Targeting High Value Customers While Under Resource Constraint: Partial Order Constrained Optimization with Genetic Algorithm

Geng Cui a,b,⁎ & Man Leung Wong c & Xiang Wan d

a Dept. of Marketing and International Business, Lingnan University, Tuen Mun, Hong Kong
b Yunshan Scholar, School of Management, Guangdong University of Foreign Studies, Guangzhou, Guangdong Province 510006, China
c Dept. of Computing & Decision Sciences, Lingnan University, Tuen Mun, Hong Kong
d Department of Computer Science, Baptist University, Kowloon, Hong Kong

Abstract

To maximize sales or profit given a fixed budget, direct marketing targets a preset top percentage of consumers who are the most likely to respond and purchase a greater amount. Existing forecasting models, however, largely ignore the resource constraint and render sup-optimal performance in maximizing profit given the budget constraint. This study proposes a model of partial order constrained optimization (POCO) using a penalty weight that represents the marginal penalty for selecting one more customer. Genetic algorithms as a tool of stochastic optimization help to select models that maximize the total sales at the top deciles of a customer list. The results of cross-validation with a direct marketing dataset indicate that the POCO model outperforms the competing methods in maximizing sales under the resource constraint and has distinctive advantages in augmenting the profitability of direct marketing.

© 2014 Direct Marketing Educational Foundation, Inc., dba Marketing EDGE.

Keywords: Direct marketing; Profit maximization; Partial order function; Constrained optimization; Genetic algorithms; Return on investment

Introduction

Resource constraint such as a limited budget is prevalent in marketing operations. How to maximize sales or profit given a limited budget is a common challenge for marketing managers, making constrained optimization a research topic of intensive interest. This problem is especially relevant and acute in direct mail marketing (Gönül and Shi 1998), where managers usually have a fixed budget for a campaign, thus can only contact a pre-set percentage of customers, e.g., the top percentile of a customer list (e.g., top 10%). Conventional models of customer selection and forecasting, however, are mostly based on the ordinary least square (OLS) approach, and do not consider such constraint or focus on the high value customers. While researchers have attempted to arrive at better estimates of response probabilities and expected profit (e.g., Rao and Steckel 1995), most customer selection models do not address the problem of maximizing sales or profit given a specific mailing depth. Since customer response probabilities may have a low or even negative correlation with the expected purchase amount, models with a high accuracy of predicting customer response may not generate maximum sales. Although such models can be calibrated for a certain budget or mailing depth, directly incorporating the resource constraint is preferred and can generate superior result (Prinzie and Van den Poel 2005).

These models following the ordinary least squares (OLS) approach typically produce a complete order of all cases based on their expected sales or profit. However, a complete order of all cases is neither necessary nor does it guarantee superior result because it focuses on fitting a model to the overall population using the conditional mean to generate the expected profit for the “average” customer, while most of a company’s profit usually comes from a small group of high value customers, which are ignored as outliers. Although a few studies have investigated

⁎ Corresponding author at: Dept. of Marketing and International Business Lingnan University 8 Castle Peak Road Tuen Mun, Hong Kong.
E-mail address: gcui@ln.edu.hk (G. Cui).

http://dx.doi.org/10.1016/j.intmar.2014.09.001
1094-9968/© 2014 Direct Marketing Educational Foundation, Inc., dba Marketing EDGE.
the problem of resource constraint in direct marketing and adopted the constrained optimization approach (Bhattacharyya, 1999; Yan and Baldasare, 2006), they have adopted different dependent variables and used objective functions that are neither effective for profit maximization nor efficient for model optimization. Thus, the lack of a plausible objective function and an efficient and robust solution to optimize the selection of models presents a significant challenge.

We formulate the problem of targeting the high value customers given a fixed budget as one of constrained optimization using the number of customers to be selected as the constraint. For an optimal solution that satisfies the constraint while maximizing the total profit, we propose a model of partial order constrained optimization (POCO) that directly integrates the mailing depth (i.e., a certain percentile) as the constraint and adopt a penalty weight that represents the marginal penalty for selecting one more customer from a list. This allows us to convert the problem of complete order into a simpler partial order optimization problem with soft constraint. To ensure the efficiency and robustness of the solution, we adopt an advanced genetic algorithm (GA) with Cauchy mutation as a method of stochastic optimization to estimate the parameters and to select the models that maximize revenue at the top percentiles of a customer list. We apply the POCO model to a large direct marketing dataset and compare our results with those of competing methods at different constraint levels. The results indicate that the POCO model outperforms the nonparametric boosting method of AdaC2 and the constrained DMAX model (Bhattacharyya, 1999), proves to be a viable solution to the problem of constrained optimization in direct marketing. The POCO model can help managers augment the return on marketing investment under resource constraint.

Relevant Literature

Customer Selection and Forecasting

Despite the technological advances, direct mail remains a popular marketing vehicle due to its ease of production, personal touch and quick response. The primary objective of modeling consumer responses in direct marketing is to identify customers who are the most likely to respond. This requires the researchers to produce a greater number of true positives in the upper deciles in the testing or rollout data. Direct marketing forecasting has focused on predicting consumer purchase probability using a variety of consumer behavior and background variables. Sophisticated models include the Customer Response-Based Iterative Segmentation Procedures (De Sarbo and Ramaswamy, 1994), tree models such as CART and CHAID (Haughton and Oulabi, 1997), the beta-logistic model (Rao and Steckel, 1995), and the hierarchical Bayes random-effects model (Allenby, Leone, and Jen, 1999). When comparing the alternatives, a better model should correctly classify a greater percentage of responders at the top deciles of a customer list. Since customers with higher probabilities to respond may have a lower expected purchase amount, models with a high accuracy of predicting customer responses may not generate maximum sales or profit.

Another stream of research aims at improving the profitability of direct marketing. Bult and Wansbeek (1995) propose a profit maximization approach to direct marketing — sending catalogs to customers as long as their expected profit is greater than the cost or the expected marginal profit is greater than zero. This profit maximization approach, which includes all the customers with a positive expected marginal profit, is not realistic given a fixed budget. Rao and Steckel (1995) propose a model that combines the predicted probabilities and the expected profit to re-rank the customers to float the high value customer to the top of a file to improve profitability. Meanwhile, researchers have adopted the lifetime value (LTV) approach to select high value customers based on their long-term contribution of profit to a company in the future (Fader, Hardie, and Lee, 2005; Venkatesan and Kumar, 2004).

Recently, researchers have adopted joint distribution models such as neural networks and Bayesian networks for customer selection and revenue forecasting in direct marketing (Bose and Chen, 2009; Cui, Wong, and Lui, 2006; Zahavi and Levin, 1997). A number of scholars proposed various machine learning methods such as genetic algorithms and evolutionary programming to optimize the performance of customer selection models (Bhattacharyya, 1999; Cui, Wong, and Lui, 2006; Jonker, Piersma, and Van den Poel, 2004). Cui, Wong, and Wan (2012) proposed cost-sensitive learning that places greater weight on the high value customers in the customer selection process. Miguéis, Benoit, and Van den Poel (2014) applied the Bayesian quantile regression for selecting customers. Despite the low cost of contacting customers, low response rate is common for direct marketing campaigns, for instance, around 5% for catalog mailings, making improving profitability and return on investment a top priority. Even a small percentage of improvement in predictive accuracy can result in a significant increment in profitability.

To date, customer selection models including the OLS regression and its variants (logit) have largely followed the principle of mean squared error (MSE) and focus on the mean estimate for the entire population (Hao and Naiman, 2007). Under a typical condition, i.e., when the errors are assumed to have precisely the same distribution, OLS regression is sufficient to describe the relationship between the covariates and the response distribution (Fitzenerger, Koenker, and Machado, 2002). In addition, conditional-mean models lead to estimators that possess attractive statistical properties, are easy to calculate, and straightforward to interpret (Hao and Naiman, 2007). OLS is efficient and powerful for generating the predicted profit of customers and can “generally” distinguish the high value customers from low value customers.

However, direct marketing typically has a low response rate. With a 5% response rate, for instance, most customers (i.e., 95%) are non-buyers with a negative profit. Among the small portion of buyers (e.g., 5%), a majority of them (e.g., 4%) contributes very small profit while the minority (1% or less) contributes high profit. In general, customer profits exhibit along-tailed, skewed distributions with the high-value customers as the outliers as shown in Fig. 1. This has been referred to as the so-called 80/20 rule in that 80% of a company’s profit often comes from 20% of its customers. Thus, firms mostly depend on a small set of
customers for its profit, while the rest of customers have a low, and in some cases, negative profit or loss (Mulhern 1999). This type of distribution appears to be prevalent in marketing.

In light of the majority non-buyers and the long-tailed distribution of profit data, OLS regression has several limitations for targeting high value customers. Given the highly skewed customer profit data, several critical model assumptions of OLS regression including distribution normality and homoscedasticity are often not tenable (Hao and Naiman 2007; Mulhern 1999). Violation of such an assumption results in biased estimates, e.g., p-values and confidence intervals. Furthermore, both the segment of high value customers and the budgetary constraint in direct market point to a non-central location, somewhere in the upper tail (percentile). When summarizing the response variable for fixed values of predictor variables, the conditional model by definition focuses on the average consumers (e.g., average profit = $5) and cannot be readily extended to non-central locations (Hao and Naiman 2007), thus making it unsuitable for targeting the high profit customers. As the low value customers greatly outnumber the high value ones, such distribution-bound methods usually have a bias in favor of the majority of the nonbuyers or low value customers because their goal is to construct a function to make as less errors as possible when predicting the profit of new samples.

In light of the long-tailed, highly skewed profit distribution, OLS regression treats the upper tail of high value customers as outliers as shown in Fig. 2. These outliers of high value customers are in fact the target of direct marketers (Hao and Naiman 2007). The conventional way of handling outliers in OLS regression is to eliminate them. As a result, the predicted values for high value customers would be much lower than their true values (the blue line) and seriously “under-estimate” the upper tail of high value customers. However, the outliers of high value customers are of essence for targeted marketing to augment profitability. By eliminating the outliers from analysis, OLS regression cannot arrive at an accurate measure of the high value customers.

Since OLS-based methods do not focus on the high value customers and do not consider the resource constraint, they may produce sub-optimal results in maximizing profitability given the resource constraint. Until recently, the issue of resource constraint has been largely neglected in the direct marketing literature because constrained optimization problems are inherently complex and difficult to solve.

Resource Constraint

Resource constraints such as budget, manpower or manhours are prevalent in marketing and business operations. Recently, there has been a growing body of literature on resource allocation and constrained optimization in marketing operations, including marketing mix strategies and resource allocation (Graham and Ariza 2003; Gupta and Steenburgh 2008), new product development (Peng, Li, and Wan 2012; Reich and Levy 2004), and services marketing (Thomas, Teneketzis, and Mackie-Mason 2002). In the direct marketing and CRM literature, researchers have emphasized the problem of resource constraint and suggested constrained optimization as a solution (e.g., Mulhern 1999). Bitran and Mondschein (1996) consider the problem of constrained optimization in the case where companies manage their cash flow while incorporating the financial impact of carrying inventory. Given resource constraint, Gönül and Shi (1998) proposed a dynamic programming model that can help selectively mail to customers based on their purchase history to optimize the revenues from customers over time.

To date, a few researchers have treated customer selection for target marketing as a problem of resource constraint or have directly incorporated this constraint in generating models to maximize sales given such constraint. Bhattacharyya (1999) applied a scoring function of expected profit and used a genetic algorithm to select models that maximize profitability at a given depth of file. The decile analysis indicates that the model has good performance as evidenced by the total profit at the top decile. However, the model was not as stable as initially believed. The total profit shows unstable performance through the deciles, i.e., profit values do not decrease steadily through the deciles. The “jumps” in several deciles indicate that it may
not be sufficiently robust for decision support. According to the author, although the DMAX model’s maximization approach focuses on the upper decile levels, it is necessary to develop models that can perform well at other deciles or constraint levels and to avoid local optima in the model selection process.

Prinzie and Van Den Poel (2005) recognized the problem of profit maximization given resource constraint and adopted the Weighted Maximum Likelihood Estimation (WMLE) to arrive at better estimates of the response probability of the top deciles. They used model fitness as a performance criterion, placed greater weights on those customers with the highest predicted probabilities, and performed a number of iterations of the WMLE to achieve more accurate parameter estimates for the top deciles. They compared the confidence intervals of the parameter estimates to establish the superiority of the proposed method. However, they did not specify the number of iterations thus lacking a suitable termination criterion. This method may be computationally intractable. They did not perform any cross-validation or comparison alternative methods. Although they attempted to improve the accuracy of predicting consumer response at the upper deciles, their study was not concerned with maximizing sales or profit given the budget constraint.

Since maximization of the true positive rate is different from the maximization of sales or profit, optimization of a classifier for predicting purchase incidents does not necessarily maximize sales. To solve this problem, Yan and Baldasare (2006) proposed a nonlinear scoring function that uses neural networks and the gradient descent on a sigmoid function to maximize the ROI, i.e., sales at a specific depth of file under the budget constraint. By comparison with several classification, regression, and ranking algorithms, they find that the new algorithm may result in improved ROI. However, complete ordering of all the cases and the gradient search method using the sigmoid function are computationally inefficient. Since the objective function is non-convex, the descent method is more prone to local optima. The predicted results are ranking scores instead of estimates of response or profit.

While these studies attempted to improve performance at a top percentile, they used different dependent variables, expected sales or profit by Bhattacharyya (1999) and Yan and Baldasare (2006) and response probability by Prinzie and Van Den Poel (2005). Only Bhattacharyya (1999) addressed the problem of resource constraint by directly incorporating the constraint in the model. Using this approach, model parameters are estimated in an iterative manner. At each iteration, the model with estimated parameters is applied to produce a complete order of all cases and to calculate the total sales/profit at the targeted percentile to determine whether the current estimation is better. Each iteration may take a few minutes. However, this presents a significant challenge for optimization or machine learning methods, especially with large datasets leading to very large search spaces of model parameters and thereby numerous iterations. Like many optimization models, it may take hundreds and sometimes up to thousands of iterations to reach the point of optimum. Since a complete order of all cases is required at each iteration, it may take many hours or days to converge even for a simple greedy search algorithm. For other more complex algorithms such as those based on tournament method that compares dozens of models at one time, it can be computationally intensive and even intractable.

In summary, an unconstrained model such as OLS regression is efficient in estimating conditional mean for the whole population, which may not be accurate for the outliers of high value customers. Moreover, it does not consider the resource constraint, at a level of which profit is to be maximized. On the other hand, if a constrained optimization method has to produce a complete order of all cases at each iteration, it is inefficient and subject to local optima. Thus, despite these efforts, finding a suitable objective function that focuses on the high value customers and developing an efficient and robust optimization solution remain significant challenges for maximizing sales or profit given the resource constraint.

Partial Order Constrained Optimization

To maximize profit given the resource constraint of a fixed budget, we propose to identify the high value customers using partial order constrained optimization (POCO). In the following sections, we elaborate on each component of this approach: 1) a partial order function to identify the high value customers, 2) penalty method for satisfying the resource constraint, and 3) stochastic optimization using genetic algorithm.

Partial Order Function to Identify the High Value Customers

In a typical situation, a marketing manager has a fixed budget that allows targeting only a specific percentile of high value customers to maximize sales. Any top percentile such as the first and second deciles may represent the resource constraint. While considering a specific level of constraint or mailing depth, let \( E = \{x_1, x_2, \ldots, x_N\} \) be the set of \( N \) potential customers and \( m(e), 1 \leq i \leq N \) be the amount of money spent by the customer \( e_i \). Suppose that a fixed budget allows only \( r \) or a top percentile of the customers to be solicited, one can estimate a function to predict their expected purchase amount or induce a ranking function to arrange the cases in a descending order according to their expected revenue. Then, the first \( \lceil N \ast \frac{r}{N} \rceil \) cases are selected in hope to maximize the total expected sales from the solicited customers. The conventional approach based on OLS regression estimates a linear function for the whole population, which may not be accurate for the outliers of high value customers. Moreover, it does not consider the resource constraint, at a level of which profit is to be maximized. This approach, however, focuses on the average customer and would make errors in predicting the high value customers. Moreover, it does not consider the problem of resource constraint, and thus renders suboptimal performance in maximizing profit given the constraint.

This and other full order models attempt to estimate the profit of each customer as accurately as possible in hope that the high profit customers will float to the top for the final selection. For the purpose of maximizing profit given the constraint, however, full ordering of all cases over-complicates the problem and in fact it is not necessary because more accurate ordering above or below the threshold (constraint) is immaterial and will not produce more profit above the threshold. For example, if we have 10 customers and a model’s
Consider a case where the top percentile of customers to be targeted is irrelevant. Using the above example, if the resource constraint allows to target only the top 20% of customers and one already identifies the top 2 customers, whether the order is $10, $9 or $10 is immaterial, while the order below the threshold (from $4 to $2) is also irrelevant. In fact, perfect ordering or ranking of cases above or below the threshold is of a lesser concern to managers. Moreover, full order models are rather cumbersome for constrained optimization and may result in poor performance given a large data set of hundreds of thousands of customers. Ordering is a branch of mathematics and computing science, which we cannot go into details (Gavanelli 2001).

To develop an objective function that is better suited for identifying the high value customers above a certain constraint, one may directly incorporate the resource constraint into the objective function by controlling the percentile of customers to be included in the group above the constraint level while maximizing the total expected sales or profit. To achieve this objective, instead of a full order function, we propose a partial order function to control the number of customers in the top percentile and make the process more efficient and reliable while achieving the objective of maximizing profit under the resource constraint. Partial order is a subbranch of the order theory in mathematics and computing science and can help solve constrained optimization problems (Gavanelli 2001).

For this constrained optimization problem in direct marketing, we suggest learning a scoring function $f$ that divides the $N$ cases into 2 classes: $U$ and $L$, to represent the “upper” group of high value customers above the constraint level for targeting and the “lower” group of low value customers below the constraint level, not to be contacted. Then, $r$ represents the resource constraint, i.e., the top percentile of customers to be targeted. The sizes of $U$ and $L$ should be $\left[ \frac{N}{r} \right]$ and $\left( N - \left[ \frac{N}{r} \right] \right)$ respectively. Consider a case $e_i$ in $U$, its $f(e_i)$ must be greater than the values of all cases in $L$. Moreover, the total expected benefit of all cases $e_i$ in $U$, i.e., from the high value customers, is to be maximized. In so doing, the budget constraint in terms of the depth of file is directly incorporated into the target function. To accomplish the objective of maximizing profit at a pre-determined constraint level, we formulate the learning process as the following constrained optimization problem,

**Find a scoring function $f$ that**

$$
\begin{align*}
\max Z &= \sum_{e_i \in U} m(e_i) \\
\text{s.t.} \\
\forall e_i \in U & \ni \exists e_j \in L (f(e_i) \leq f(e_j)) \\
|U| &= \frac{N}{r} \\
|L| &= N - \frac{N}{r}.
\end{align*}
$$

(1)

As such, the partial order function helps to avoid the cumbersome problem of complete ordering of all the cases and converts the problem of revenue or profit maximization into a much simpler problem, i.e., to determine whether a customer belongs to the high value group ($U$). Instead of ranking all the customers using their expected sales or profit, we only need to select those customers that are deemed to belong to the group of high value customers above the constraint level, i.e., the specific percentile. In contrast to a full order model, the partial order function tackles the problem of resource constraint directly and is simpler and more effective. According to the Occam’s razor principle, a simpler model often produces better results given the same objective.

**The Penalty Approach for Satisfying the Resource Constraint**

As the resource constraint of a fixed budget is incorporated into the partial order function, we need a solution to optimize the function. Since the orderings of all cases in $U$ and all cases in $L$ are insignificant to our objective, it is sufficient to learn the scoring function that achieves an optimal partial ranking (ordering) of cases and that maximize sales above the constraint level in $U$. Among the methods of constrained optimization such as the barrier method and Lagrange multipliers, the penalty approach is an effective and popular method for turning such a hard constraint problem into one of relaxed soft constraint and can solve the time-consuming optimization problem (Smith and Ceit 1996). The penalty method converts a constrained optimization problem into a series of unconstrained problems whose solutions converge to the solution of the original constrained problem. This is achieved by adding a term or condition to the objective function that consists of a penalty parameter and a measure of violation of the constraint. The penalty for violation is nonzero when the constraint is violated. It is zero in the region where the constraint is not violated or removed. In our case, we adopt the penalty approach to control the number of cases to be included above the constraint level, i.e., in the $U$ group, and formulate it as the following optimization problem,

**Find a scoring function $f$ and a threshold $\tau$ that**

$$
\max Z = \left\{ \sum_{e_i \in U} m(e_i) - \text{penalty\_weight} \cdot \max(0, |U| - \frac{N}{r}) \right\}
$$

(2)

where $U = \{e_i | f(e_i) > \tau\}$ and penalty\_weight represents the marginal penalty for selecting one more customer from the list. Since the budget limits the number of customers to a specific percentage or amount, the cost of not including a “high value” customer in the $U$ would be the potential sales lost. For example, a marketing program has been budgeted to select 20% of 10,000 customers. Suppose that the average sales of a customer in terms of dollar amount around the constraint level is $20, the penalty\_weight would be 20. The total penalty will be 0 if the number of solicited customers is smaller than or equal to 2,000 = 10,000 * 20%. On the other hand, the total penalty will be 8,000 = 20 * 400 if 2,400 customers are solicited. The specific penalty weight can be determined by the threshold value near the selected percentile or constraint level and be adjusted accordingly.
**Stochastic Optimization Using Genetic Algorithm**

To maximize the ROI of a marketing program constrained by a preset budget, we constrain the predictive model to a top percentile of a customer list, for instance, the second decile or 80% percentile. The task of constrained optimization can be formulated as a search problem to select models that can maximize sales using the above objective function of partial order and the penalty approach (Eq. (2)). In this study, we adopt a scoring function similar to linear regression and rely on the partial order function and the penalty method to determine whether a customer has a score that is higher than the threshold level, e.g., the top 20 percentile. Assuming that the model parameters are unknown, constrained optimization searches for a model with parameter estimates that generate maximum sales or profit at the constraint level.

Although the partial order method has simplified the search procedure and reduced the computation burden, the search space can still be extremely large given a big data set. Thus, strong heuristics are required to guide the process. Most conventional methods of optimization such as branch and bound algorithms are based on a greedy search strategy and generate a sequence of solutions until a good solution is found. However, such strategies and heuristics are not always effective because they may be trapped in local maxima. To overcome this problem, non-greedy strategies are preferred. Genetic algorithm (GA), which is well-known for its flexibility and ability to perform both exploitation of the most promising solutions and exploration of new models, represents one of the alternative stochastic search strategies (Goldberg 1989). To solve this problem of constrained optimization, we adopt genetic algorithm (GA) as a stochastic optimization algorithm to learn the parameters of the scoring function $f$ given the threshold $\tau$.

Genetic algorithm (GA) is one of the evolutionary computational methods originated in the field of artificial intelligence and has been increasingly adopted as a stochastic optimization method in many fields, including marketing. It emphasizes the behavioral linkage between parents (existing models) and their offspring (new alternative models). GA is suitable for difficult combinatorial and real-valued function optimization problems in which the fitness landscapes are rugged and have many locally optimal (i.e., suboptimal) solutions. GA models the relationship between the iterations of models in successive generations. It is a useful method of optimization when other techniques such as gradient descent or direct analytical methods are less effective. In the pseudo code shown in Fig. 3, $x_i$ is a vector of target variables evolving and $\eta_i$ controls the vigor of mutation of $\bar{x}_i$.

In general, to learn the optimal POCO model, the GA process takes four steps: 1) fitness value evaluation (steps 3 and 8), 2) parent selection (step 5), 3) crossover and mutation (steps 6 and 7 respectively), and 4) the replacement schema designed for parallel execution (step 9). For the first task, fitness value evaluation determines the “goodness” of individual models as one of the core components of the GA. After each evolution, the fitness value of each model in the current population is calculated. The result is passed onto the later steps of the learning process. Each model returns a fitness value by feeding the objective function $Z$ with the target variables of the individual model. This evaluation process usually consumes most of the computational time. For the second task, the selection process determines which models will be selected as parents to reproduce offspring. The selection operators in genetic algorithms are not specific, however the fitness value of a model usually induces a probability of being selected (Goldberg 1989). Tournament selection, which generally yields better performance for large populations, tends to be the mainstream selection strategy. Thus, the POCO model employs a stochastic tournament for selection of parent models. The second task of the selection operator focuses on searching promising regions of the solution space.

However, conventional genetic algorithms based on greedy search may suffer from local optima and lead to overfitting because new solutions that are not in the current population may not emerge. In order to escape from local optima and introduce larger population diversity, we employ single point crossover and the Cauchy mutation method to search for alternative models and better parameter estimates (Yao, Liu, and Lin 1999). Then, the selected parents will undergo the process of crossover and mutation to generate new offspring. Moreover, unlike binary-code GA used in the DMAX model, we use real-code GA with the new mutation operator that is more suitable for finding real-value parameters. Finally, we replace the population by comparing each model with its corresponding offspring. The one having the best fitness value replaces the model in the next iteration. The optimization process continues to iterate until the predefined termination criterion is satisfied.

**Comparison with Conventional Models**

Competing methods are subject to various assumptions and use different functions and criteria to produce the results that are used to sort the unseen cases. Table 1 provides a summary comparison of the key characteristics of the OLS approach and the POCO model. While OLS regression uses predicted profit to do a full sort of unseen cases, POCO uses the predicted scores of belonging to the high value group to do a full sort of the testing data. The OLS method estimates the model parameters based on model fitness measure. By directly incorporating the resource constraint into the objective function, the POCO approach selects a model that produces the maximum total profit. In the following sections, we apply the proposed POCO model to a large direct marketing dataset and discuss the results in comparison with those of alternative methods.

**Method and Results**

**Data and Measures**

To test the viability of the proposed approach and compare its performance with those of competing methods, we conducted experiments with a large direct marketing dataset from a U.S.-based catalog company, which is provided by the Direct
Marketing Educational Foundation. It has been used by researchers to test direct marketing models. The catalog marketing company carries numerous lines of merchandise including gifts, apparel, and consumer electronics. This dataset contains the records of 106,284 consumers from a recent promotion as well as their purchase history from the past twelve years. There are 361 variables for each customer. The most recent promotion sent a catalog to every customer in this dataset and had a 5.4% response rate, representing 5,740 buyers.

With customer response or purchase from the most recent promotion as the dependent variable, logistic regression was first used for data reduction using the forward selection method to select the variables for model building. We also use logistic regression as the baseline model and then compare the performance of the proposed POCO model with several benchmark models, including three unconstrained linear models: 1) an OLS regression model of profit, 2) the unconstrained linear GA model without penalty, i.e., the penalty weight = zero, 3) a nonparametric boosting model using AdaC2. Then, we compare the performance of the constrained POCO model with that of the constrained DMAX model using the scoring function by Bhattacharyya (1999). The detailed information for the genetic algorithm used in the POCO model is enclosed in Appendix A. For the setting specified in Appendix A (i.e. population size = 256, generation number = 2,000), the number of models evaluated was $256 \times 2,000 = 512,000$. The fitness values of models selected by POCO improve quickly to a very high point around the 10th generation/iteration and the solution converges at about 300 generations. Thus, the POCO model is effective in avoiding local maxima and very efficient.

To measure the performance of a model at different depths of file, researchers adopt the “response lift,” which is the ratio of the number of true positives to the total number of records identified by the classifier in comparison with that of a random

Table 1
Comparison of the OLS approach with the POCO model.

<table>
<thead>
<tr>
<th>Areas of comparison</th>
<th>The OLS approach</th>
<th>POCO model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumptions</td>
<td>Normality, homoscedasticity, one-model for all</td>
<td>Distribution-free, optimized solution by genetic algorithm</td>
</tr>
<tr>
<td>Handling of outliers</td>
<td>Ignored</td>
<td>Focus on high value outliers</td>
</tr>
<tr>
<td>Resource constraint</td>
<td>Not considered</td>
<td>Directly incorporated (r) in the model</td>
</tr>
<tr>
<td>Objective function</td>
<td>The conditional mean, to produce a full order of cases</td>
<td>Partial order, to perform binary classification (U vs. L)</td>
</tr>
<tr>
<td>Output (sorting criterion)</td>
<td>Expected profit</td>
<td>A score of belonging to the high value group</td>
</tr>
<tr>
<td>Estimation method and criterion</td>
<td>Maximum likelihood method based on model fitness (of the data)</td>
<td>Stochastic optimization of a linear model with the penalty function, based on the total profit at the constraint level</td>
</tr>
</tbody>
</table>
model at a specific decile of the file (Bhattacharyya 1999). The figure is then multiplied by 100. For instance, a lift of 200 in the first decile indicates that a specified model performs twice as well as a random model when the top 10% of customers are selected. Thus, we follow the convention of decile analysis in direct marketing research and compare the performance of models across the top deciles using cumulative lifts. Profit lift is measured in exactly the same way. In other words, the cumulative profit lift measures the improvement of proposed models in profitability over a random model at a specific depth of file. In this study, profit lift is a more important metric for model performance than response lift. Lifted profit is the actual amount of extra profit generated by the new method over that generated by a random method.

As a single split or hold-out validation is apparently insufficient as a validation procedure, we compare their performance using stratified 10-fold cross-validation (Mitchell 1997). For the stratified 10-fold cross-validation experiments, we followed the standard practice and split the complete data into 10 disjoint subsets using random sampling. We applied stratification of the original cases in the process so that these subsets of data have approximately the same number of responders (574) and non-responders (10,054). Then, we trained and tested each method 10 times, using each of the 10 subsets in turn as the testing set and all of the remaining datasets combined as the training set. All the methods use exactly the same ten testing and validation datasets, each of which is 10% of the entire data. The results for each method are the average of the 10-fold cross-validation.

Results

First, nine variables were selected by logistic regression using the forward selection criterion (p = .05): recency (Recency for R), i.e., the number of months lapsed since the last purchase, frequency of purchase in the last 36 months (Frequency for F), monetary value of purchases in the last 36 months (Money for M), average order size (Ordsize), lifetime orders (number of orders placed — Liford), lifetime contacts (number of mailings sent — Lifcont), and whether a customer typically places telephone orders (Tele), makes cash payment (Cash), and uses the “house” credit card (Hcrd) issued by the catalog company. Thus, the model includes the RFM variables, lifetime variables, and purchase and payment behaviors. In subsequent experiments, all models use exactly the same nine variables.

The results based on the 10-fold cross-validation in Table 2 indicate that among the unconstrained models, the baseline model using logistic regression achieves an average response lift of 376.4 and 262.4 in the first two deciles. Its average profit lifts are 589.9 in the first decile and 364.8 in the second decile. The OLS model produces average response lift of 352.8 in the first decile and 269.4 in the second decile, and its profit lifts in the top two deciles are 610.6 and 379.4. By comparison, the unconstrained GA model achieved response lifts of 151.3 and 135.1 and profit lifts of 248.9 and 191.5 in the top two deciles. Thus, using the partial order function alone (without the penalty method for optimization), the linear GA model’s performance is much below those of logistic regression and OLS regression. However, these are all parametric or semi-parametric models. Thus, we also produce the results of a popular nonparametric model using the boosting method (AdaC2), which produces response lifts in the top two deciles as 375.0 and 263.1 (Table 3). Its profit lifts are 593.6 and 367.7, only slightly better than those of the logit model. Thus, without incorporating constrained optimization, a non-parametric model alone does not render superior performance.

For the constrained optimization models, we adopt the 80 percentile or top 20% as the benchmark constraint. The DMAX model generates average response lifts of 335.2 in the first decile and 249.2 in the second decile. Its average profit lifts are 596.7 in the first decile and 364.0 in the second decile. Thus, the DMAX model produces higher response lifts than the unconstrained GA model without penalty but only slightly higher profit lift in the first decile than logistic regression (596.7 vs. 589.9).

By comparison, the POCO model records much higher response lifts (369.1 and 274.7) than the DMAX model. It also generates significantly higher profit lifts (629.1 and 385.1) than the DMAX model. (p ≤ .1 and p ≤ .05 respectively, based on t-tests of the 10-fold cross-validation results). In terms of computing efficiency, the proposed POCO model takes an average of 1 h 40 min for a 10-fold cross-validation experiment, compared to the DMAX model, which requires about 6 to 8 h of computing time (Mitchell 1997).

Table 2
Response and profit lifts of unconstrained models.

<table>
<thead>
<tr>
<th>Models</th>
<th>Decile</th>
<th>Response lift (no constraint)</th>
<th>Profit lift (no constraint)</th>
<th>OLS (no constraint)</th>
<th>Response lift</th>
<th>Profit lift</th>
<th>Linear GA model (no constraint)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Response lift</td>
<td>Profit lift</td>
<td>Response lift</td>
<td>Profit lift</td>
<td>Response lift</td>
<td>Profit lift</td>
</tr>
<tr>
<td>1</td>
<td>376.4</td>
<td>151.3</td>
<td>248.9</td>
<td>352.8</td>
<td>610.6</td>
<td>151.3</td>
<td>248.9 (32.63)</td>
</tr>
<tr>
<td>2</td>
<td>262.4</td>
<td>135.1</td>
<td>191.5</td>
<td>269.4</td>
<td>379.4</td>
<td>135.1</td>
<td>191.5 (20.27)</td>
</tr>
<tr>
<td>3</td>
<td>217.2</td>
<td>128.5</td>
<td>170.2</td>
<td>215.8</td>
<td>273.2</td>
<td>128.5</td>
<td>170.2 (11.52)</td>
</tr>
<tr>
<td>4</td>
<td>185.0</td>
<td>122.4</td>
<td>154.5</td>
<td>173.3</td>
<td>209.2</td>
<td>122.4</td>
<td>154.5 (13.22)</td>
</tr>
<tr>
<td>5</td>
<td>161.5</td>
<td>118.0</td>
<td>141.8</td>
<td>184.1</td>
<td>217.0</td>
<td>118.0</td>
<td>141.8 (10.81)</td>
</tr>
<tr>
<td>6</td>
<td>145.2</td>
<td>112.9</td>
<td>130.1</td>
<td>159.3</td>
<td>149.1</td>
<td>112.9</td>
<td>130.1 (7.80)</td>
</tr>
<tr>
<td>7</td>
<td>130.2</td>
<td>110.4</td>
<td>122.9</td>
<td>139.0</td>
<td>149.1</td>
<td>110.4</td>
<td>122.9 (5.79)</td>
</tr>
<tr>
<td>8</td>
<td>118.7</td>
<td>106.7</td>
<td>115.0</td>
<td>123.3</td>
<td>117.4</td>
<td>106.7</td>
<td>115.0 (3.04)</td>
</tr>
<tr>
<td>9</td>
<td>108.7</td>
<td>104.4</td>
<td>108.5</td>
<td>110.4</td>
<td>117.4</td>
<td>104.4</td>
<td>108.5 (1.01)</td>
</tr>
<tr>
<td>10</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>110.4</td>
<td>117.4</td>
<td>100.0</td>
<td>100.0 (0)</td>
</tr>
</tbody>
</table>

Note: The means of the ten-fold cross-validation followed by standard deviations in brackets.
in comparison with 21 h that took the DMAX model using the same i7 computer. These superior profit lifts and the computing time together lend support to the effectiveness and efficiency of the partial order function using the advanced genetic algorithm.

To test the robustness of the POCO model, we also perform the same experiments using constraint levels at the other top percentile (i.e., top 10%, 30%, and 40%) to compare the performance of the DMAX and POCO models. The results in Tables 4 and 5 show that the POCO model records higher response lifts in the top deciles than the DMAX model at all constraint levels. Moreover, the POCO model generates higher profit lifts in the top deciles than the DMAX model. The top decile profit lifts of DMAX models are between 591.7 (constrained at top 10%) to 581.6 (constrained at top 40%) with the other two constrained models having higher profit lifts (596.7 and 597.6). The profit lifts of the POCO model range from 630.8 to 600 in a smooth descending order, exhibiting greater stability. The superior performance of the POCO model is also reflected in the actual lifted profit, which ranges from 3.5% to 8.5% improvement over the DMAX model (Tables 4 and 5). In a campaign that contacts hundreds of thousands or even millions of potential consumers, this kind of improvement will generate tremendous incremental sales or profitability. For example, for a campaign that contacts one million customers (94.09 times as big as our testing dataset of 10,628), the POCO model would generate additional profit in the amount of $61,874 and $58,308 respectively over the DMAX model at the second decile for the constraint levels of top 10% and top 20%. Overall, the POCO model performs significantly better than the DMAX model in terms of both response and profit lifts at different constraint levels.

Discussion

Findings and Implications

Since the problem of budget constraint is prevalent in marketing operations, the partial order function (POCO) provides a viable solution to the constrained optimization problems and has several distinctive advantages with meaningful implications for marketing managers to improve the performance of targeted marketing. The results of our experiments indicate that the POCO model is a simple and effective approach to improving the ROI of marketing campaigns given the constraint of a fixed budget as it outperforms the conventional unconstrained model as well as the constrained DMAX model in terms of extra profit given the same budget constraint. It proves to be flexible, intuitively appealing,
computationally efficient, and straightforward for implementation at different constraint levels.

Given the popularity of direct marketing as a tool for promotion and customer relationship management, marketers need methods of intelligent decision support for customer selection and augmenting the profitability of targeted marketing with limited resources. The POCO model is particularly relevant for targeting the high value customers. This approach can help enhance operational efficiency and cost containment of marketing operations, thus affecting the bottom line of companies. It may also alleviate the problem of customer fatigue, a problem due to frequent promotions. This solution also applies to many other situations in marketing and customer relationship management when managers have a limited budget or manpower, for instance, 1) to provide incentives to those customers who are the most likely to upgrade their services and contribute greater profit, 2) to identify and prevent the most valuable customers from switching to a competitor, and 3) to predict the small minority who default on their large loans or credit card bills. Overall, the constrained optimization approach can be applied to a whole array of problems of targeted marketing and CRM operations to improve the ROI of marketing.

Limitations and Directions

Although the POCO method is effective and can improve the profitability of customer selection models, it is essentially a partial order model. Its insight into customer characteristics is not straightforward in comparison with that of parametric models. Researchers can explore sophisticated methods such as quantile regression to differentiate the high value customers from the low value ones. Applying the estimates of conditional quantiles, say from a percentage of high value customers, to predicting the responses of an entire population presents a significant challenge to marketer researchers. Such methods, however, can potentially help to select target customers and to raise flags for specific management actions such as customer retention, behavioral intervention, and upgrading or downgrading customer relationships. Researchers may also consider approaching this task as a problem of multi-objective or combinatorial optimization to improve model performance in both response and profit lifts and the cost-effectiveness of marketing operations.

This study applies the POCO method to only one dataset in direct marketing. Although these principles are relevant for other scenarios where managers have a limited budget or manpower, future studies may apply the POCO method to other datasets and problems in which maximum benefits are expected given the resource constraint. The linear scoring function as the underlying model to predict customer profit from a single promotion is relatively simple in comparison with the constrained optimization function. Future research may test the proposed method using more sophisticated revenue forecasting tools, such as the customer lifetime value (CLTV) models. Researchers may also consider the dynamic nature of customer values over time and incorporate them into these models (Ching et al. 2004; Göniil and Shi 1998). Given that resource constraints and optimal allocation of resources are prevalent in devising marketing strategies (Gupta and Steenburgh 2008), such studies can help to validate the benefits of the constrained optimization methods such as the POCO method and fine-tune their applications.

Appendix A. Additional Information of the Genetic Algorithm for the POCO Model

Population size = 256
Mutation rate = .1
Crossover rate = 1.0
Tournament size = 10
Generation Number = 2,000
The number of models evaluated (256 * 2,000) = 512,000

References


