Media Planning by Optimizing Contact Frequencies

Nanda Piersma[1]

Sander Kapsenberg[2]

Peter Kloprogge[3]

Albert P.M.Wagelmans[4]

18 December 1998

Econometric Institute Report EI 9856/A

Abstract

In this paper we study a model to estimate the probability that a target group of an advertising campaign is reached by a commercial message a given number of times. This contact frequency distribution is known to be computationally difficult to calculate because of dependence between the viewing probabilities of advertisements. Our model calculates good estimates of contact frequencies in a very short time based on data that is often available. A media planning model that optimizes effective reach as a function of contact frequencies demonstrates the usefulness of the model. Several local search procedures such as taboo search, simulated annealing and genetic algorithms are applied to find a good media schedule. The results show that local search methods are flexible, fast and accurate in finding media schedules for media planning models based on contact frequencies. The contact frequency model is a potentially useful new tool for media planners.

Keywords: media planning, optimization, effective reach, contact frequency

[1] Corresponding author: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: piersma@few.eur.nl.

[2] PLS Logistics, Rotterdam, The Netherlands, currently: CMG, The Netherlands.

[3] PLS Logistics, Rotterdam, The Netherlands

[4] Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

1. Introduction

Media planners are faced with the decisions how often and when to broadcast a commercial message. A media schedule gives a detailed account of the advertising effort in terms of the broadcasting times and chosen media. The aim is to find a media schedule that maximizes the advertising effect given a limited budget and preference restrictions on this schedule. A number of mathematical models have been developed for this planning process (see Little and Lodish (1969), Aaker (1975), Rust (1985), Ratham et al. (1992) and the overview by Leckenby and Ju (1989)), using different objective functions to model the advertising effect.

Contact frequency is the number of exposures that the target group of an advertising campaign receives for a given media schedule. Many studies have been conducted on the desired frequency to convey the advertising message, for different products and media classes (overviews in Little (1979), Naples (1979) and Got Plenty (1997)).

Our main contribution is a mathematical model that accurately estimates the contact frequency distribution for target groups. The data that is needed consists only of the average probability to see one commercial message, from a set of planning options, and the average probability to see any combination of two messages from this set. Furthermore it is possible to search for media schedules that optimize objective functions based on contact frequency. We show that all commonly used objectives can be seen as a function of contact frequencies. Therefore the contact frequency model can be used as the basis for media planning models that include many earlier models.

Media planning based on contact frequency results into a non-linear optimization model. This planning model can be efficiently solved using local search techniques. Earlier media planning models have been solved by simple local search (Little and Lodish (1969), Aaker (1975)). However, our second contribution is to apply the powerful local search techniques that have emerged in the recent literature such as taboo search, simulated annealing and genetic algorithms. The combination of the proposed model and the solution methods seems to work very well. The model is flexible and easy to apply in practice. In this paper we will focus on a basic media planning model (with few restrictions), for which we compare the performance of the various local search techniques. In the last section some extensions of the model, which can be implemented without much difficulty, are mentioned.

In the next section, we will define an estimation procedure for the contact frequency distribution. Also, a media planning model is introduced that optimizes effective reach (as a function of contact frequencies). The local search methods are discussed in Section 3 and applied to the planning problem for a test set in Section 4. In Section 5 we will show how the results can be used in the media planning process. In the remainder of the paper we will limit ourselves to TV advertisements.

2. A mathematical model for block selection

2.1. The contact frequency distribution

The past exposure of a target group to a broadcasted media schedule can be estimated by collecting data on the viewing behavior of a number of representative target group members. Predictions of the exposure to future advertisements have to be estimated in some way. If a target group is homogeneous then the exposure of the target group can be represented by an aggregate viewing probability. Otherwise the target group is segmented into homogeneous subgroups and the future exposure of every subgroup is modeled separately.

TV commercials are usually shown in blocks consisting of several commercials, the so-called commercial breaks. We will express future exposure by the probability to see the entire block. The planning horizon limits the number of blocks that the planner can choose to show a commercial. The media planning problem is to select a subset of blocks that will contain the commercial of interest to achieve maximal (effective) reach.

Calculating the contact frequencies for target groups is computationally difficult because there is dependence in the viewing probabilities of advertisements. If the dependence is ignored then the following method can be used to calculate the contact frequencies. Assume that the viewing probabilities of single advertisements are available. The probability to see exactly j advertisements when there are n broadcasted advertisements is denoted by $f(j|n)$ for $j=1,\ldots,n$. If an extra advertisement is broadcasted the probability to see exactly j advertisements from these $n+1$ broadcasts given by

$$f(j \mid n+1) = f(j \mid n)P(\text{the extra advertisement is not seen}) + f(j-1 \mid n)P(\text{the extra advertisement is seen})$$

This recursive equation can be evaluated in polynomial time, quadratic in the number of advertisements. The independence that is assumed by this "quadratic" method is not correct. However, for viewing probabilities of individuals this method may give satisfactory estimates. To estimate the contact frequencies for a target group one needs to combine the contact frequencies of a large number of representative individuals. Therefore the methods becomes rather slow. From our experience we notice that applying this method directly to aggregate viewing probabilities for target groups leads to really bad estimates of the contact frequency. To our knowledge no satisfactory estimation method for contact frequencies of target groups has been published in the literature.

In mathematical terms the contact frequency distribution for target groups can be defined as follows.
N is a set of *n* blocks that is selected by the planner from a set of options consisting of *K* blocks (*K* depends on the planning horizon). Let G_i be the event that block *i* is seen by the target group and let $\overline{G_i}$ be the event that block *i* is not seen, where the event of a target group seeing a block is represented by the viewing behavior of an average member of the target group. Denote the number of blocks that are seen by the target group by *X*. Then the contact frequency *f (j)*, the probability that the target group is confronted with exactly *j* blocks, is given by

$$f(j) = P(X = j) = \sum_{\substack{C \subseteq N \\ |C| = j}} P\Big(\big(\bigcap_{i \in C} G_i\big) \bigcap \big(\bigcap_{i \notin C} \overline{G_i}\big)\Big)$$

In the appendix we will prove the following theorem.

Theorem 1

The frequencies *f (j)*, for *j = 0,...,n*, can be calculated by the recursive relationship

$$f(j) = \sum_{\substack{C \subseteq N \\ |C| = j}} P\big(\bigcap_{i \in C} G_i\big) - \sum_{k=1}^{n-j} \binom{j+k}{j} f(j+k), \qquad \text{for } j = 1,\dots,n,$$

and

$$f(0) = 1 - \sum_{j=1}^{n} f(j).$$

ÿ

The probability to watch all blocks, that is $f(n) = P(\bigcap_{i \in N} G_i)$, has to be calculated directly from the viewing probabilities. With the functional relationship between the contact frequencies, all the other frequencies can be calculated backward (starting with *f (n),* going down to *f (1)* and *f (0)*). Notice that this recursive formula for the contact frequency does not require assumptions on the set of blocks, such as independence between the blocks.

Evaluation criteria for media schedules that have been used in other studies are reach, opportunity to see (OTS) and gross rating point (GRP). In mathematical terms we can express these quantities as a function of the contact frequencies:

- The reach is defined as the probability that the target group sees at least one block: $\text{reach} = P(X > 0) = 1 - f(0)$. With a slight abuse of the mathematical notation and assuming that the target group is sufficiently large, the frequency *f (j)* can be interpreted as the fraction of the target group that is exposed to the advertisement exactly *j* times.
- The OTS is the average number of blocks seen by the viewer, given that the viewer sees at least one block:

  $$\text{OTS} = E(X \mid X > 0) = \sum_{j=0}^{n} \frac{jP(X = j)}{P(X > 0)} = \frac{1}{1 - f(0)} \sum_{j=1}^{n} jf(j).$$
- The GRP is defined as the sum of the viewing probabilities for each of the blocks:

  $$\text{GRP} = \sum_{i \in N} P(G_i).$$

  The GRP can be calculated as the product of the OTS and the reach. The next theorem shows the relationship between GRP's, the contact frequencies and the viewing probabilities of the blocks.

THEOREM 2

$$\text{GRP} = \text{OTS} * \text{reach} = \sum_{j=1}^{n} jf(j).$$

PROOF

Using Theorem 1 we see that

$$f(1) = \sum_{i \in N} P(G_i) - \sum_{k=1}^{n-1} \binom{1+k}{1} f(1+k)$$

$$= \sum_{i \in N} P(G_i) - \sum_{k=1}^{n-1} (k+1) f(k+1).$$

Thus, by definition,

$$\text{OTS} * \text{reach} = \sum_{j=1}^{n} jf(j) = f(1) + \sum_{j=2}^{n} jf(j)$$

$$= \sum_{i \in N} P(G_i) - \sum_{k=1}^{n-1} (k+1) f(k+1) + \sum_{j=2}^{n} jf(j)$$

$$= \sum_{i \in N} P(G_i) = \text{GRP}.$$

ÿ

2.2. Estimating the contact frequency distribution

Part of the formula for the contact frequency is to calculate the probability that a person sees a given collection *C* consisting of *j* commercials from the set *N*

$P(\bigcap_{i \in C} G_i)$, for *CÍ N*, *|C|=j*.

In practice one could only hope for the exposure data of one block and of any combination of two blocks. To collect the exposure information of any combination of three or more blocks is too expensive and is hardly ever done. To assume independence between the probabilities to watch a commercial block is not correct, since it is more likely that two or more commercial breaks within the time frame of the same program are seen.

In our model we assume that the dependence between the viewing probabilities of the commercial breaks can be modeled according to the Markov principle from probability theory.

Assumption

We assume that the blocks can be ordered according to the time of showing.
The probability that a person sees a block only depends on the fact whether or not this person has seen the block previous to this one and is independent of earlier blocks.

From a practical viewpoint this assumption is not unrealistic, since we are primarily interested in the TV channel that a person is currently watching. We need to know only the last block that the person has seen before the current block. How long the person is watching this channel (i.e. which blocks this person has seen before the last block) or which channel this person was watching before is of lesser importance.

Let set C consist of the blocks $i_1,...,i_j$, ordered according to the showing time. The Markov assumption results into the following formula for the probability that every block in set C is seen by the target group,

$$P(\bigcap_{i\in C} G_i)) = \frac{\prod_{k=1}^{j-1} P(G_{i_k} \bigcap G_{i_{k+1}})}{\prod_{k=2}^{j-1} P(G_{i_k})}, \qquad \text{for } j=|C| \geq 3.$$

However, the calculation of the frequencies is still very time consuming since we need to calculate the above quantity for each subset C of N. There are $\sum_{j=1}^{n}\binom{n}{j}$ of these subsets for a fixed j which is exponential in the number of blocks with the commercial. Therefore, if the planner selects more than 20 commercials it becomes very time consuming to calculate the frequencies. If we replace the probabilities of a single block and of a combination of two blocks by their average probability then we can approximate the probability that all the blocks in C are seen with

$$P(\bigcap_{i\in C} G_i)) = \binom{n}{j}\frac{p_2^{j-1}}{p_1^{j-2}},$$

where

$$p_1 = \frac{1}{n}\sum_{i\in N} P(G_i) \quad \text{and} \quad p_2 = \frac{1}{\binom{n}{2}}\sum_{\substack{C\subseteq N\\ |C|=2,}} P(\bigcap_{i\in C} G_i).$$

The final estimation of the frequencies is thus given by

$$f(j) \approx \binom{n}{j}\frac{p_2^{j-1}}{p_1^{j-2}} - \sum_{k=1}^{n-j}\binom{j+k}{j} f(j+k).$$

The recursive relationship can then be used to approximate the values of *f (j)* in polynomial time.

The key issue is now the quality of the estimation. We have tested the model as follows. Since there is no data on the viewing probabilities for every possible block combination for a target group, we generated this data by simulation. First the probability to see a particular block or a combination of blocks is taken from a real data set for Dutch television supplied by Intomart Cko. The probabilities to see combinations of three or more blocks were drawn randomly from a set $[0,d]$ where d is the minimum of the probability to see a subset of the blocks in the combination. We generated these probabilities for a large number of distributions. Next, for a random selection of blocks, the contact frequency distribution was determined exactly. This was repeated for a large number of block selections. The resulting contact frequencies were then compared to the contact frequency estimation based on the Markov approximation and based on Markov approximation in combination with average probabilities.

In Tables 1 and 2 we show the average deviation for all the probability distributions with a set N consisting of three blocks and a set consisting of twenty blocks.

= Insert Table 1 =

= Insert Table 2 =

Hence, our tests indicate that our approximation is quite accurate.

2.3. Optimization

When the objective is (effective) reach we can define the media planning problem as follows. Recall that in our time frame there are in total K blocks to choose from. The size (denoted by n) of the selection is assumed to be fixed. With only a budget constraint (block i has price a_i, the budget is B) the problem can be modeled using the variables

$$x_i = \begin{cases} 1 & \text{if block } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}, \quad j = 1,2,\ldots K.$$

as follows

$$g = \max \sum_{j=l}^{u} f(j)$$

subject to

$$\sum_{i=1}^{K} a_i x_i \leq B$$

$$\sum_{i=1}^{K} x_i = n$$

$$f(j) = \binom{n}{j} \frac{p_2^{\;j-1}}{p_1^{\;j-2}} - \sum_{k=1}^{n-j} \binom{j+k}{j} f(j+k)$$

$$p_1 = \frac{1}{n} \sum_{i=1}^{K} P(G_i) x_i$$

$$p_2 = \frac{1}{\binom{n}{2}} \sum_{i,j} P(G_i \cap G_j) x_i x_j$$

$$x_i \in \{0,1\},\ \ i = 1,\ldots,K$$

The upperbound u and the lowerbound l on the reach can be chosen depending on the application. Note that this model has a non-linear objective function.

3. Optimization methods

Since the objective function is non-linear, we need solution methods that can handle such an objective function.

We will show that the optimization problem can be solved for large instances by local search methods. These methods are relatively simple optimization procedures that can be applied to a wide variety of problems. We will apply pure random search, steepest ascent, simulated annealing, taboo search, and two genetic algorithms to the media planning problem and compare these methods with respect to performance and speed.

3.1. Pure Random Search

The set S of feasible solutions is

$$S = \{x \in \{0,1\}^K : \sum_{i=1}^{K} a_i x_i \leq B, \sum_{i=1}^{K} x_i = n\}.$$

This set can be sampled directly by choosing a block selection and checking its feasibility. However many selections obtained this way will be infeasible. It is recommended to use a more advanced sampling by sub-sampling on price classes of blocks. For example, with a budget of Dfl. 300.000 and 20 blocks to select one can randomly select 2 blocks in the price range Dfl. 30.000 or higher, 10 blocks in the price range Dfl. 10.000 to Dfl.30.000 and 8 blocks cheaper than Dfl.10.000. The pure random search algorithm randomly selects solutions from the feasible set *S* and determines the function value of the solution. The best solution is returned after a fixed number of feasible selections.

3.2. Ascent methods

For every solution *x* in the feasible set *S*, represented as a 0-1 vector, define the neighborhood *U(x)* by all solutions *y* that differ from *x* by a two-exchange: $y_k=1-x_k$, $y_l=1-x_l$, for some *k* and *l* $y_i=x_i$ for all other *i*.

A first ascent method starts by selecting a current solution x_c and determining its neighborhood $U(x_c)$. The neighborhood is systematically searched for a better solution. The first better solution that is found becomes the new current solution and subsequently the neighborhood of this new solution is searched in a similar way. When no better solution can be found in the neighborhood the method terminates.

A steepest ascent method first searches for the best solution in the entire neighborhood $U(x_c)$. If this solution is better than the current solution, then this solution becomes the new current solution and its neighborhood is searched in a similar way. If no better solution can be found the algorithm terminates.

Ascent methods have successfully been applied to media planning (Little and Lodish (1969), Aaker (1975)).

3.3. Simulated Annealing

An alternative to the ascent algorithms is to allow selection of a worse solution from the neighborhood of the current solution with some positive probability. This probability is reduced during the course of the algorithm. Simulated annealing has its origin in condensed matter physics (annealing processes) where the energy of solid masses is minimized by quick heating followed by slow cooling. A simulation of the annealing process based on Monte Carlo techniques results in local search with a stochastic component. This approach was developed by Metropolis et al. (1953), and has become very popular since the 1980’s. A vast number of applications has shown that simulating annealing can find good solutions for optimization problems (Aarts et al. (1996)).

Simulated annealing can be seen as an ascent method with a different acceptation model for alternative solutions. Start with a high temperature that results in a high probability to accept a (better or worse) solution as the new current solution. Decrease the temperature (decrease the acceptance probability) after a fixed number of iterations. Continue this process until some stopping criterion is satisfied, for instance when the temperature drops below a predetermined value.

SIMULATED ANNEALING

Input: ***starting temperature t_0, temperature factor A,***
MAXITER, MAXSUCC and MAXNTEMP
(MAXSUCC is the maximum number of accepted solutions within one iteration
MAXITER is the maximum number of considered solutions within one iteration
MAXNTEMP reflects a stopping criterion for the temperature decrease process)

Step 0 ***Select $x_{start} \in S$, $x_{best}=x_{start}$, $x_c=x_{best}$. Temp:=t_0, $n_{Temp}=1$.***
(x_{start} is the starting solution
x_{best} is the best solution found throughout the algorithm
x_c is the current solution)

Step 1 i:=1, Succ:=0.

Step 2 Select y from the neighborhood $U(x_c)$ of x_c, i:=i+1.

Step 3 Accept solution y if $g(y) \geq g(x_c)$ or
accept solution y with probability $e^{[g(y)-g(x_c)]/Temp}$ if $g(y)<g(x_c)$:
x_c:=y, Succ:=Succ+1.
If $g(x_c)>g(x_{best})$ then $x_{best}:=x_c$.

Step 4 If i>MAXITER or Succ > MAXSUCC then Step 5 otherwise Step 2.

Step 5 If n_{Temp} < MAXNTEMP then
Temp:=Temp*A, $n_{Temp}:=n_{Temp}+1$, goto Step 1
else return x_{best} and $g(x_{best})$.

3.4. Taboo Search

Taboo search also differs from the ascent methods by the possibility to a worse alternative solution. The neighborhood of the current solution is completely searched and the best alternative solution *y* is identified. If *y* is better than *x* then the neighborhood of *y* is searched. If *y* is not better than *x* then *y* is still selected if solution *y* is not on a list of solutions that are (temporarily) taboo. The taboo list prevents the algorithm from cycling between two solutions. The use of some sort of memory to record taboo solutions is an essential feature of taboo search. Although the method has no solid theoretical founding explaining its good performance, it has shown remarkable effectiveness in practice (Glover and Laguna (1997)).

TABOO SEARCH

Input: Length of the Taboo list, MAXITER

Step 0 Select $x_{start} \in S$, $x_{best}=x_{start}$, $x_c=x_{best}$. ***Empty taboo list, i:=1.***

Step 1 Search in $U(x_c)$ ***for solution y such that***

$g(y) \geq g(u)$ ***for all u in*** $N(x_c)$.

Step 2 If $g(y) \geq g(x_c)$ ***then goto Step 3.***

Otherwise if y is not taboo or y is taboo but $g(y) > g(x_{best})$

then goto Step 3.

Otherwise find the second best solution in the neighborhood $U(x_c)$ ***and goto Step 2.***

Step 3 $x_c:=y$

If $g(x_c) > g(x_{best})$ ***then*** $x_{best}=x_c$.

Update Taboo list, i:=i+1.

Step 4 If $i< MAXITER$ ***then Step 1 else Return*** x_{best} ***and*** $g(x_{best})$.

3.5 Genetic algorithm

Evolutionary algorithms have been applied to optimization problems since the 1960s. Genetic algorithms are based on selection, mating and mutation. In the selection phase solutions are selected from a set (population) that have strong or preferred characteristics (such as high objective values). New solutions that hopefully combine the preferred characteristics are generated through crossover (mating). Mutation is applied to the new solutions to introduce new characteristics that are not present in the current population. Adding the new solutions and deleting the weaker solutions generates the new solution set (generation). Genetic algorithms have been successfully applied to a wide range of optimization problems and are presently very popular (Goldberg (1989)).

Here we present two versions of genetic algorithms. An 0-1 vector can represent a solution, where 1 on position *j* reflects that block *j* is selected and 0 otherwise. In computer implementations more advanced datastructures should be used to prevent extensive memory use.

GENETIC ALGORITHM (1)

Input: ***Starting set of solutions P, MAXITER, number of pairs k, mutation probability p.***

Step 1 Draw with replacement 2k solutions from the set P, where each solution x has probability $\frac{g(x)}{\sum_{y \in S} g(y)}$***. Divide the solutions randomly into k pairs.***

Step 2 In each pair randomly two crossover points are selected as the indices of the 0-1 vectors representing the solutions. The part after the crossover index in the first solution is interchanged with part after the crossover index of the second solution.

Step 3 Both new solutions are mutated by randomly selecting indices. The value in the solution at this index is changed from 0 to 1 or from 1 to 0 with some (small) probability p.

Step 4 For non-feasible solutions randomly remove blocks until the solution is feasible.

Step 5 The new solutions are added to the set P and the worst k solutions are removed.

This process is repeated a number of times (MAXITER), the best solution found is returned.

An alternative genetic algorithm combines the crossover and mutation phase and uses a different selection criterion.

GENETIC ALGORITHM (2)

Input: ***Starting set of solutions P, MAXITER, percentages a and b, selection probabilities*** p_b, p_n, ***and*** p_o.

Step 1 Order the solutions from the set P according to non-decreasing objective function value. Select best a% and the worst b% of the solutions and randomly divide them into k pairs.

Step 2 For each pair create one new solution. If a block is (not) selected in both the solutions in the pair then the block is (not) selected in the new solution with probability p_b (p_n)***.***
If the block is selected in only one of the solutions then the block is selected in the new solution with probability p_o***.***

Step 3 For non-feasible solutions randomly remove blocks until the solution is feasible
Step 4 Add the new solutions to the set P and remove the k worst solutions.
This process is repeated a number of times (MAXITER), the best solution found is returned.

The size of a solution is assumed to be fixed for the other search methods, but for the genetic algorithms we only start with solutions of predetermined size. In the crossover and mutation phase the solution size can be changed. Therefore the genetic algorithm combines two aspects of the optimization process that can increase the effective reach: the solution size and the block selection. The other local search methods optimize only the block selection and need to be applied to a number of different solution sizes. We purposely split the optimization of solution size and block selection for the other algorithms. Without this distinction it is hard to generate feasible solutions in the neighborhood of the current solution and much time is spent generating and rejecting alternative solutions.

The genetic algorithm has a longer running time, but since the other methods require multiple runs with different solution sizes the running times are comparable.

4. Test Results

The algorithms were compared using a real data test set. The test sets of 100 blocks consisted of real life data for Dutch television commercials in 1995 and was supplied by Intomart Cko, The Netherlands via Point Logic Systems, The Netherlands. The test set consisted of a specification of the probabilities to see a commercial block and of the probabilities to see a combination of two commercial blocks.

As the objective function of the optimization problem we used effective reach between 4 and 8 blocks (l = 4, u = 8). The algorithms were applied to the optimization problem, where the number of selected blocks was set between 25 and 45 blocks. These choices were based on pre-testing that showed that the optimum number of blocks would be between 30 and 40 for the test sets. The genetic algorithms were tested on a start set S with 30 blocks. Each algorithm was applied 5 times to the test set. Steepest ascent has no stochastic component and was applied only once. The best performance was achieved with the following parameter values.

Algorithm	Parameter values
Pure Random Search	stop after generating 100 feasible solutions
Simulated Annealing	t_0=10, A=0.95, MAXNTEMP=0.005, MAXSUCC=100, MAXITER=1000
Taboo Search	Length taboo list = 5, MAXITER=1000
Genetic Algorithm (1)	$\|S\|$=250, MAXITER=20, p=0.0001
Genetic Algorithm (2)	$\|S\|$=250, MAXITER=20, p_b=p_n=0.95, p_o=0.5, a=25, b=5

These parameters resulted in a running time of approximately 15 minutes for all the algorithms. The effect of each algorithm was recorded as the mean function value of the 5 times that the algorithm is applied to the test problem.

= Insert Figure 1 =

Figure 1 shows that taboo search produces selections with the highest effective reach for every size of the selection. For the solution sizes between 25 and 45 blocks the effect of an advertisement campaign will increase as the number of selected blocks increases, but the increase in the effect will become smaller as the number of selected blocks increases. Of course the budget constraint will limit the size of the block selection. From Figure 1 we see that with 30 selected blocks an effect of at most 0.75 can be reached, whereas 45 blocks can result in an effect of 0.8. In a separate test we extended the taboo search procedure to 60 selected blocks, but the increase in effect was negligible. For the test set the variance remains approximately the same for all block selections, but is reasonably small. The simulated annealing algorithm has a comparable behavior to steepest ascent for larger selection sizes.
Pure random search can not match the performance of the other algorithms.

= Insert Figure 2 =

= Insert Figure 3 =

The genetic algorithms show that an effective reach of 0.75 can be reached within 10 (iterations) generations. The last 10 iterations cannot significally improve the objective value. Taboo searched can generate an effective reach of 0.80, but needs to find the optimal size of block selection through enumeration. The second version of the genetic algorithm quickly adds solutions with on average 37 blocks where the first version slowly moves to solutions with more blocks. Hence the first version algorithm replaces solutions with fewer blocks with solutions with more blocks without a significant improvement of the objective function value. The genetic algorithms also show that the optimal number of blocks is between 30 and 40 blocks.

5. Optimal block selection in practice

The optimization problem defined in Section 2.3 seems somewhat arbitrarily chosen. We used this model as the simplest problem that can be defined that still resembles the block selection process. In practice, this model is not so realistic. One needs to add other characteristics of the planning process to obtain practical block selections. However, most of these characteristics can easily be incorporated into the basic model. We will distinguish between three dimensions: multiple goals, dynamic planning and constrained planning.

5.1. Multiple goals

Besides the generated effective reach other goals can be incorporated into the model such as the reach of the solution and the cheapest solution. It seems wise to give the planner control over a trade-off between effective schedules, cheap schedules and schedules with a larger reach. An objective function can be formulated as a weighted sum of multiple goals. The planner has control over the weights (together with the size of the selection). A typical objective function then looks like

$$\max\left[\mathrm{w}_1 \sum_{j=4}^{8} f(j) + w_2 \sum_{j\geq 1} f(j) + w_3 \sum_{i=1}^{K} a_i x_i \right].$$

The local search methods can easily be adjusted for these objective functions.

5.2. Dynamic planning

In practice, there are a number of obstacles in the execution of a media plan. One such complication is that, although the planner has chosen a specific set of blocks in which the commercial should be shown, some blocks may longer be available at the time of purchase. Usually the planner gives an order for a specific set of blocks at a certain point in time, and after some time the planners is notified which subset of the selected blocks is accepted and what blocks are unavailable. The planner is then left with a different optimization problem, where some blocks are already pre-assigned and some blocks are unavailable.

The local search algorithms can deal with this optimization problem by adjusting the feasible set,

$$\{x \in \{0,1\}^K : x_i = 1 \text{ for } i \in I_1, x_i = 0 \text{ for } i \in I_2, \sum_{i=1}^{K} x_i = n, \sum_{i=1}^{K} a_i x_i \le B\},$$

where I_1 is the set of pre-assigned blocks and I_2 is the set of unavailable blocks. The neighborhood of a solution is defined as all alternative solutions that result from x through a two-exchange for elements in S but not in set $I_1 \bigcup I_2$. All local search algorithms can be easily adjusted for this situation.

5.3. Constrained Optimization

In practice, there can be a number of other constraints that are specific for the planning situation. Two examples are

1. There can be a customer preference for some blocks, other than price or effect. This preference could relate to channel preference, time preference or contracts with media

$$\sum_{i \in I_s} x_i \ge s.$$

agencies. It can be modeled as a selection of a minimum number of blocks from a specific set I_s

2. Often the advertising company asks for a specific number of GRP. The planner is asked to buy blocks until a pre-defined number of GRP is attained. With the relationship between GRP and contact frequencies from Theorem 2, this constraint can easily be added to the basic problem. Even though the restriction is non-linear in x_i, once the contact frequencies are calculated, the restriction is linear in $f(j)$.

For both constraints, the feasible set in the local search methods is further restricted. The methods easily apply to the extensions.

6. Concluding remarks

In this paper we have studied a media planning model with the objective to optimize the contact frequencies themselves instead of related measures such as reach. The mathematical formulation of the contact frequency is tested on real data combined with randomly generated data. The model for the contact frequency can reflect the true contact frequencies within a small error range.

The optimization of the media planning model uses local search techniques. It is shown that these techniques are flexible and are well suited for this non-linear media planning model. To be useful in practice, the media planning model from this paper should be extended to include other objectives and more restrictions. The local search techniques seem to be flexible enough to include a number of commonly used extensions. Further improvements to our approach can be found in hybrid algorithms. Then a genetic algorithm can be combined with another local search technique such as taboo search by applying taboo search to the best solutions in a given generation of the genetic algorithm.

The mathematical formulation of contact frequencies relies on good estimates of future viewing behavior, modeled as the probability that a target group will be exposed to specific commercial blocks. The recent interest in household level exposure models (Abe (1997)) are of importance for media selection models, and shows a need for a new generation of selection models. Our model needs only the probabilities to see one block or a combination of two blocks. It can also be applied on a household level. Previous research (Headen et al. (1977), Rust and Klompmaker (1981), Leckenby and Kishi (1982)) showed that the beta binomial model also performs well in predicting future viewing behavior on an aggregate level even though it assumes independence between the commercial blocks that is not correct. Still new exposure models are needed, both on an aggregate and household level, to assure a better performance of the media planning models.

Literature

Abe, M. (1997), “A Household-Level Television Advertising Exposure Model”, *Journal of Marketing Research*, 34, 394-405.

Aaker, D. (1975), “ADMOD: An Advertising Decision Model”, *Journal of Marketing Research*, 12, 37-45.

E.L. Aarts and Lenstra J.K. (1996), *Local Search in Combinatorial Optimization*, Wiley, NY.

Glover, F. and Laguna M. (1997), *Tabu Search*, Kluwer Academic Pub., Boston.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.

Got Plenty, Special issue of *Journal of Adverting Research* (1997), issue April.

Headen, R.S., Klompmaker, J.E. and Teel Jr, J.E. (1977), “Predicting Audience Exposure to Spot TV Advertising Schedules”, *Journal of Marketing Research*, 14, 1-9.

Leckenby, J.D., and Ju, K.H. (1989), “Advances in Media Decision Models”, *Current Issues and Research in Advertising*, 2, 311-357.

Leckenby, J.D. and Kishi, S. (1982), “Performance of Four Exposure Distribution Models”, *Journal of Advertising Research*, 22, 35-44.

Little, J.D.C. (1979), “Aggregate Advertising Models: The State of the Art”, *Operations Research*, 27, 629-667.

Little, J.D.C., and Lodish, L.M. (1969), “A Media Planning Calculus”, *Operations Research*, 17, 1-35.

Metheringham, R.A. (1964), “Measuring the Net Cumulative Coverage of a Print Campaign”, *Journal of Advertising Research*, 4, 23-28.

Metropolis, N., Rosenbluth A., Rosenbluth M., Teller A. and Teller E. (1953), “Equation of State Calculations by Fast Computing Machines”, *Journal of Chemical Physics*, 21, 1087-1092.

Naples, M.J. (1979), *“Effective Frequency: The Relationship between Frequency and Advertising Effectiveness”*, New York: Association of National Advertisers.

Rathnam, S., Arun, M.R., Chaudhury, A. and Shukla, P.R. (1992). “MUDRAPLAN-A DSS for Media Planning: From Design to Utilization”, *Interfaces*, 22, 65-75.

Rust, R.T. (1985). “Selecting Network Television Advertising Schedules”, *Journal of Business Research*, 13, 483-94.

Rust, R.T. and Klompmaker, J.E. (1981), “Improving the Estimation Procedure for the Beta Binomial TV Exposure Model”, *Journal of Marketing Research*, 18, 442-448.

APPENDIX

Theorem 1

The frequencies *f (j)*, for $j = 0,\dots, n$, can be calculated by the recursive relationship

$$f(j) = \sum_{\substack{C\subseteq N\\ |C|=j}} P(\bigcap_{i\in C} G_i) - \sum_{k=1}^{n-j} \binom{j+k}{j} f(j+k) \qquad \text{for } j=1,\dots,n$$

and

$$f(0) = 1 - \sum_{j=1}^{n} f(j).$$

Proof

Let $N=\{i_1,i_2,\dots,i_n\}$ with i_1 £ i_2 £ i_3 £.. £ i_n. Consider the relationship between *f (n)* and *f (n-1)*. If *C* is a set consisting of *n-1* blocks, then there is only one block in *N* that is not in the set *C*, say block *b*. We have

$$P(\bigcap_{i\in C} G_i) = P[(\bigcap_{i\in C} G_i)\bigcap \overline{G}_b] + P[(\bigcap_{i\in C\bigcup\{b\}} G_i)]$$

$$= P[(\bigcap_{i\in C} G_i)\bigcap \overline{G}_b] + P[\bigcap_{i\in N} G_i].$$

Thus the summation over all subsets of *N* with *n-1* elements yields:

$$\sum_{\substack{C\subseteq N\\ |C|=n-1}} P(\bigcap_{i\in C} G_i) = \sum_{\substack{C\subseteq N\\ |C|=n-1}} \left(P[(\bigcap_{i\in C} G_i)\bigcap \overline{G}_b] + P[\bigcap_{i\in N} G_i] \right) = f(n-1) + nf(n).$$

We see the required relationship for *j=n-1*.

If the set *C* consists of *j* elements then the relationship becomes

$$\sum_{\substack{C\subseteq N\\ |C|=j}} P(\bigcap_{i\in C} G_i) = \sum_{\substack{C\subseteq N\\ |C|=j}} P\left[\left(\bigcap_{i\in C} G_i\right)\bigcap\left(\bigcap_{i\notin C} \overline{G}_i\right)\right] + \sum_{k=1}^{N-j} \sum_{\substack{C\subset N\\ |C|=j}} \sum_{m=1}^{k} \sum_{\substack{l_m=1\\ l_m\notin C}}^{n} P\left[\left(\bigcap_{i\in C\bigcup\{l_1,\dots,l_k\}} G_i\right)\bigcap\left(\bigcap_{i\notin C\bigcup\{l_1,\dots,l_k\}} G_i\right)\right]$$

$$= f(j) + \sum_{k=1}^{n-j} \binom{j+k}{j} f(j+k),$$

because there are $\binom{n}{j}\binom{n-j}{k}$ terms, but only $\binom{n}{j+k}$ different combinations of *j+k* elements from the set *N*. Hence each *f(j+k)* is counted $\binom{j+k}{j}$ times.