $r$ if action $i$ is applied to segment $D_k$
- $f_{ks}$: absorption of dedicated resource $s$ if action $i$ is applied to segment $D_k$
- $q_{ij}$: absorption of dedicated resource $j$ if action $i$ is applied to segment $D_k$.

The decisions associated to campaign optimization can be expressed as binary variables:

$$ x_{ik} = \begin{cases} 1 & \text{if action } i \text{ is applied to segment } D_k \\ 0 & \text{otherwise} \end{cases} \quad (3) $$

We can now formulate the campaign optimization problem (COP) as a capacitated integer programming model:

$$ \max ROMI = \sum_{i=1}^{J} \sum_{k=1}^{K} v_{ik} x_{ik} \quad (COP) $$

s.t. \quad \sum_{i=1}^{J} \sum_{k=1}^{K} w_{rk} x_{ik} \leq W_r \quad r = 1, 2, \ldots, R \quad (4) $$

$$ \sum_{i=1}^{J} f_{ks} x_{ik} \leq F_{ks} \quad k = 1, 2, \ldots, K ; \ s = 1, 2, \ldots, S \quad (5) $$

$$ \sum_{k=1}^{K} q_{ij} x_{ik} \leq Q_{ij} \quad i = 1, 2, \ldots, I ; \ j = 1, 2, \ldots, J \quad (6) $$

$$ \sum_{i=1}^{J} x_{ik} = 1 \quad k = 1, 2, \ldots, K \quad (7) $$
Constraints (4), (5) and (6) express the capacity constraints for global resources, resources dedicated to segments, and resources dedicated to actions, respectively. Constraint (7) is a multiple choice condition, stating that for each segment exactly one action has to be selected.

Problem COP is expressed in a form known in the literature as Generalized Assignment Problem (GAP) [3], [9]. This type of problems have been extensively investigated to derive both exact and approximate solution algorithms. We have resorted to two different classes of methods for solving problem COP:

- an approximate algorithm based on a truncated branch and bound method. In this case the CPLEX optimization library was utilized, and the parameters available to the users were the driving factor for achieving the approximate algorithm, by indirectly letting the branch and bound scheme to explore only a partial portion of the whole enumeration tree;
- a heuristic method based on an initial sequential assignment, followed by a subsequent local improvement scheme, along the lines described in [9].

The whole proposed procedure has been applied to different campaign optimization problems arising in real world contexts, in the telecom and finance industries. Although the number of individuals originally contained in the customer base were in the range between 300,000 and 1,000,000 for the different marketing tasks considered, the segmentation drastically reduced these huge numbers to manageable sizes. For instance, in a retention analysis modeled there were 4 alternative actions considered, including the option of “no-action”. The preliminary segmentations for each action and for the retention analysis led to a number of segments ranging between 5 and 15, with 4 segments of profitability. The final segmentation, obtained by intersecting the 6 previous segmentations, led to around 15,000 segments, after some elimination of segments of negligible size and aggregation of homogeneous segments. The corresponding optimization model resulted therefore in 60,000 variables and 200,000 constraints. The approximate solutions were generated by both approaches in the order of minutes. In general, our tests showed a better accuracy obtained through the first technique, i.e. the truncated branch and bound approach, whereas the local search method was generally faster. Notice that, due to the original size of the customer base, the formulation of the optimization problem COP at the level of individual customers would have been much harder to solve, eventually leading to much inaccurate solutions. In particular, to further validate the proposed two steps procedure, the solutions achieved through the campaign optimization
model have been compared to those obtained using a myopic heuristic, which essentially consists of assigning to each segment the marketing action locally maximizing the ROMI, without violating any of the capacity constraints. The comparisons showed that the solutions achieved by our model improved significantly the total ROMI with respect to myopic solutions.

5 Conclusions

We have considered in this paper the campaign optimization problem, addressing a generic marketing task such as customer acquisition, retention, cross sell and up sell. A modeling framework for its solution has been proposed. It takes advantage of the combined use of data mining techniques and mathematical programming methods. Specifically, data mining classification algorithms, based on a recently proposed technique of classification trees with multivariate splitting rules, are utilized for generating a segmentation of the customer base. This has the advantage of dramatically reducing the size of the optimization problem formulated in the second stage of the procedure, by addressing segments instead of individual customers. Furthermore, the segmentation also helps the marketing managers in determining the parameters of the models and interpreting the results obtained. In the second stage of our approach an integer programming model is formulated, in the form of a generalized assignment problem, for which several efficient approximate solution techniques have been devised. Although an extensive testing of our solution framework has not yet been performed, the two steps procedure has been used to model some campaign optimization real-world problems arising in the telecom and finance industries, showing encouraging results. In particular, the solutions achieved through the campaign optimization model appeared definitely superior in terms of ROMI with respect to those obtained using a myopic heuristic which consists of assigning to each segment the marketing action which locally maximizes the ROMI.

Future extensions of the proposed approach will proceed along two main directions: the extension of the model to capture a multiple time horizon, in which also the timing of the actions should be selected, in the perspective of continuous marketing cycles; second, an extensive testing to further validate the approach.

References


