



# Joint Optimization of Customer Segmentation and Marketing Policy to Maximize Long-Term Profitability

## Abstract

With the advent of one-to-one marketing media, e.g. targeted direct mail or internet marketing, the opportunities to develop targeted marketing campaigns are enhanced in such a way that it is now both organizationally and economically feasible to profitably support a substantially larger number of marketing segments. However, the problem of what segments to distinguish, and what actions to take towards the different segments increases substantially in such an environment. A systematic analytic procedure optimizing both steps would be very welcome.

In this study, we present a joint optimization approach addressing two issues: (1) the segmentation of customers into homogeneous groups of customers, (2) determining the optimal policy (i.e., what action to take from a set of available actions) towards each segment. We implement this joint optimization framework in a direct-mail setting for a charitable organization. Many previous studies in this area highlighted the importance of the following variables: R(ecency), F(requency), and M(onetary value). We use these variables to segment customers. In a second step, we determine which marketing policy is optimal using markov decision processes, following similar previous applications. The attractiveness of this stochastic dynamic programming procedure is based on the long-run maximization of expected average profit. Our contribution lies in the combination of both steps into one optimization framework to obtain an optimal allocation of marketing expenditures. Moreover, we control segment stability and policy performance by a bootstrap procedure. Our framework is illustrated by a real-life application. The results show that the proposed model outperforms a CHAID segmentation.

**Key words:** direct marketing, econometric models, sample selection, target selection, endogeneity.

# 1 Introduction

The concept of segmentation is central to marketing. A search on this keyword in article titles only resulted in more than thirty articles in the *Journal of Marketing*, more than fifty in the *Journal of Marketing Research*, etc. In the early marketing applications, the process of dividing a population of customers by means of clustering techniques into homogeneous groups was often done without the use of a dependent/target variable (Frank et al. (1972)). However, marketers realized that segmentation should not be an end in itself, but rather a means to an end. As most companies want to maximize profits (or some other quantity, e.g. sales), marketers quickly realized that a segmentation should ensure that 'better' customers are separated from other customers. This largely explains the popularity of clustering techniques using a dependent variable such as (CH)AID or CART (Breiman et al. (1984); Wedel and Kamakura (2000)).

The goal of separating the profitable segments of customers from the non-profitable ones, is to be able to differentiate marketing activities towards these segments. This requires that the segments are identifiable and targetable. Both properties are offered by the use of direct marketing media such a direct mail, catalogs and internet. They create the opportunity to use alternative marketing mix sets per segment or even per individual. Bass and Wind (1995) highlight the need for more research in direct marketing to achieve empirical generalizations in this field. Given that it takes substantial effort to design coherent marketing mixes, most companies still use the notion of a segment, rather than real one-to-one marketing. With the advent of one-to-one marketing media, e.g. digital printing for targeted direct mail or e-mail marketing, the opportunities to develop targeted marketing campaigns are enhanced in such a way that it is now both organizationally and economically feasible to profitably support a substantially larger number of marketing segments. However, the complexity and the dimension of the segmentation and of the identification of the appropriate marketing activity increases substantially in such an environment. A systematic analytic procedure optimizing both steps would be very welcome.

The quality of a specific segmentation is generally measured by within-segment homogeneity and between segment heterogeneity (Wedel and Kamakura (2000)). The marketing activity is optimized with the assumption that the segmentation is optimally determined and can remain fixed. However, the objective for the segmentation procedure may not be optimal for the overall performance of the direct marketing effort. It can be more desirable to determine a segmentation that performs well on other criteria than with respect to within-segment homogeneity. For instance, if we look at a direct marketing context, the (direct) marketer is interested in maximizing direct-mail returns. In this case, the most preferred segmentation is the one that maximizes profits and not within segment homogeneity.

In this study, we present a joint optimization approach addressing two issues: (1) the *segmentation* of customers into homogeneous groups of customers, (2) determining the *optimal policy* for marketing activities towards each segment.

The first step determines the segmentation of the customers. The proposed segmentation can be classified as an a-priori descriptive method (see Wedel and Kamakura (2000)): we determine segments before-hand, using only variables that describe personal characteristics of the customers. Hence, we do not look at the relationship between a dependent variable and some explanatory variables.

In a second step, we determine the optimal marketing policy using Markov decision processes, following similar applications such as Bitran and Mondschein (1996), Gönül and Shi (1998) and Piersma and Jonker (2000). The attractiveness of this stochastic dynamic programming procedure is based on the long-run maximization of expected average profit. Moreover, we control segment stability and policy performance by a bootstrap procedure.

The performance (= profitability) of the (optimal) marketing policy provides an indication of the quality of the segmentation determined in the first step. By using a local search method, multiple segmentations are evaluated, resulting in an optimal segmentation of the customers (optimal with respect to generated profits). We implement this joint optimization framework in a direct-mail setting for a charitable organization.

To summarize, we contribute to the existing literature by the specification of an opti-

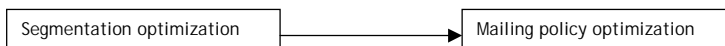


Figure 1: The general procedure to go from a segmentation to an optimal marketing strategy.

mization framework offering the integration of the two-step process of segmentation and targeting in one framework to obtain an optimal allocation of marketing expenditures.

This paper is structured as follows. In Section 2, we discuss our methodology in detail. In Section 3 we present an operationalization of the proposed methodology in a direct mailing context. The results of an empirical application of our model are presented in Section 4. We finish with a detailed discussion in Section 5.

## 2 Methodology

In marketing, segmentation is usually not a goal in itself, but rather a means to an end. The actual goal is to determine the most appropriate (= most profitable) marketing mix for each segment. Usually, these segments are determined first and then, on the basis of these segments, the marketing mix is determined (see Bitran and Mondschein (1996) and also Piersma and Jonker (2000)<sup>1</sup>), see Figure 1.

There is, however, no guarantee that the segmentation determined in the first step will lead to maximized profits. We propose a joint optimization routine that determines the most profitable customer segmentation and marketing strategy in terms of the best overall long-term performance. The optimization process considers a series of segmentations of the customers. For each segmentation, the best marketing policy in terms of long-term impact (= revenues) is determined using stochastic dynamic programming. Other segmentations are considered in a local search environment of the current segmentation until a (local) optimum is found or a stopping criterion is triggered. The stochastic dynamic programming procedure needs parameter estimates, especially transition proba-

---

<sup>1</sup>Both in Bitran and Mondschein (1996) and Piersma and Jonker (2000) first define a state space, which can be viewed as a segmentation if each state is seen as a segment. Secondly, the optimal marketing action is determined.

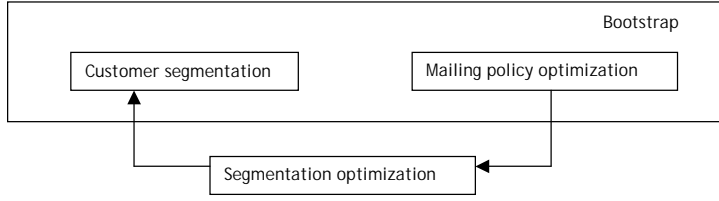


Figure 2: The proposed procedure to go from a segmentation to an optimal marketing strategy.

bilities between the segments over time and the expected response per customer segment. The marketing strategy optimization is bootstrapped to account for estimation errors. A genetic algorithm is applied to the segment optimization in combination with the optimization procedure for the marketing strategy.

In conclusion one can describe the proposed optimization process as shown in Figure 2.

## Step 1 - Segmentation

The segmentation we propose does not consider a relationship between certain (explanatory) variables and a dependent variable (such as response). We propose to use an a-priori descriptive method (see Wedel and Kamakura (2000)), where the segments are determined before-hand. The quality of the resulting segmentation is evaluated later, by the performance of the (optimal) marketing policy that results from this segmentation.

The customer segmentation is described in terms of breakpoints of variables. Variables are used that are expected to influence the response behavior of customers, such as socio-demographic variables or variables that describe past response behavior<sup>2</sup>.

For each variable  $x$  we define a limited number  $K(x)$  of breakpoints,  $B_{x,1}, \dots, B_{x,K(x)}$ ,  $x = 1, \dots, L$ . These breakpoints define the segmentation. For instance, all customers with scores below the first breakpoint  $B_{x,1}$  for variable  $x$  are separated from the customers with scores above this first breakpoint. Likewise all customers with a score for variable 1 between breakpoint  $C_{1,4}$  and  $C_{1,5}$ , a score between breakpoint  $C_{2,7}$  and  $C_{2,8}$  for variable

---

<sup>2</sup>Note that we do not explicitly model the relationship between these "explanatory" variables and a "dependent" response variable. But in order to determine "sensible" segments it is important that the chosen variables are expected to have a relation with the explanatory variable.

2,..., and a score for variable  $L$  above the largest breakpoint  $C_{L,K(L)}$  are within different segments. Notice that there are exactly  $U$  segments, with

$$U = \prod_{x=1}^L (K(x) + 1). \quad (1)$$

The values of the breakpoints thus fully specify the segmentation.

Standard segmentation techniques related to CHAID can be used to obtain a segmentation of the customers with respect to some objective. However, these partitioning tree techniques only ensure that segments are different with respect to some dependent variable (e.g. the total amount purchased during some period), i.e., they do not consider how customers move between segments over time. Moreover, they do not ensure 'optimal' splitting, because they do not re-evaluate this choice once a split variable and split value are chosen (Friedman and Fisher (1999)).

The model we propose determines the quality of a segmentation by looking at the profit of the optimal mailing policy for this segmentation. This policy is determined in the next step.

## Step 2 - Optimization

In this step the optimal marketing policy is determined for the given segmentation. There are different ways to determine the optimal strategy. One can choose to optimize short-term profits by using (non) linear regression models or long-term profits by using for instance a Markov decision model. When short-term profits are maximized, one is only interested in the profits that can be earned in the next period. It is not taken into account that the customer can or will be approached again in the future.

One can also choose to maximize long-term profits. The idea is to determine a strategy that maximizes the total discounted net (future) revenues earned from an individual customer. These discounted net revenues are often called the Customer Lifetime Value (or CLV) (Blattberg and Deighton (1996), Berger and Nasr (1998)). A Markov Decision model can be used to maximize CLV (see Bitran and Mondschein (1996) and Piersma and Jonker (2000)).

## Step 3 - Local Search Methods

Our goal is to determine a segmentation (and its associated optimal marketing strategy) that maximizes profits. In order to find a new candidate segmentation we propose to adopt a local search method. This method will look for an alternative segmentation in the neighborhood<sup>3</sup> of the current segmentation. For practical reasons we propose to use a local search method as there are no clear limits to the space of possible solutions (the number of breakpoints for each variable is unbounded). It is also rather straightforward to formulate a neighborhood of a current segmentation, for instance a small interval for each breakpoint for each variable. The search methods that can be used are for instance simulated annealing, tabu search and genetic algorithms (see Laarhoven and Aarts (1987), Glover and Laguna (1997), Holland (1975) and Davis (1991)).

## 3 Operationalization for Direct Marketing

To date, the largest part of the direct marketing literature focuses on address selection for a single mailing or catalog (target selection). A review of this literature can be found in Baesens et al. (2002) or Jonker et al. (2002). Apart from this stream, we find several approaches to optimizing direct marketing messages that are somewhat related to our proposed framework.

The approach of DeSarbo and Ramaswamy (1994) seems comparable to the one we propose. They consider combining the segmentation and targeting issue in one framework (their model is in essence a latent class formulation of a probit model). They use the model to determine which customers should be included on the mailing list for the next period. However, their model does not determine an optimal (in terms of profits) mailing strategy, as they only model the binary (yes or no) response. Second, the model has a single-period horizon and does not take future mailings into account.

Three models that do have a multi-period horizon are Gönül and Shi (1998), Bitran and Mondschein (1996) and the one discussed in Piersma and Jonker (2000). All three

---

<sup>3</sup>A neighborhood is defined as a number of alternative solutions whose breakpoints differ only slightly from those of the current solution.



determine the optimal (long-term) mailing strategy using a Markov decision model. However, the segmentation and the mailing policy decision step are considered separately. The state space (or segmentation) is determined before-hand, and given this segmentation, the optimal mailing policy is determined. Our approach considers several segmentations and determines the one that is optimal.

This section describes how our model is operationalized in a direct mailing context. First, we look at the segmentation, and motivate which variables are used for segmentation. Next, we explain the method that is used to obtain the optimal mailing policy, the Markov Decision process. Lastly, we describe the search algorithm that is used to look for different segmentations (Genetic Algorithm).

## **Step 1 - Segmentation**

### **Bases for Segmentation**

In direct marketing, the RFM variables are the most likely candidates as bases for segmentation (see also Gönül and Shi (1998), Bult and Wittink (1996), Bitran and Mondschein (1996)). The RFM variables measure consumer response behavior in three dimensions (see also Jonker et al. (2002)). The first dimension is Recency, which indicates how long it has been since the customer has last responded. The second is Frequency, which provides a measure of how often the customer has responded to received mailings. And finally Monetary Value measures the amount of money or the number of products that the customer has spent in response to the mailings.

Various operationalizations of the RFM variables have been used in the literature. The specific definition usually depends on a number of case specific elements (for instance catalogs vs. mailings). But, as we have indicated in Donkers et al. (2001), it is important to formulate the variables such that they measure response behavior of individuals and not (past) mailing behavior of the direct mailer. Therefore we propose to use the following RFM variables:

1. Recency ( $R$ ), the number of mailings without a response since the last response of the customer.

2. Frequency ( $F_1$ ), the response percentage over the last two years, defined by the total number of responses over the past two years divided by the number of mailings in the past two years.
3. Frequency ( $F_2$ ), the response percentage over the whole period of registration.
4. Monetary ( $M_1$ ), the average size of the responses over the past two years.
5. Monetary ( $M_2$ ), the average size of the responses over the whole period of registration.

Note that  $F_1$  and  $M_1$  explicitly measure (relative) short term response behavior, whereas  $F_2$  and  $M_2$  measure long-term response behavior. We have deliberately chosen to include variables which measure short-term as well as long-term response behavior in order to rule out the effects of the mailing strategy as much as possible.

For example, suppose individual  $i$  always responds to a mailing that (s)he receives and (s)he then spends 10 guilders. The  $R$  is always equal to zero,  $F_1$  and  $F_2$  are equal to one and  $M_1$  and  $M_2$  are equal to 10. If the decision rule of the direct mailer indicates not to send a mailing to  $i$  (for instance because of a budgetary constraint) in year  $t$  and also in year  $t + 1$ , the score on  $F_1$  and  $M_1$  will have become zero. But  $R$  will also be zero, because  $i$  has responded to the last mailing (s)he received. Now if the direct mailer does not incorporate long term measures of response behavior (such as  $F_2$  and  $M_2$ ) when determining the selection,  $i$  probably won't be selected again, whereas (s)he is a profitable customer (if we assume the cost of sending a mailing is equal to 1). Hence, valuable response behavior might be lost if long-term response measures are not incorporated.

### **Defining a Segmentation**

As described in Section 2, a segmentation is defined by the breakpoints for the variables. The number of segments increases rapidly as the number of breakpoints increases. For instance, if each of the five RFM variables from the previous Section has one breakpoint, there will be ( $2^5 =$ ) 32 segments, as each variable is divided in two parts when one has one breakpoint. Similarly, if there are two breakpoints for each variable, there will be ( $3^5$

=) 243 segments. Hence, the size of the search space remains limited, as a reasonable sized state space is already obtained with a small number of breakpoints<sup>4</sup>.

There is another reason to limit the number of breakpoints. If there are too many segments, the transition probabilities will become difficult to estimate as there can be very few observations per transition. This makes the transition probabilities less reliable. Hence, a trade-off has to be made between the number of breakpoints (= segments) and the reliability of the estimates of the parameters. In order to prevent the presence of segments which only have a few members, our program combines segments with too few members with neighboring segments.

## Step 2 - Optimization

Some recent models describe this dynamic process by means of a Markov decision process (Gönül and Shi (1998), Bitran and Mondschein (1996), Piersma and Jonker (2000), Van den Poel and Van Kenhove (2002)). These models describe the migration of an individual customer in terms of transitions between states. A state records the segment that the customer belongs to at a decision moment. Migration between the states will occur as a result of the mailing policy and the response of the customer in a time period between two decision moments. For each customer the states are recorded at a number of decision moments. Let  $a$  be the mailing decision at a decision moment for all customers who are observed in state  $s$  ( $a = 1, \dots, A$ ,  $s = 1, \dots, S$ ). The transition probability to observe a migration from state  $s$  to state  $t$  after mailing decision  $a$  is denoted by  $p_{s,t}(a)$  ( $s, t = 1, \dots, S$ ). At each decision moment, action  $a$  triggers an immediate reward  $r_s(a)$  for all customers observed in state  $s$  at the decision moment. This reward is given by the response of the customer till the next decision moment. The direct marketing firm however may be more interested in long-term profits. We will consider the long-term discounted reward as the objective for mailing optimization. Bitran and Mondschein (1996) show that this objective is closely related to lifetime value of the customers. Other long or short term objectives can be used as desired. The Markov decision model will determine

---

<sup>4</sup>The segmentation described here can also be seen as a state space formulation of the Markov model, which will be discussed in the next subsection.

the optimal mailing policy by the well-known value iteration or policy iteration algorithm or by linear programming (Ross, 1983, p. 40). The parameters of the model that have to be estimated before the optimization routine can start are the transition probabilities for each action  $a$  and the immediate reward in each state when action  $a$  is chosen. The mathematical formulation of the model can be found in Piersma and Jonker (2000).

The estimation of the parameters of the Markov decision model introduces an estimation error in the expected profit for each mailing policy. Most critical is correct estimation of the transition probabilities  $p_{st}(a)$ . Sparse data sets are known to obstruct reliable estimation for the transition probabilities. To account for these problems in the parameter estimation we use a bootstrap technique (Efron (2000)). More specifically, we want to assess the stability of a solution by calculating the bootstrap mean and standard deviation. In our setting, we only accept a solution as 'better' if it outperforms a given solution (means are higher), and if (mean - std.dev.) of the new solution is higher than the (mean - std.dev.) of the old solution. In general, bootstrapping takes a random sample with replacement of the same size as the original database. Bootstrapping enables estimating the distribution of any statistic (Efron and Tibshirani, 1993, p. 55-56, 162-177). For each bootstrap sample the model considered is estimated separately. In our application, we use the procedure to obtain a more reliable estimate of the average performance of the selected segmentation scheme (Hruschka (2002)). This is accomplished by drawing bootstrap samples from the dataset, followed by assigning the customers to the segmentation scheme under investigation and optimizing the mailing policy.

### **Step 3 - Local Search Methods**

Since its introduction by Holland (1975) genetic algorithms (GA) have shown good results in numerous applications. There is a vast literature on GA (see e.g. Davis (1991)), including studies on its theoretical and practical performance and many extensions of the basic algorithm. For our application we consider the Simple Genetic Algorithm (SGA) as defined by Goldberg (1989). Although we realize that other, possibly more sophisticated GA formulations exist, we feel that SGA is best suited for our application because its

simplicity will lead to a feasible running time of the algorithm. The key feature of GA has been formulated by (Balakrishnan and Jacob, 1996, p.1108) as: "A key feature of GA is that the search is conducted from a population rather than a single point thus increasing the exploratory power of GA. In contrast with other gradient search techniques which need additional knowledge such as the differentiability of the function, GA use the objective function or payoff directly, in that the candidate [solutions] are evaluated based on the specific objective."

The genetic algorithm starts with a population of randomly selected solutions for the segmentation of the customers based on the RFM variables. A solution can be denoted by the specification of the breakpoints for each of these variables. All solutions have two breakpoints for every variable, but a breakpoint can have values outside of the range of the variable indicating that the breakpoint is inactive. We also include the CHAID solution in the starting population.

One then evaluates the goodness of each solution with respect to the current population. This evaluation consists of finding the optimal mailing policy for this segmentation and determining the expected profit of this policy. Solutions that result into a higher profit are given more chance to "reproduce" than others. In particular, we draw with replacement  $2k$  solutions from the population  $P$  where each solution has probability

$$\frac{f(z)}{\sum_{y \in P} f(y)} \quad (2)$$

where  $f(z)$  indicates the profitability of solution  $z$ .

### Crossover

Divide the subpopulation into pairs. For each pair: Randomly select a variable number (number between 2 and 5), call this number index. Make two new solutions: The first new solutions consists of the variable breakpoints smaller to index of the first solution and the variable breakpoints greater or equal to index of the second solution. The second new solution consists of the variable breakpoints smaller to index of the second solution and the variable breakpoints greater or equal to index of the first solution.

## Mutation

Both new solutions of each pair are mutated as follows: For every variable the breakpoints are mutated with a certain small probability as follows:

- Recency ( $R$ ): the value of the breakpoints is raised by one (if possible)
- Frequency ( $F_1, F_2$ ): the value of the breakpoints is lowered with a random value between 0 and 0.1 (if possible)
- Monetary value ( $M_1, M_2$ ): the value of the breakpoints is raised to the next giftsize (if possible)

As a result there are  $2k$  new solutions that are complemented with the  $n - 2k$  best solutions from the previous population. Repeat the process of selection, crossover and mutation a given number of times.

For our study we emphasize the importance of a general search technique for the segmentation. GA is well suited for this application because it considers multiple segmentations in each generation. Moreover, it uses part of the most promising segmentations for its search for new segmentations. Since there is no intuition on the best combination of breakpoints for the RFM variables, it is better to search in multiple directions at the same time. Other local search techniques such as simulated annealing or taboo search consider only one segmentation in each step of the algorithm. The local search can consider a less profitable choice for one of the breakpoints while searching for a local optimum for a long time.

## 4 Empirical Application

The model is applied to a database from a large Dutch charitable organization. This database contains information on mailings sent between 1994 and 1999 to about 600,000 individuals. For illustrative purposes, we provide descriptive statistics on the five RFM variables  $R$ ,  $F_1$ ,  $F_2$ ,  $M_1$  and  $M_2$ , see Table 1. From this table we see that, on average, a person has received five mails since the last mail to which (s)he responded. An average

donor responds to 4 out of 10 mails, and in the last two years (s)he has responded to 3 out of 10 times. When a donor responds, (s)he donates (on average) 16 guilders. In the last two years, a donor has donated on average 12 guilders per response.

The transition probability matrices  $P_{i,j,a}$  are constructed from the data. They describe the probability of going from state  $i$  in period  $t$  to state  $j$  in period  $t + 1$ , given a certain action  $a$ . This action  $a$  is the number of mailings that were received in period  $t$ . The states  $i$  and  $j$  are described by the five RFM variables. The total number of states depends on the number of breakpoints, see equation (1).

The data are also used to construct the revenues earned while being in a certain state.

## Genetic Algorithm

The genetic algorithm is tested with the following parameters:

The starting population of our model consists of 200 solutions (the solution of a CHAID analysis and 199 other randomly generated solutions). A solution consist of a maximum of two breakpoints per variable, the associated fitness of the solution, and the standard deviation of this fitness. The fitness and standard deviation are calculated over 100 (bootstrap) samples.

The first generation of (almost all) randomly generated solutions has an average fitness of 1,877, with a minimum fitness of 585 and a maximum of 4,429. After 50 generations, the average fitness has significantly increased to 4,848, with a minimum of 1,906, and a maximum of 7,447. This is also a big improvement compared to the fitness of the CHAID solution (1,970 with a standard deviation of 230).

Table 3 shows that the fitness of the best twenty solutions is considerably higher than that of the CHAID solution. However, the standard deviation of the solution of the CHAID solution is also lower. The problem is that the standard deviation always has to be considered relative to the mean. Therefore we look at average fitness divided by the average standard deviation (to get something like a t-value). The "t-value" of the CHAID solution is 8.57, which is comparable to the values found for the best twenty solutions (see column four of Table 3).

Table 1: Descriptive statistics of RFM variables.

	Average	t-value
Recency		
Number of missed mails	5.32	(0.65)
Frequency		
Over entire period	0.42	(0.63)
Over last two years	0.34	(0.91)
Monetary Value		
Over entire period	16.38	(2.15)
Over last two years	12.20	(2.93)

Table 2: Input Parameters Genetic Algorithm.

Parameter	Value
Bootstrap size	100
Population size	200
Mutation probability	0.1
Size of subsample	20
Number of generations	100



When we compare the "t-values" of generation 1 and 50, we see that it has decreased from 10.8 to 8.1. Although the absolute value of the standard deviation has increased from 174 to 597, the relative quality of the solution has not decreased.

Next we compare the breakpoints of our solutions in generation 50 with the breakpoints of the CHAID solution. The CHAID solution is:

- $F_1 = 0.3650$ ;  $F_2 = 0.0700$  &  $0.7300$ ;  $M_1 = 5.835$

The CHAID solution has no breakpoints for  $R$  and  $M_2$ . If we look at the values of the breakpoints for the CHAID solution and the 20 best solutions obtained in generation 50 (see Table 4), we see that CHAID does not use a breakpoint for  $R$  and  $M_2$ , whereas all 20 best solutions do use (at least) one.

## Breakpoints

When we look at Table 4 at a glance, we see that there are only a limited number of different values for the breakpoints. A large variation in the values of the breakpoints would be a strong indication that the method converges to a number of varied (local) optima. Fortunately, this does not seem the case here. Next, we want to evaluate more precisely how large the variation in the values of the breakpoints is for the best twenty solutions of generation 50.

First, we examine the values of the breakpoints for each of the five variables for the best twenty solutions in generation 50. Looking at  $R$  (Recency), we see that of the best twenty solutions, sixteen have only one breakpoint. In ten cases the breakpoint is equal to 3 and in five it is equal to 10. Four of the other five solutions can be seen as mutations of these two prevailing solutions. In one case there is one breakpoint, which is equal to 4. In the local search method, this point is a direct neighbor of the frequently encountered solution with the breakpoint equal to 3. The same goes for three cases where there are two breakpoints: one is equal to 3 or 10 and the other is equal to one. Hence, nineteen of the twenty best solutions can be contributed to (mutations of) two values for the breakpoints.

For the second variable  $F_1$  (Frequency measured over the last two years), sixteen have only one breakpoint. In eight cases the breakpoint is equal to 0.5909, in five cases it is

equal to 0.6364 and in two cases it is equal to 0.4545. In three of the four cases where there are two breakpoints, they are equal to 0.0455 and 0.4091. Although there is somewhat more variation in the outcomes than for  $R$ , it is clear that in most cases there is only one breakpoint with a value between 0.45 and 0.64.

Next, we look at  $F_2$  (Frequency measured over the entire period of registration). There, seven of the best twenty solutions have one breakpoint. In six of these, the breakpoint is equal to 0.5909. Of the thirteen solutions with two breakpoints, four have breakpoints 0.4091 and 0.7273 and two have breakpoints 0.5909 and 0.7727. The remaining seven solutions are relatively varied. The general picture is that either there is one breakpoint (around 0.5909) or there are two breakpoints, where one is between 0.67 and 0.77 and/or one is between 0.3 and 0.45.

All best twenty solutions have two breakpoints for  $M_1$  (Monetary value measured over the last two years). In eight cases, the breakpoints are equal to 12.37 and 21.56. In three cases, the breakpoints are 5.71 and 12.37 and in two cases 12.4 and 21.67. More generally speaking, there are thirteen cases where there is a breakpoint around 12.4 and there are twelve cases where there is a breakpoint between 21 and 25.

Again, for  $M_2$  (Monetary value measured over the entire period of registration), all best twenty solutions have two breakpoints. In eight cases the breakpoints are equal to 1.57 and 7.67. There are also two cases where the breakpoints are 1.6 and 7.71. Of the remaining ten solutions, there are seven with one breakpoint between 4.75 and 5. The most found combinations are 4.75 and 33.12 (three times) and 4.75 and 6 (two times). Hence, there seem to be two more general solutions: one with breakpoints 1.57 and 7.67 and one with a breakpoint around 4.75 in combination with another breakpoint.

In conclusion, we can say that when looking at the individual variables, the breakpoints do converge to a limited number of values (see also Tables 5-7). Generally speaking,  $R$  has one breakpoint, with a value of 3 or 10.  $F_1$  has one breakpoint, equal to 0.5.  $F_2$  has either one breakpoint (at 0.5909) or two breakpoints (between 0.67 and 0.77 and/or one between 0.3 and 0.46).  $M_1$  has two breakpoints, one around 12,4 and one between 21 and 25. Finally,  $M_2$  has two breakpoints: 1.57 and 7.67 or 4.75 and another breakpoint. If we

look at the twenty solutions and compare them, we see that seven match all these criteria. However, seventeen of the best twenty solutions violate at most one assumption (if we do not consider mutations as a violation). This supports the claim that most solutions have breakpoints that are in the same range.

Two solutions occur three times:

- $R = 3$ ;  $F_1 = 0.5909$ ;  $F_2 = 0.4091$  &  $0.7273$ ;  $M_1 = 12.37$  &  $21.56$ ;  $M_2 = 1.57$  &  $7.67$
- $R = 10$ ;  $F_1 = 0.6364$ ;  $F_2 = 0.5909$ ;  $M_1 = 12.37$  &  $21.56$ ;  $M_2 = 1.57$  &  $7.67$

This illustrates that there is considerable convergence of the solutions to a limited number of (comparable) solutions.

This is confirmed if we compare all 200 solutions of generation 1 and all 200 of generation 50. In Table 5, Table 6 and Table 7 we see that the values for the breakpoints are more clustered for generation 50 than for generation 1.

## Chaid

It has to be noted that solutions 4–11 have a rather high standard deviation. This can have two causes (i) the sample used in the bootstrap causes more uncertainty in the outcomes. If this is the case, the high standard deviations would have to occur more or less independently of the values of the breakpoints. However, as we can see in Table 4, solutions 6, 7, 8 (and 10) are (almost) identical. This could indicate some instability of the solution. However, even though the standard deviations are somewhat high, the fitness is still highly significant.

## Improvement over Generations

Next we want to evaluate how quickly the fitness improves. Figure 3 shows the fitness of each of the 200 solutions for every other five generations. The solutions are ordered from lowest (number 1) to highest (200) fitness. The figure clearly shows that the fitness improves over the generations and that the increase in improvement declines in the later

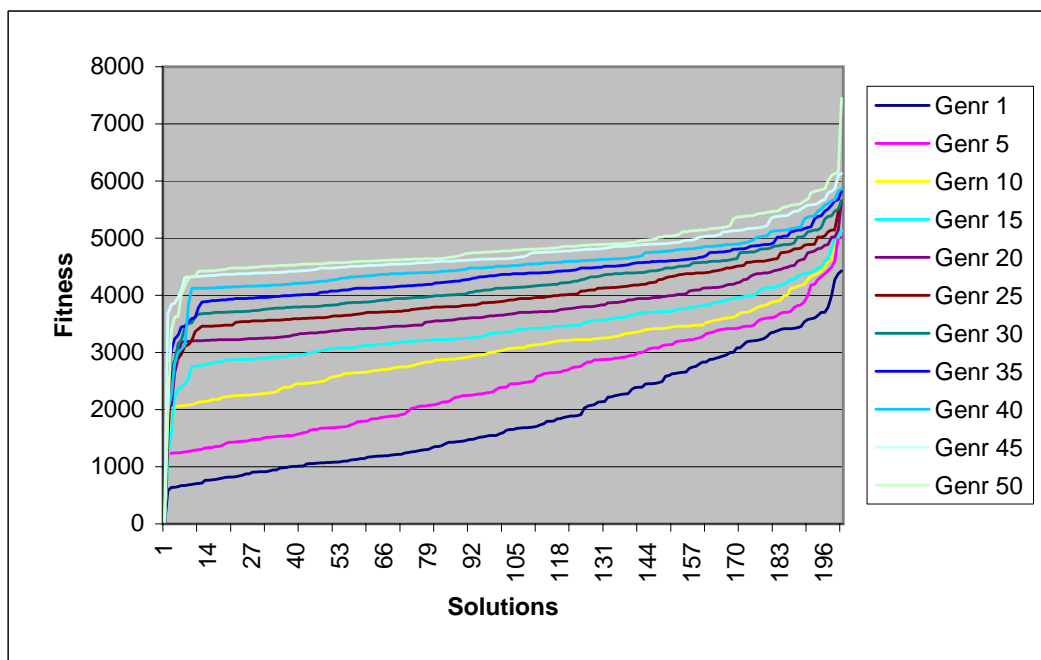


Figure 3: Fitness all solutions for different generations.

generations. This indicates that the fitness is converging to a maximum. However, the best 30 solutions are still improving for the higher generations.

## 5 Discussion

Both segmentation and optimization of marketing actions have been studied extensively in marketing, see Wedel and Kamakura (2000) and Sasieni (1989). However, until now these two processes have not been considered in one unifying framework. Usually, the optimization is preceded by the segmentation. The segments are determined using a certain criterium, and then the optimal marketing actions are determined for each segment using another criterium. In this paper we propose a method that determines the optimal segmentation and the optimal actions using the same criterium.

The proposed method consists of three steps. First we determine a segmentation. This initial segmentation can be randomly determined, or one can use the outcome of a preliminary analysis such as CHAID. Second, we determine the optimal action for each

segment. Here we use an optimization technique (such as a Markov decision model) and determine a strategy (a collection of actions) that maximizes profits. And finally, in the third step, we look for an alternative segmentation. Various (local) search techniques such as genetic algorithms can be used to determine new segmentations. We then return to step two and investigate the profitability of the alternative segmentation. This iterative search process is stopped when we have found an optimal solution.

We have applied this method in a direct mailing framework. The goal here is to determine the optimal mailing strategy, where we define a mailing strategy as the number of mailings each segment will receive during one year and optimal means a strategy that maximizes the discounted future net revenues.

The results show that our model leads to a significant improvement over CHAID, a model that determines an optimal strategy given a segmentation. We also see that the best segmentations proposed by our method are almost identical. This indicates that our method does not converge to various different local optima.

However, this study also has some limitations and there remain a number of issues for future research. A limitation is that the proposed method requires a couple of years of information, otherwise the results will become less reliable. There are also a number of elements which could be further examined. First, one could consider alternative definitions of a segmentation. In this study, a segmentation was defined by the "breakpoints". An alternative could be to combine the scores on the three RFM dimensions into one variable, and then use this new variable to segment the customers. A problem there would be to determine the weights for each dimension. A second issue for future research would be to look at global search methods to determine the optimal segmentation.

Table 3: Fitness, standard deviation and "t-values" of the best twenty solutions from generation 50.

Number	Fitness	Standard deviation	"t-value"
1	7447.3	531.5	14.01
2	6154.5	754.3	8.16
3	6128.6	605.3	10.12
4	6097.5	1660.8	3.67
5	5998.3	2746.9	2.18
6	5863.2	1278.6	4.59
7	5849.5	1284.1	4.56
8	5833.6	1291	4.52
9	5821.8	1723	3.38
10	5795.6	1350.6	4.29
11	5676.5	1310.8	4.33
12	5654.7	517.4	10.93
13	5608.9	101.8	55.10
14	5590.5	80.1	69.79
15	5582.5	516	10.82
16	5577.4	1807.9	3.09
17	5553.6	481.9	11.52
18	5548.6	341	16.27
19	5511.5	318	17.33
20	5475	3040.8	1.80

Table 4: Best twenty solutions from generation 50.

Recency		Frequency 1	Frequency 2		Monetary 1		Monetary 2			
1	3	0.4545	0.2727	0.3636	5.71	12.37	4.75	33.12		
2	3	0.5909	0.5909		7.50	15.63	4.80	118.50		
3	1	10	0.4545	0.2727	0.3636	5.71	12.37	4.75	33.12	
4	3	0.5909	0.0455	0.4545	11.67	33.00	2.86	20.67		
5	3	0.5909	0.4091	0.7273	12.37	21.56	1.57	7.67		
6	10	0.6364	0.5909		12.37	21.56	1.57	7.67		
7	10	0.6364	0.5909		12.37	21.56	1.57	7.67		
8	10	0.6364	0.5909		12.37	21.56	1.57	7.67		
9	3	0.5909	0.4091	0.7273	12.37	21.56	1.57	7.67		
10	10	0.6364	0.5909		12.40	21.67	1.6	7.71		
11	3	0.5909	0.4091	0.7273	12.37	21.56	1.57	7.67		
12	10	0.6364	0.5455		6.86	425.00	4.75	6.00		
13	1	3	0.0455	0.4091	0.5909	0.7727	12.37	21.56	1.57	7.67
14	1	3	0.0455	0.4091	0.5909	0.7727	12.37	21.56	1.57	7.67
15	3	0.0455	0.4091	0.4091	0.7273	12.40	21.67	1.60	7.71	
16	2	6	0.5455	0.6364	0.1170	0.2054	10.33	24.75	2.43	28.67
17	3	0.5909	0.5909		6.86	425.00	4.75	6.00		
18	3	0.5909	0.3396	0.6785	11.25	107.14	5.00	33.57		
19	4	0.5000	0.6818		5.71	12.37	4.75	33.12		
20	3	0.5909	0.0455	0.4545	22.86	23.86	1.00	1.83		

Table 5: Breakpoints for  $R$  for generation 1 and generation 50.

Value	Generation 1	Generation 50		
0	20	16	35	46
1	17	20	23	37
2	25	18	16	31
3	17	14	61	7
4	16	13	20	11
5	15	19		3
6	15	15	18	5
7	16	21	10	3
8	13	13	1	5
9	9	13	10	1
10	18	22	1	42
11	9	9		9



Table 6: Breakpoints for  $F_1$  and  $F_2$  for generation 1 and generation 50.

Value	Generation 1				Generation 50			
	$F_1$	$F_2$	$F_1$	$F_2$	$F_1$	$F_2$	$F_1$	$F_2$
0.0	11	12	6	7	52	10	21	
0.01-0.10	17	20	20	14	26	6	5	6
0.11-0.20	16	18	12	13	8	2	16	2
0.21-0.30	13	20	16	25	5	25	39	34
0.31-0.40	20	11	19	13	42	22	9	31
0.41-0.50	26	23	21	22	47	41	26	37
0.51-0.60	21	15	23	16	41	16	30	27
0.61-0.70	15	14	17	21	25	12	19	24
0.71-0.80	20	15	20	22	4	3	30	6
0.81-0.90	15	24	10	17				
0.91-1.00	22	20	31	22			16	5

Table 7: Breakpoints for  $M_1$  and  $M_2$  for generation 1 and generation 50.

Value	Generation 1				Generation 50			
	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$	$M_1$	$M_2$
0-10	42	49	54	47	31	47	127	90
11-20	47	46	52	37	81	83	35	6
21-30	18	22	18	30	67	3	21	49
31-40	11	11	15	19	3	19	11	49
41-50	18	8	10	3	13		1	
51-60	9	9	3	10				
61-70	5	8	5	6		1		
71-80	10	6	6	18				
81-90	7	2	6	6				4
91-100	3	7	5	4		1		
101-110	2	5	2	3		25		
> 110	28	27	24	17	5	21	5	2

## References

- Baesens, Bart, Stijn Viaene, Dirk Van den Poel, Jan Vanthienen and Guido Dedene (2002), 'Using bayesian neural networks for repeat purchase modelling in direct marketing', *European Journal of Operational Research* **138**(1), 191–211.
- Balakrishnan, P.V. (Sundar) and Varghese S. Jacob (1996), 'Genetic algorithms for product design', *Management Science* **42**(8), 1105–1117.
- Bass, Frank M. and Jerry Wind (1995), 'Introduction to the special issue: empirical generalizations marketing', *Marketing Science* **14**(3), G1–G6.
- Berger, Paul D. and Nada I. Nasr (1998), 'Customer lifetime value: Marketing models and applications', *Journal of Interactive Marketing* **12**(1), 17–30.
- Bitran, Gabriel R. and Susana V. Mondschein (1996), 'Mailing decisions in the catalog sales industry', *Management Science* **42**(9), 1364–1381.
- Blattberg, Robert C. and John Deighton (1996), 'Manage marketing by the customer equity test', *Harvard Business Review* pp. 136–144.
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen and Charles J. Stone (1984), *Classification and Regression Trees*, Chapman and Hall, New York.
- Bult, Jan Roelf and Dick R. Wittink (1996), 'Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising', *International Journal of Research in Marketing* **13**, 215–226.
- Davis, Lawrence (1991), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- DeSarbo, Wayne S. and Venkatram Ramaswamy (1994), 'Crisp: Customer response based iterative segmentation procedures for response modeling in direct marketing', *Journal of Direct Marketing* **8**(3), 7–20.

- Donkers, Bas, Jedid-Jah J. Jonker, Richard Paap and Philip Hans B.F. Franses (2001), Deriving target selection rules from endogenously selected samples, ERIM Report ERS-2001-68-MKT, Erasmus University Rotterdam.
- Efron, Bradley (2000), ‘The bootstrap and modern statistics’, *Journal of the American Statistical Association* **95**(452), 1293–1296.
- Efron, Bradley and Robert J. Tibshirani (1993), *Introduction to the Bootstrap*, Chapman and Hall, New York, NY.
- Frank, Ronald E., William F. Massy and Yoram Wind (1972), *Market Segmentation*, Prentice Hall, Englewood Cliffs, NY.
- Friedman, Jerome H. and Nick I. Fisher (1999), ‘Bump hunting in high-dimensional data’, *Statistics and Computing* **9**(2), 123–143.
- Glover, Fred and Manuel Laguna (1997), *Tabu Search*, Kluwer Academic Publishers, Boston, MA.
- Goldberg, David E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA.
- Gönül, Füsün and Meng Ze Shi (1998), ‘Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models’, *Management Science* **44**(9), 1249–1262.
- Holland, John H. (1975), *Adaption in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.
- Hruschka, Harald (2002), ‘Market share analysis using semi-parametric attraction models’, *European Journal of Operational Research* **138**(2), 212–225.
- Jonker, Jedid-Jah J., Philip Hans B.F. Franses and Nanda Piersma (2002), Evaluating direct marketing campaigns; recent findings and future research topics, ERIM Report ERS-2002-26-MKT, Erasmus University Rotterdam.

- Laarhoven, Peter J.M. and Emile H.L. Aarts (1987), *Simulated annealing: theory and applications*, Reidel, Dordrecht.
- Piersma, Nanda and Jedid-Jah J. Jonker (2000), Determining the direct mailing frequency with dynamic stochastic programming, Econometric Institute Report EI2000-34/A, Erasmus University Rotterdam.
- Ross, Sheldon M. (1983), *Introduction to Stochastic Dynamic Programming*, Probability and Mathematical Statistics, A Series of Monographs and Textbooks, Academic Press, San Diego, CA.
- Sasieni, Maurice W. (1989), 'Optimal advertising strategies', *Marketing Science* **8**(4), 358–370.
- Van den Poel, Dirk and Patrick Van Kenhove (2002), Customer relationship management by optimising database marketing actions toward segments to maximise long-term profitability, Working paper, Ghent University.
- Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Dordrecht.