Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Information Systems 29 (2004) 327–341

www.elsevier.com/locate/infosys

Exploiting response models—optimizing cross-sell and up-sell opportunities in banking

M.-D. Cohen*

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA

Abstract

The banking industry regularly mounts campaigns to improve customer value by offering new products to existing customers. In recent years this approach has gained significant momentum because of the increasing availability of customer data and the improved analysis capabilities in data mining. Typically, response models based on historical data are used to estimate the probability of a customer purchasing an additional product and the expected return from that additional purchase. Even with these computational improvements and accurate models of customer behavior, the problem of efficiently using marketing resources to maximize the return on marketing investment is a challenge. This problem is compounded because of the capability to launch multiple campaigns through several distribution channels over multiple time periods. The combination of alternatives creates a complicated array of possible actions. This paper presents a solution that answers the question of what products, if any, to offer to each customer in a way that maximizes the marketing return on investment. The solution is an improvement over the usual approach of picking the customers that have the largest expected value for a particular product because it is a global maximization from the viewpoint of the bank and allows for the effective implementation of business constraints across customers and business units. The approach accounts for limited resources, multiple sequential campaigns, and other business constraints. Furthermore, the solution provides insight into the cost of these constraints, in terms of decreased profits, and thus is an effective tool for both tactical campaign execution and strategic planning.

Keywords: Database marketing; Cross-selling; Up-selling; Profit optimization; Assignment problems; Constrained optimization; Response models

1. Introduction

The new mantra of database marketing in banking and financial services is "the right product to the right customer at the right time". However, a practical and effective implementation of this goal is not easy to accomplish. What makes this particularly difficult is that companies have multiple products and operate under a complex set of business constraints. Choosing which products to offer to which customers in order to maximize the marketing return on investment and meet the business constraints is enormously complex. This paper outlines a framework for solving this problem and presents an example.

Many banks have made deliberate efforts to become customer-focused institutions, as opposed to vertical product driven companies. Typically the goal is to provide products that help their

*Tel.: +1-919-531-7830.
E-mail address: marc.cohen@sas.com (M.-D. Cohen).

0306-4379/$ - see front matter
doi:10.1016/j.is.2003.08.001

customers improve their financial situations. A direct consequence of this goal is that marketing campaigns are multiple product campaigns as opposed to single product campaigns, in order to present customers with more appropriate offers and increase the ROI of the campaign. This transforms the data mining and campaign targeting process from a fairly simple application of individual response models into a significantly more complex problem of choosing which product, if any, to offer to which customer and through which channel. The benefit is that campaigns are more customer-focused than in the past.

1.1. Business problem

The database marketing community has changed significantly over the last several years. In the past, database marketers applied business rules to target customers directly. Examples include targeting customers solely on their product gaps or on marketers' business intuition. Marketers have also applied RFM-type analysis where general *recency*, *frequency*, and *monetary* measurements as well as product gaps are used to target customers for specific offers. The current approach, which has widespread use, relies on predictive response 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 offering. However, simply knowing a customer's probability of responding to a particular offer is not enough when a company has several products to promote and other business constraints to consider in its marketing planning.

Marketing departments also face the problem of knowing which *product* to offer to a *customer*, not just which customer to offer a product. In practice, many ad hoc rules are used. Prioritization rules based on response rates or estimated expected profitability measures have been used; business rules to prioritize products that can be marketed are sometimes used; and product response models to select customers for a particular campaign are also used. One approach that is easily implemented but, for reasons outlined later, may not produce optimal customer contact plans relies on a measure of expected offer profitability (the estimated probability of response multiplied by the profit given customer response less direct costs) to choose which products to offer customers. However, a shortcoming of this approach is its inability to effectively handle complex constraints on the customer contact plan.

1.2. Business constraints

Database marketing departments face several types of business constraints. Typically, there are:

- restrictions on the minimum and maximum number of product offers that can be made in a campaign,
- requirements on minimum expected profit from product offers,
- limits on channel capacity,
- limits on funding available for the campaign,
- customer specific 'do not solicit' and credit risk limiting rules, and
- campaign return-on-investment hurdle rates that must be met.

These are a sample of the constraints that marketing departments must meet when executing a campaign. Ad hoc approaches are also typically used in an attempt to meet these constraints.

The opportunity costs of the business constraints are generally not known. Constraints are usually negotiated between marketing, product lines and delivery channel management. If the cost of a constraint was known, then the company could choose to tighten or to relax the constraint by removing or adding more resources. For example, channel capacity could be increased if it were known that there was a significant return on the investment by doing so. Knowledge of the opportunity costs could help evaluate these management decisions. How the proposed framework contributes to these discussions will be addressed in the "Strategic Uses" section of this article.

Ultimately, the database marketer needs a concrete framework to effectively act on "the right product to the right customer at the right time" mantra. The approach we take is to transform the database marketing problem into an optimization problem that is designed to generate the maximum incremental profit from a limited amount of resources subject to the necessary business constraints. This paper will describe an actionable framework that will address this business problem.

2. Problem description and solution framework

It is helpful in understanding the solution framework to understand the current standard approaches to solving this problem and the data that are available for marketing campaign planning. We assume that there has been a thorough analysis of historical marketing campaigns and that accurate response probability models exist for all products. The result of these data mining exercises is a data set that contains an expected profit for each product for each customer, where the expected profit is derived from the customer specific probability of response and the profit generated given a customer response. The approach we take does not rely on the methods used to obtain the response and profitability measures. Needless to say, the data sets that result from analysis of customer data can be rather large. It is not unusual to have over 5 million customer records in such a data set.

2.1. Problem description

A formal description of the decision problem will help to clarify current practices and the proposed optimization approach. Let i denote the subscript on customer, let j denote the subscript on product, and let k denote the subscript on channel. The response models result in a probability p_{ijk} that customer i accepts offer j via channel k, assuming of course that the offer is made. Consider the random variable $a_{ijk} = \{1$ with probability p_{ijk}, 0 otherwise$\}$. When $a_{ijk} = 1$ then customer i accepts offer j via channel k. The profitability models are used to estimate an expected value of v_{ij}, a random variable that takes the value of the financial value of the customer to the institution given that the customer accepts the offer. Define $r_{ijk} = E(a_{ijk}v_{ij})$ and let $\hat{r}_{ijk}$ denote an estimate of r_{ijk}.

The problem of deciding which particular offer to make to each customer can be expressed as characterizing a decision function or decision variable as $d_{ijk} = \{1$ if offer j is made to customer i on channel k and 0 otherwise$\}$, such that $\sum_{jk} d_{ijk} \leqslant 1$ for each i. This inequality restricts the number of product offers to each customer to be at most 1.

The database marketing community has been refining the science around estimating r_{ijk} for a number of years. The KDD community has been one source of energy for these advancements but there have been many others as well. One could argue that *data mining* has been significantly economically motivated by the need to obtain response and profitability estimates for the database marketing community. However, the discussion on how to exploit these estimates in decision-making seems to be in its infancy within this community.

2.2. The current solution

There are many possible characterizations for d_{ijk}. The database marketing community, particularly in banking, uses an approach that could be called a greedy approach because it uses the "best available choice" first. Suppose that we know the number of offers for each product and an order for the offer types. Then starting from the first offer type, the Greedy algorithm assigns the offer to the customer who has the highest expected return, then the second highest, etc. until the number of offers for that type have been exhausted. Then repeat this process for the remaining offer types. More rigorously, consider the following characterization of the greedy algorithm for specifying d_{ijk}.

Algorithm G. Given an ordering to the offers $j_1, j_2, j_3, \ldots$, the number of customers desired for each offer $N_{j_1}, N_{j_2}, N_{j_3}, \ldots$, and Ω the total number of customers, let $n = 1$ and $d_{ijk} = 0$ for all i,j, and k.

1. Order the customers so that $\hat{r}_{i_1 j_n k_1} \geqslant \hat{r}_{i_2 j_n k_2} \geqslant \hat{r}_{i_3 j_n k_3} \geqslant \cdots$ over all k and i.

2. Set $l \leftarrow 0$ and $m \leftarrow 1$.
3. If $\sum_{jk} d_{i_m jk} = 0$ then set $d_{i_m j_n k_m} \leftarrow 1$ and go to 5.
4. Set $m \leftarrow m+1$ if $m \leqslant \Omega$ then go to 3 else not enough customers to satisfy N_{j_n} requirement.
5. Set $l \leftarrow l+1$ if $l \leqslant N_{j_n}$ then go to 3.
6. Set $n \leftarrow n+1$. Go to 1 unless all offers are exhausted.

Let the decision function d_{ijk} determined with Algorithm G be denoted by d^G and let the expected value of the campaign following this decision function be $T(d^G) \equiv \sum_{ijk} \hat{r}_{ijk} d^G_{ijk}$. Note that if Algorithm G is applied for only for one product j_m then the greedy solution is optimal in the sense that any other solution for d_{ijk} would result in a smaller value for $T(d_{ijk}) \leqslant T(d^G)$.

While easy to understand and implement the Greedy algorithm has a number of drawbacks, the main one being that it results in sub-optimal assignments of offers but it also ignores the business constraints and requires that the number of product offers be known precisely beforehand. Even in the case where there is only one product the result may not account appropriately for constraints.

2.3. The ideal solution

The ideal approach to solving for decision variables is to find d_{ijk} that maximizes the objective $T(d) = \sum_{ijk} \hat{r}_{ijk} d_{ijk}$ subject to the business constraints as outlined in the previous section. If we assume that the constraints are linear then this type of problem can be expressed as an integer program I. An example follows:

$$
\begin{aligned}
\max \quad & \textstyle\sum_{ijk} \hat{r}_{ijk} d_{ijk} \\
\text{s.t.} \quad & \textstyle\sum_{ijk} d_{ijk} c_{ijk} \leqslant B \ \text{ campaign budget}, \\
& \textstyle\sum_{ij} d_{ijk} \leqslant C_k \ \text{ channel } k \text{ capacity } \forall k, \\
& \textstyle\sum_{ik} d_{ijk} \geqslant N_j \ \text{ minimum offers of product } \forall j, \\
& \textstyle\sum_{ik} d_{ijk} \leqslant M_j \ \text{ maximum offers of product } \forall j, \\
& \textstyle\sum_{ijk} d_{ijk} \hat{r}_{ijk} \geqslant \sum_{ijk} d_{ijk} c_{ijk} (1+R) \ \text{ corporate hurdle}, \\
& \textstyle\sum_{jk} d_{ijk} \leqslant 1 \ \text{ at most one offer per customer } \forall i, \\
& d_{ijk} \in \{0,1\},
\end{aligned}
\tag{I}
$$

where c_{ijk} is the cost of offering customer i product j over channel k; C_k is the capacity of channel k; and R is a corporate hurdle rate. This formulation contains constraints on the decision variables that reflect the types of business constraints discussed above. The ability to express and capture these business requirements in this way gives approach I a distinct advantage over Algorithm G. Denote the optimal solution to problem I as d^I. Because formulation I accounts for constraints it is possible that $T(d^I) < T(d^G)$ even though approach I is an optimal approach compared to Algorithm G. However, this has not been our experience in practice as will be discussed later.

The particular formulation above captures only the bare elements of the problem. It does not account for multiple campaigns composed of different products as an example. However, the model can easily be extended to cover most typical business constraints encountered in practice, but the basic formulation remains the same. This ideal formulation is difficult to solve because of its scale. For 1 million customers and 10 products there are 10 million integer decision variables d_{ijk}, this yields $2^{10{,}000{,}000}$ possible customer–offer combinations. Using standard optimization methods to solve a problem of this size can, in principle, result in a branch and cut tree of as many nodes. Because of this problems of this size are extremely difficult to solve, so we propose an alternative approach. While not providing a strictly optimal solution, the alternative approach does provide an approximately optimal solution that in preliminary studies has shown to be a good approximation.

2.4. The practical solution

Although it is not practical to solve problems formulated in this ideal way, it is possible to approximate the ideal formulation and arrive at a formulation that is practical to solve. There are numerous ways to approach this approximation; one approach is to sample from the customer base and use that sample as representative for the optimization. However, the sampling approach will tend to under-represent customers in the tails of the distribution of $\hat{r}_{ijk}$. Another approach (and the one that we take) is to aggregate customers based on the coefficients $\hat{r}_{ijk}$ and c_{ijk} in the ideal formulation. Aggregation can be considered natural in this setting particularly when we understand that much of the data is consistent across sets of customers as a result of the estimation, response modeling, and statistical modeling processes that are commonly employed. The implementation of this framework is loosely coupled to the chosen form of the predictive response models. As long as the customer/offer specific response rate is represented as a probability and therefore the expected profit of the offer can be calculated, the proposed framework can handle it. Our banking and financial services customers use standard, accepted statistical and data mining approaches to obtain these estimates.

The aggregation process we use involves conversion of the raw data into a form that can be used *naturally* in a linear programming optimization model. The key is to group the raw data $\hat{r}_{ijk}$ and c_{ijk} and use the aggregate as a single data point. The purpose here is not the identification of customer segments or to differentiate groups of customers, but to aggregate customers into similar groups. This is an important point to keep in mind since many statistical aggregation and clustering methods are most frequently used to distinguish customer, not to group them. If the aggregates are internally consistent, then the aggregate centroids can be used as representative of the data for all the customers within a single aggregate.

This aggregation process enables the problem to be reformulated as a linear program so that rather than assigning offers to individual customers, as the ideal integer program does, the program identifies proportions within each aggregate for each product offer. This can be accomplished with similar constraints to those of the ideal formulation. Moreover, the linear program is much smaller and much easier to solve. Note however, the solution may require that multiple products are offered to proportions of customers within a single aggregate. When that happens, a new problem is defined that is a simple assignment problem at the level of the aggregate, where multiple offers are to be assigned within the aggregate, and it is relatively easy to solve.

2.4.1. Formulation

Consider the following variables defining raw data as input into the solution algorithm. Let y_{ljk} be the number of customers in aggregate l that are offered product j over channel k; let A_l be the set of customers in aggregate l; let $\hat{r}'_{ljk}$ be the estimated expected profit given that customer in aggregate l is offered product j over channel k; let c'_{ljk} be the cost of offering a customer in aggregate l product j and channel k; let R be the corporate hurdle rate. Then, a very simplified version of the problem can be expressed as finding the y_{ljk} that satisfy

$$
\begin{aligned}
\max \quad & \textstyle\sum_{ljk} y_{ljk}\hat{r}'_{ljk} \\
\text{s t.} \quad & \textstyle\sum_{jk} y_{ljk} \leqslant |A_l| \ \text{ number customers in aggregate } \forall l, \\
& \textstyle\sum_{ljk} y_{ljk}c'_{ljk} \leqslant B \ \text{ campaign budget}, \\
& \textstyle\sum_{lj} y_{ljk} \leqslant C_k \ \text{ channel } k \text{ capacity } \forall k, \\
& \textstyle\sum_{lk} y_{ljk} \geqslant N_j \ \text{ minimum offers of product } \forall j, \\
& \textstyle\sum_{lk} y_{ljk} \leqslant M_j \ \text{ maximum offers of product } \forall j, \\
& \textstyle\sum_{ljk} y_{ljk}\hat{r}'_{ljk} \geqslant \sum_{ljk} y_{ljk}c'_{ljk}(1+R) \ \text{ corporate hurdle rate}, \\
& y_{ljk} \geqslant 0.
\end{aligned}
\tag{A}
$$

Let y^*_{ljk} be the optimal values that satisfy the formulation above. They are the optimal number of customers within aggregate that are offered a specific product. For example, suppose that y^*_{ljk} is the total number of customers in aggregate l, namely $y^*_{ljk} = |A_l|$. Then, every customer in aggregate l should be offered product j via channel k. Alternatively, suppose that for a given l, $y^*_{ljk} > 0$ and $y^*_{lj'k} > 0$ for $j \neq j'$. Then, for aggregate l, y^*_{ljk} customers must be offered product j and $y^*_{lj'k}$ customers must be offered product j'. The optimal way to distribute these offers over the aggregate is to solve a simple assignment problem using the estimated expected profit $\hat{r}_{ijk}$ for the individual customers and not the aggregate centroids as is used in formulation A. The formulation below shows that assignment problem with the objective function limited over A_l.

$$\begin{aligned} \max \quad & \textstyle\sum_{i \in A_l, jk} \hat{r}_{ijk} d_{ijk} \\ \text{s.t.} \quad & \textstyle\sum_{i \in A_l} d_{ijk} = y^*_{ljk} \quad \forall jk, \\ & d_{ijk} \in \{0, 1\}. \end{aligned} \tag{1}$$

The total expected return from the campaign becomes $T(d^P)$ where $d^P \equiv \bigcup_l d^l$. The relationship between $T(d^P)$ and $T(y^*_{ljk})$ is not easy to characterize since it depends on the distribution of $\hat{r}_{ijk}$ within each of the aggregates. If all the aggregates are each perfectly homogeneous with respect to the $\hat{r}_{ijk}$ then $T(d^P) = T(y^*_{ljk})$. It is important to note that some of the constraints may be violated as a result of solving this assignment problem particularly if the aggregate centroids used in the linear program formulation are involved in a tight constraint and not consistent within the aggregate. This problem with constraint violation can be remedied to some extent but at some additional potential loss in the expected value by adding the constraints to the assignment problem and then adjusting the right-hand sides of each of the assignment problems to account for the left-hand sides realized by all previous assignment solutions. The details of that are not presented here.

Finally, an algorithm for finding the solution d^P is given below.

Algorithm P.

1. Aggregate the customers into M aggregates using the $\hat{r}_{ijk}$ and c_{ijk}, such that the variation in the values within aggregates is minimized.
2. Calculate the centroids $\hat{r}'_{ljk}$ and c'_{ijk} for each of the $l = 1, \ldots, M$ aggregates.
3. Solve A for y^*_{ljk}.
4. For $l = 1, \ldots, M$ solve $T(d^l)$.

As mentioned above it is possible that $T(d^P) < T(d^G)$, however this has not been our experience in practice. In fact, our experience is that $T(d^P) \gg T(d^G)$. It is also been our experience that for the constraints used in practice the approximation does not result in serious violations and is a significant improvement over standard practices.

3. Example

We demonstrate this approach with data from a marketing department in a large international commercial bank and using existing procedures within the SAS system to implement the Algorithm P described above. The details of the SAS code will not be given.

Eleven unique offers were to be considered: five investment, three lending and three day-to-day banking offers. The investment offers included GICs, mutual funds, education savings program (RESP), and two unique discount brokerage offers. The lending offers included a mortgage and two credit card offers. The day-to-day banking offers included one of two online banking service offers and a deposit account acquisition. The term campaign is used here to imply one large pro-active customer contact campaign that it comprised of 11 distinct offers; it can be thought of as 11 single product campaigns that are being offered at generally the same time to a non-overlapping set of customers. For the purposes of this paper the detailed product offer descriptions will not be given. Approximately 2.5 million customers were included in the potential universe for the campaign.

Ultimately, the goal of marketing campaigns is to produce a positive return on investment for the company that exceeds the corporate investment hurdle rate. Although the timeframe upon which this investment should be measured may be debatable, the goal is fundamental to the bank. To achieve this specific objective, the bank can execute marketing campaigns that are designed to maximize the expected incremental profit through making one of several offers to some of its customers, or potential customers.

3.1. Response models

The expected incremental profit of a specific offer to a customer is an estimate based on response models and detailed product profitability calculations. These response models are used to estimate *the probability that a customer will accept a specific offer*. A data warehouse has detailed account level profitability calculations for all of the bank's products. This profitability information is used to estimate the near-term incremental profit *given that the customer accepts the specific offer*. Once a specific offer is made to a customer there are two possible outcomes: the customer can accept or reject the offer. Using the offer specific response models the probability of both states is known for each customer. The incremental profit for both states is also known; it is zero if the customer rejects the offer and the mean near-term profitability for new accounts of the specific type if the offer is accepted. With this information, the expected incremental profit of the offer can be calculated for each customer/offer combination. This is the estimate $\hat{r}_{ij}$. The cost of making each offer is also known and is largely dependent on the channel through which the offer is made.

3.2. Channels

The bank has several distribution channels through which campaigns can be executed. The main channels for direct marketing are direct mail, retail branch centers and call centers. For this example we assume that leads sent to the branch officers and call centers are follow-ups from a direct mail piece and that offers designated as direct mail are direct mail only. The use of the branch and call centers for follow-ups has been shown to have a positive effect on the probability of response to the offer when compared to direct mail alone. Of course, the lead delivery costs vary with the channel used. In this example we have used costs per lead of \$3.00, \$1.50 and \$1.00 for the branch, call center and direct mail only channels, respectively.

3.3. Business constraints

Several practical issues surround the campaign execution process that affects the customer/offer selection process, for this application to be acceptable for implementation these business constraints must be maintained. The following business rules have been translated into constraints that can be applied to the optimization model:

- Campaign costs cannot exceed \$1 million.
- The campaign must have a return on investment of at least the corporate hurdle rate. In this example we have used 20%, which is not necessarily the bank's actual corporate hurdle rate.
- The branch and call center channels have a certain capacity constraint for timely processing of campaign generated leads. In the example, the call center can accept up to 500,000 leads, the branch can handle up to 250,000 leads and direct mail is unlimited.
- Product offer minimums are also required to satisfy internal bank objectives. For the purposes of this example we set all offer minimums to 20,000 with two exceptions. The RESP offer, which has an extremely limited eligible universe, had a lower bound of 2500 and one of the online offers had a lower bound of 5000.
- Cannot offer products to customers who already have that product.
- The standard marketing exclusions, such as credit risk or do not solicit, must also be strictly adhered to.

3.4. Solution

The estimates for customer/offer expected incremental profit, costs and business constraints serve as inputs to the profit optimization phase of the campaign design. The profit optimization phase is independent of the construction of these inputs. This means that as response models, profit estimates or costs are refined as long as the results are represented in the same manner, the optimization phase will be able to accept them as inputs. This property is important as the bank is constantly testing and refining these inputs as the marketplace is ever changing.

3.4.1. Results

The result of this algorithm is an allocation, of a specific offer, or no offer, to each customer. Also output is the associated expected incremental profit by customer making that offer. This solution is a SAS data set that has a customer identifier, the expected return, offer and channel designation. The full data set is 2.5 million records; Fig. 1 shows the first 25 records.

To better understand the solution, it is useful to look at several charts that summarize the solution and a report that is produced by the algorithm.

Customer Id	Product	Channel	Return
00182723			.
00200688			.
00032937			.
00722119			.
02137391			.
00992639			.
00060721			.
00483601	Offer 2	Direct Mail	0.0005
01164964			.
00025469			.
01008244			.
00179891			.
00410488	Offer 10	Direct Mail	3.1852
01484008			.
00184804			.
03350118			.
00983111			.
00387834	Offer 5	Call Center	13.0782
01100914			.
01507075			.
01559899			.
00309931			.
00657640			.
01095694			.
02075404			.

Fig. 1. Sample of the solution data set.

3.4.2. Offers

The Offer Frequencies by Channel chart, Fig. 2, provides a graphical representation of the distribution of offers by channel and is useful for understanding the solution. This figure shows that few of the contacts have a branch follow-up treatment; some have call center follow-ups and most have just the direct mail treatment. Offer 10 has the largest quantity of contacts and is spread across the direct mail and call center channels. The Constraint Report, Fig. 3, provides significantly more insight into the nature of the derived solution.

3.4.3. Business constraints

The Constraint Report summarizes the constraints applied to the problem as well as outlines the chosen level and marginal costs associated with each constraint.

- The last line of the Constraint Report in Fig. 3 shows that the objective function, expected profit, was maximized at \$3.58 million from an expenditure of \$1 million. This results in a 258% return on investment for the campaign.
- The first two lines in Fig. 3, *Branch Capacity* and *Call Center Capacity*, summarize the results of the branch and call center capacity constraints, respectively. The branch capacity for follow-up contacts was limited to 250,000 and the call center to 500,000. In the solution, only 2180 contacts were assigned to the branch, and 202,258 to the call center, for follow-up calls. This low quantity of follow-up contacts, at either the branch or the call center, is due to the conservative estimate of the increased response rates resulting from the follow-up and the significantly higher cost, \$3.00 for branch and \$1.50 for call center, as compared to direct mail only, \$1.00.
- The third constraint, *Campaign Cost*, limits the total costs for the campaign to \$1 million. This constraint is in fact tight, meaning that the optimal solution was restricted by the condition. The marginal value of the constraint is \$1.53, this means that an additional \$1 spent on the campaign would result in a \$1.53 increase in expected profit.

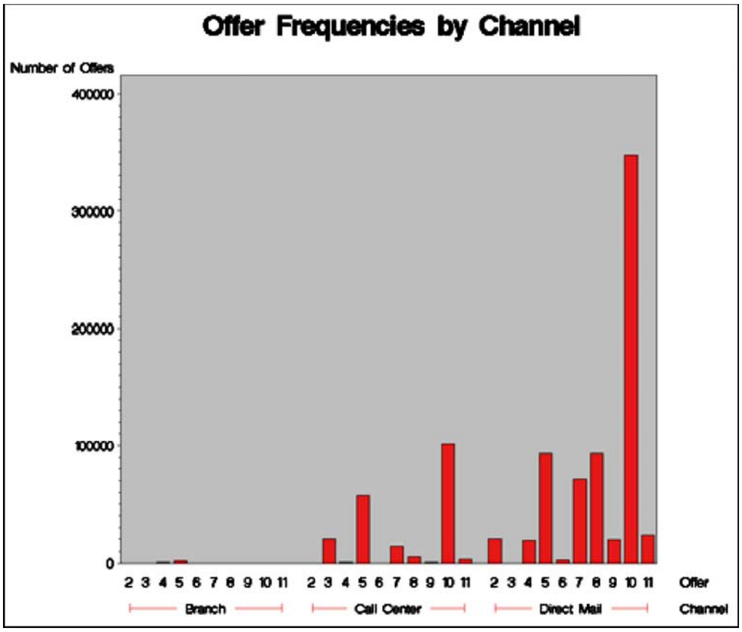

Fig. 2. Offer Frequencies.

Constraint Name	Lower Target	Level	Upper Target	Marginal Value
Branch Capacity	.	2180.00	250000	0.00000
Call Center Capacity	.	202258.00	500000	0.00000
Campaign Cost	.	1000000.00	1000000	1.52605
Offer 10	20000	448325.00	.	0.00000
Offer 11	5000	7046.00	.	0.00000
Offer 12	20000	20000.00	.	-1.12965
Offer 2	20000	20000.00	.	-1.50959
Offer 3	20000	20000.00	.	-0.58726
Offer 4	20000	20000.00	.	0.00000
Offer 5	20000	152468.00	.	0.00000
Offer 6	2500	2500.00	.	-1.98643
Offer 7	20000	85665.00	.	0.00000
Offer 8	20000	98507.00	.	0.00000
Offer 9	20000	20000.00	.	-1.47605
Objective	.	3583503.25	.	0.00000

Fig. 3. Constraint Report.

- The *offer* constraints show the lower bounds for each of the specific offers. Notice that five of the 11 products were limited by their lower bounds. In the solution, *Offer 9* was only made to 20,000 people. If that constraint were to be *decreased* by one unit, e.g. to 19,999, then the objective function, expected profit, would be *increased* by \$1.48—so the cost, on expected profit, of this business constraint is clear.

3.4.4. Profitability

The estimated Profitability by Channel report, Fig. 4, clearly reveals the quality of leads that are sent to the respective channels. The branch follow-up leads have a significantly higher expected incremental profit than for call center follow-up leads or direct mail alone. The call center is also sent leads that are more profitable than direct mail alone. A few comments about the way that the channel effects were modeled are necessary to fully understand this result.

The differential effects of the various channels enter the response models as a *main effect*. A call center follow-up treatment increases the probability of response as compared to no call center follow-up treatment. The branch follow-up treatment has the same directional effect as the call center, although larger. As such, everything else the same, the expected profit from making an identical offer to a customer with a branch or call center follow-up is greater than without the follow-up. As such we would expect to see a higher expected rate of return when applying the additional follow-up treatments, although not at this magnitude. There is some other factor driving the higher rate of return, in fact, it is the channel selection by the optimization routine.

Recall that the channel costs are fixed at \$1.00, \$1.50 and \$3.00 for direct mail only, call center follow-up and branch follow-up, respectively. The incremental expected profit enters the equation through an increase in response rate. For a branch follow-up to be more profitable than an offer made without a branch follow-up it must be the case that:

$$
\begin{aligned}
&Prob_{Branch} \cdot Profit - Cost_{Branch} \\
&\quad > Prob_{DM} \cdot Profit - Cost_{DM} \\
&\quad \Rightarrow Profit \cdot (Prob_{Branch} - Prob_{DM}) \\
&\quad > Cost_{Branch} - Cost_{DM} \\
&\quad \Rightarrow Profit \cdot (Prob_{Branch} - Prob_{DM}) > 3 - 1 \\
&\quad \Rightarrow Profit \cdot (Prob_{Branch} - Prob_{DM}) > 2.
\end{aligned}
$$

Either the profit given that the customer accepts (*Profit*) or the incremental effect on the response probability of the Branch follow-up ($Prob_{Branch}$–

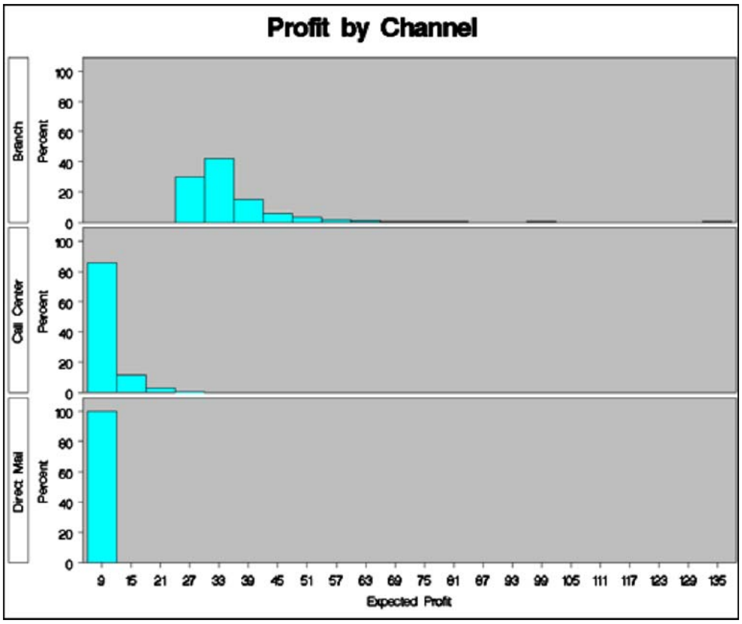

Fig. 4. Profitability Report.

$Prob_{DM}$) is large enough to overcome the \$2 increase in contact costs ($Cost_{Branch}$–$Cost_{DM}$). The greater the difference in this inequality the more beneficial the follow-up contact is. The difference can be large if a highly profitable product is being offered, or the increase in the probability of response is high. For reasonable values of the probability of response from the direct mail ($Prob_{DM}$), the increase in response rate due to a branch follow-up ($Prob_{Branch}$–$Prob_{DM}$) will rise with an increase in the probability of response from the direct mail ($Prob_{DM}$). From a business perspective, this means that either highly profitable offers and/or customers who are most likely to respond to the offer will tend to be given a follow-up treatment. This is a particularly important result when dealing with the business owners of the call center and branch channels.

3.5. *Summary of tactical uses*

In summary, the solution provided an approximately optimal solution to the ideal capacitated assignment problem. The output is a decision, for each customer as to which, if any, product to offer and through what channel. The campaign expected profit is \$3.58 million on \$1 million invested, for a return on investment of 258%. Using an ad hoc approach, that utilized response models and near-term profit, and met all the business constraints the most profit that could be generated was \$2.65 million on \$1 million invested for a campaign return on investment of 165%. The boost in the campaign return on investment of 93% is entirely attributable to the quality of the solution produced by the optimization process as opposed to the ad hoc approach.

It should also be noted that preliminary experiments solving the approximation with varying numbers aggregate indicate that as the number of aggregates increases the value of the objective quickly rises then slowly converges to the integer relaxed solution. Further study is needed to identify a good number of aggregates that work in a "typical" setting.

3.6. *Strategic uses*

Although this technique was developed primarily for its tactical application, as described above, it has some significant strategic applications too. The strategic applications are in the area of capacity planning. Two insights will be discussed, one dealing with campaign budgeting and the other with channel capacity planning.

3.6.1. *Campaign budget allocation*

In general, campaign budgets are determined prior to the campaign design. The degree of analysis that goes into determining specific campaign budgets, or annual campaign budgets, can vary greatly from institution to institution. As a strategic tool, the optimization technique provides an opportunity to determine the effects of making different budget allocations—in the budgeting process.

For example, in determining how much money to invest in a campaign, it would be useful to know the marginal return on an additional dollar investment. This is reported at \$1.53 in Fig. 3 as the marginal value of the \$1 million cost constraint. Given the campaign definition, all of the other constraints and the current customer base, investing one more dollar would result in an increase in profit of \$1.53. If marginal return on investment is greater than the corporate hurdle rate then that supports an argument to increase the investment.

This technique provides a compelling and empirically based process for altering, positively or negatively, the campaign budget. Of course, other considerations go into budget allocations but the technique could shed light onto the impact of such decisions.

3.6.2. *Channel capacity planning*

Similarly, channel capacity planning can benefit from this technique. It is understood that in the short run, channel capacity is fixed, although in the long run, or planning mode, channel capacity can be changed.

Again, from the Constraint Report, Fig. 3, if it appears as though a specific channel is used to capacity, then we can look at the marginal value of the constraint. The marginal value of these constraints gives the increase in profit, the objective function, given a one-unit increase in channel capacity. With the cost of this increase in capacity quantified, one can determine if the additional investment in the channel is warranted. This also helps to quantify the opportunity costs of having branch staff shift away from non-campaign related work. Again, this is from the perspective of campaign execution, there are other benefits to channel capacity augmentation that would also have to be taken into account, but at least the campaign benefits would be understood.

3.6.3. Sensitivity analysis

In addition to using the solution for capacity planning the approach lends itself to analyzing the sensitivity of the solution to changing assumptions. This is particularly useful for evaluating the solution to changes in the business constraints.

For example, a scenario can be optimized with several alternative budget constraints. Fig. 5 shows the results from solving a model that is very similar to that discussed above. Each row in the figure corresponds to a different limit on the total cost of the campaign. All the other constraints and data in the model are held constant. The first scenario allocates at most $1,000,000 on the campaign, the second $1,250,000 and the third $1,500,000. For each of these scenarios the optimal solution is calculated and the expected return from that solution, the number of offers through the branch, and the ROI, are also shown in the figure.

It is interesting to note that for each of these scenarios, as for the example above, the branch capacity is 250,000 offers. However, until the size of the campaign, as measured by the cost of the campaign, reaches $1,500,000, the branch channel is not used to its capacity. This means that for campaigns with a budget less than $1,500,000 there is excess capacity in this channel and its value is lost. The additional $500,000 expenditure results in an additional $670,000 return with a 34% ROI and no excess branch capacity.

Cost	Return	Branch Activity	ROI
1,000,000	3,756,135	94,088	275%
1,250,000	4,132,821	148,518	230%
1,500,000	4,426,410	250,000	195%

Fig. 5. Effects of increased campaign size.

4. Conclusions

In conclusion we discuss some shortcomings of the approach, identify some open questions, and then summarize the advantages of the approach. The shortcomings of this approach can be broadly classified into two categories: business and technical. From a business perspective there are three perceived shortcomings of this solution. These shortcomings are related to the required inputs to the solution, changes in the types of campaign design decisions, and post-analysis, which are performed by the business user. From a technical perspective there are two general shortcomings of this approach. The first is related to the constraints, and the second is related to the acceptable problem size.

4.1. Additional business issues

The approach requires as an input the expected incremental profit associated with each offer and customer combination. Fundamentally, this implies the creation, and maintenance, of offer specific probability of response models and detailed profitability measurements. Both of these inputs require ongoing maintenance and enhancement, as the marketplace is ever changing. This requirement does not seem to be too onerous as most organizations that would consider implementing this solution are likely already are producing these inputs.

The types of decisions made by business users will be altered with the adoption of this approach. The business users will be making decisions that affect the objective function and constraints but not *directly* about how many offers to make of each type or through which distribution channel.

At first the business users might resist this solution, as the decisions that are being made by the business are more abstract than those made during the business as usual campaign design process. Although the business users are making more abstract decisions about campaign design, they are actually gaining more *effective control* over the campaign. Thus, the successful adoption of this process would likely involve some amount of business user education with respect to the design of the optimization problem.

The solution was designed to explicitly maximize the expected incremental profitability from running a campaign, not to maximize the overall response rate of the campaign. For instance, a customer might have a higher estimated response probability to a low rate credit card offer than a high rate card. However, the profitability given offer acceptance could be substantially higher for the high rate card. If the goal were to maximize the response rate, then the low rate card would be offered. Whereas, if the goal were to maximize profitability then the offer with the greatest expected profitability would be made. This is only a shortcoming in the sense that business users' expectations have to be managed as campaign response rates are more easily and quickly measured than is campaign profitability.

4.2. Technical issues

As this solution was developed in a linear programming framework the constraints must be expressed as linear functions of the choice variables. A couple of examples of plausible business constraints that are non-linear and therefore do not work with the linear programming approach are:

- The number of credit card offers must be greater than 20,000 or equal to zero.
- The cost of a specific offer is a function of the number of those specific offers being made.

The scalability of this approach has not been exhaustively explored. Tests have been run on data sets with 2.5 million customers, 11 offers and three distribution channels and the solution is generated in an acceptable amount of time and resources. The solution was explicitly designed to scale well to the number of customers. Scalability in the offer and channel dimensions is significantly more expensive than along the customer dimension. Although, the number of distribution channels that a company can utilize in an automated campaign is not too large; the number of possible products that could be offered could grow well beyond 11.

4.3. Open questions

There are numerous questions that could be posed regarding the relationship among the solutions $T(d^G), T(d^1), T(y^*_{ljk})$ and $T(d^P)$. In this section we will comment on two issues which have been of concern among those who are practicing this approach.

4.3.1. Aggregation effects

The first issue regards the relationship between $T(d^I)$ and $T(y^*_{ljk})$. Because $T(y^*_{ljk})$ is a function of the number of aggregates M, it is not too hard to show that $\lim_{M\to\infty} T(y^*_{ljk}) \geqslant T(d^I)$. This follows because when M equals the number of the customers, A is the integer relaxed version of I. In order to automate the identification of an acceptable number of aggregates we have examined this relationship in a number of settings. Fig. 6 shows data for five solutions of $T(y^*_{ljk})$ for varying numbers of aggregates on data and a model very similar to that shown above. Notice that as the number of aggregates increase the value of the optimal solution appears to be converging. This experiment provides a gauge for the approximate number of aggregates required for a satisfactory objective value. The default number of aggregates is specified by the automated process. It should also be noted that as the number of aggregates increase so does the time required to obtain a solution. Finally, there is also some speculation that the likelihood and severity of constraint violation by the resulting solution $T(d^P)$ also increases as the number of aggregates increase.

A simple experiment gives some further validation to the quality of $T(d^P)$ and its proximity to

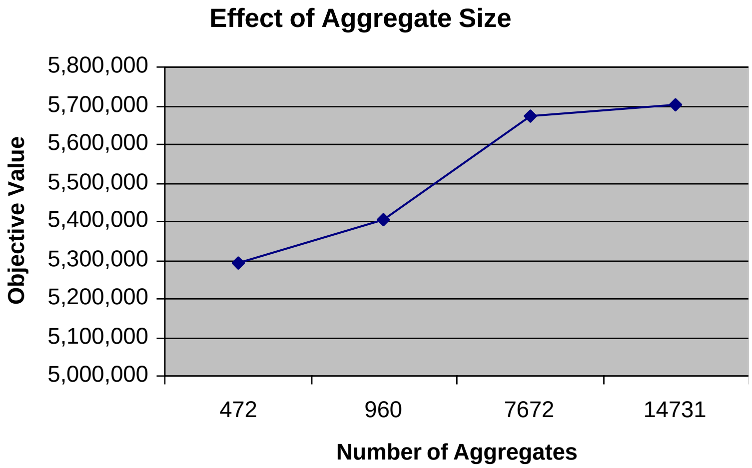

Fig. 6. Effects of aggregation size with example data.

Number of Aggregates	$T(y^*_{ljk})$	$T(d^P)$
1	58,188	2,285
2	75,359	45,667
4	90,840	77,176
8	106,190	104,724
16	113,229	110,586
32	114,243	114,059
64	115,420	115,175

Fig. 7. Effects of aggregation size in sampled data.

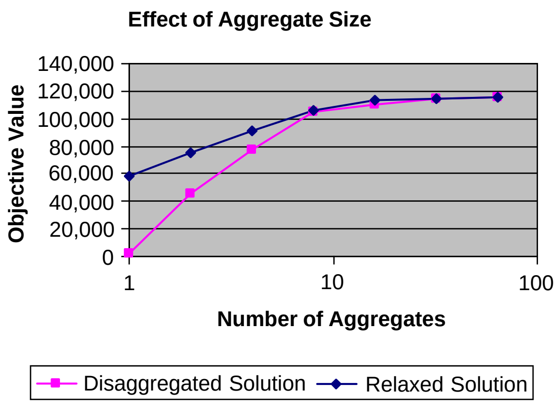

Fig. 8. Graph of aggregation size in sampled data.

$T(y^*_{ljk})$ and thus to $T(d^I)$. We sampled data for 1000 customers in a Monte-Carlo experiment that had two products and two channels. We applied Algorithm P for varying aggregation sizes and show the results in the chart in Fig. 7. Notice that for small aggregates the solutions are widely apart but they converge rapidly. One would also expect that the number of aggregates needed to achieve a fixed level of performance would increase as the number of products and channels increase.

The graph in Fig. 8 shows the same data with the logarithm of aggregate number on the horizontal axis. This shows the rapid convergence for this example.

Although not conclusive this is evidence of the efficacy of this approach and its potential for achieving near optimality. It should also be noted that for this example $T(d^G) = 123,287$. The value $T(d^G) - T(d^P) = 8112$ is the cost of meeting the constraints. A more through investigation of the relationship between aggregation number and the tightness of the constraints would be helpful in shedding additional light on the approximation expressed in Algorithm P.

4.3.2. *Variance of* $T(d)$

Another issue that has been of some concern is the relationship of the variances of the various solutions $T(d^G), T(d^I), T(y^*_{ljk})$ and $T(d^P)$. As discussed above current practice is to rely on Algorithm G. In addition to being sub-optimal and not accounting for business constraints, there is some risk associated with the variance on the return. There are several questions that can be posed.

- Is the variance of $T(d^G)$ within acceptable limits?
- Is there another decision function such that the resulting $T(d)$ has a preferable variance?
- Does the variance of $T(d^P)$ present greater risk than that of $T(d^G)$?

4.4. Summary

This offer optimization approach provides three significant improvements over other, more standard, approaches to the problem of campaign design.

- First and foremost, the developed solution produces significantly more incremental profit than competing solutions. As demonstrated in the tactical example, the campaign incremental profit is almost twice as high as that of the standard approach.
- Secondly, this technique is designed to implement multiple constraints and therefore affords the business more control over the direct marketing process. Attempting to satisfy several business constraints simultaneously using ad hoc techniques is a very labor-intensive task and generally produces poor results.
- Finally, the additional information that can be presented as a part of this solution can provide the business with more insight into the customer base, product offerings and the effects of the constraints.

This insight can be used to guide the company to craft better investment decisions in order to make future campaigns even more successful.

Acknowledgements

I am deeply indebted to Andrew Storey, Director of Analytical Services, Scotiabank, Toronto, Canada. Andy provided motivation, a sounding board, and a grounded perspective that kept this work moving forward in a way that added significant value. It is so easy to get involved in scientific and analytical work that is of interest in itself but has limited value in practice. Andy made sure that this work would continue to provide that kind of practical value that is the tradition of operations research. I cannot thank Andy enough.

I also would like to thank Dr. Yi Liao of the Management Sciences Department at SAS Institute for his careful reading of this paper and many helpful suggestions.

Further reading

[1] D. Campell, R. Erdahl, D. Johnson, E. Bibelnieks, M. Haydock, M. Bullock, H. Crowder, Optimizing customer mail streams at Fingerhut, Interfaces 31 (2001) 77–90.

[2] M. Cohen, Offer optimization, optimizing customer value, SAS Internal Technical Paper, May 2001, 2000.

[3] M. Cohen, J. Parks, Optimizing the allocation of cross-selling effort, Proceedings of the DiamondSug 2000 Conference, San Francisco, 2000.

[4] M. Cohen, Targeted marketing, optimizing the customer/event/channel assignment, SAS Internal Technical Paper, December 2000.

[5] N. Levin, J. Zahavi, The economics of selection of mail orders, Working paper. http://recanati.tau.ac.il/, (1999).

[6] G. Nemhauser, L. Wolsey, Integer and Combinatorial Optimization, Wiley, New York, 1988.

[7] D. Shepard (Ed.), The New Direct Marketing, 2nd Edition, Irwin Professional Marketing, Burr Ridge, IL, 1995.