

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/235305426>

# Solving Marketing Optimization Problems Using Genetic Algorithms

**Article** in *European Journal of Marketing* · April 1995

DOI: 10.1108/03090569510086648

---

CITATIONS

41

---

READS

305

**3 authors**, including:



**Luiz Moutinho**

University of Suffolk

**261** PUBLICATIONS **3,961** CITATIONS

SEE PROFILE

**Some of the authors of this publication are also working on these related projects:**



senior tourism [View project](#)

---

# Solving marketing optimization problems using genetic algorithms

Marketing  
optimization  
problems

39

S. Hurley

*Department of Computing Mathematics, University of Wales, Cardiff*

L. Moutinho

*Cardiff Business School, University of Wales, Cardiff, and*

N.M. Stephens

*Department of Computing Mathematics, University of Wales, Cardiff*

---

Received March 1994  
Revised August 1994  
December 1994

## Genetic algorithms

Genetic algorithms were invented by Holland[1] to mimic some of the processes of natural evolution and selection. In nature, each species needs to adapt to a complicated and changing environment in order to maximize the likelihood of its survival. The knowledge that each species gains is encoded in its chromosomes which undergo transformations when reproduction occurs. Over a period of time, these changes to the chromosomes give rise to species which are more likely to survive, and so have a greater chance of passing their improved characteristics on to future generations. Of course, not all changes will be beneficial but those which are not tend to die out.

Holland's genetic algorithm attempts to simulate nature's genetic algorithm in the following manner. The first step is to represent a legal solution to the problem you are solving by a string of genes which can take on some value from a specified finite range or alphabet. This string of genes, which represents a solution, is known as a chromosome. Then an initial population of legal chromosomes is constructed at random. At each generation, the fitness of each chromosome in the population is measured (a high fitness value would indicate a better solution than a low fitness value). The fitter chromosomes are then selected to produce offspring for the next generation, which inherit the best characteristics of both the parents. After many generations of selection for the fitter chromosomes, the result is hopefully a population which is substantially fitter than the original. The theoretical basis for the genetic algorithm is the Schemata Theorem[1], which states that the number of individual chromosomes with good, short, low-order schemata or building-blocks (i.e. beneficial parts of the chromosome) receive an exponentially increasing number of trials in successive generations.

All genetic algorithms consist of the following main components[2-4]:

- (1) *Chromosomal representation.* Each chromosome represents a legal solution to the problem and is composed of a string of genes. The binary

*European Journal of Marketing,*  
Vol. 29 No. 4, 1995, pp. 39-56.  
© MCB University Press, 0309-0566

alphabet  $\{0, 1\}$  is often used to represent these genes but sometimes, depending on the application, integers or real numbers are used. In fact, almost any representation, which enables a solution to be encoded as a finite length string, can be used.

- (2) *Initial population.* Once a suitable representation has been decided on for the chromosomes, it is necessary to create an initial population to serve as the starting-point for the genetic algorithm. This initial population can be created randomly or by using specialized, problem-specific information. From empirical studies, over a wide range of function optimization problems, a population size of between 30 and 100 is usually recommended.
- (3) *Fitness evaluation.* Fitness evaluation involves defining an objective or fitness function against which each chromosome is tested for suitability for the environment under consideration. As the algorithm proceeds we would expect the individual fitness of the "best" chromosome to increase as well as the total fitness of the population as a whole.
- (4) *Selection.* We need to select chromosomes from the current population for reproduction. If we have a population of size  $2n$ , where  $n$  is some positive integer value, the selection procedure picks out two parent chromosomes, based on their fitness values, which are then used by the cross-over and mutation operators (described below) to produce two offspring for the new population. This selection/cross-over/mutation cycle is repeated until the new population contains  $2n$  chromosomes, i.e. after  $n$  cycles. The higher the fitness value, the higher the probability of that chromosome being selected for reproduction.
- (5) *Cross-over and mutation.* Once a pair of chromosomes has been selected, cross-over can take place to produce offspring. A cross-over probability of 1.0 indicates that all the selected chromosomes are used in reproduction, i.e. there are no survivors. However, empirical studies[5] have shown that better results are achieved by a cross-over probability of between 0.65 and 0.85, which implies that the probability of a selected chromosome surviving to the next generation unchanged (apart from any changes arising from mutation) ranges from 0.35 ( $1 - 0.65$ ) to 0.15 ( $1 - 0.85$ ). One point cross-over involves taking the two selected parents and crossing them at a randomly chosen point, e.g. if we have the following parents (with  $x$  and  $y$  representing individual values within the chromosome):

parent 1:  $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$

parent 2:  $y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8$

and we cross at point 3, say, the following offspring would be produced

child 1:  $x_1, x_2, x_3, y_4, y_5, y_6, y_7, y_8$

child 2:  $y_1, y_2, y_3, x_4, x_5, x_6, x_7, x_8$ .

If we only use the cross-over operator to produce offspring, one potential problem which may arise is that, if all the chromosomes in the initial population have the same value at a particular position, then all future offspring will have this same value at this position. For example, if all the chromosomes have a 0 in position two, then all future offspring will have a 0 at position two. To combat this undesirable situation a mutation operator is used. This attempts to introduce some random alteration of the genes, e.g. 0 becomes 1 and vice versa. Typically this occurs infrequently, so mutation is of the order of about one bit changed in a thousand tested. Each bit in each chromosome is checked for possible mutation by generating a random number between zero and one and if this number is less than or equal to the given mutation probability, e.g. 0.001, then the bit value is changed.

This completes one cycle of the simple genetic algorithm. The fitness of each chromosome in the new population is evaluated and the whole procedure repeated, i.e.:

- (1) Generate random population
- (2) Repeat
  - evaluate fitness of current population
  - select chromosomes, based on fitness, for reproduction
  - perform cross-over and mutation to give new improved populationuntil *finished*

where *finished* indicates either an optimal or suitable suboptimal has been found or the maximum number of generations has been exceeded.

### **Evolutionary algorithms in management and marketing**

Nissen[6] has assessed the range of applications of evolutionary algorithms (EA) in management science. To date, the quantity and diversity of applications are still very moderate. Combinatorial optimization with a focus on job-shop scheduling is dominant. Most applications appear in an industrial setting with emphasis on production. Evolutionary programming (EP) and genetic programming (GP) do not seem to have been applied in management. We have not seen the big breakthrough of EA in management yet. Interest in these new techniques, however, has risen considerably over the last three years and will lead to a further increase in practical applications in the near future. In terms of research in management science, the dominant technique is genetic algorithms (GAs) among all the EA-variants. Nevertheless, the considerable versatility of EA has so far not been appreciated adequately in the domain of management science. Most researchers have been attracted by GAs, while evolution strategies (ES), EP and GP are not so widely known. GP is the newest technique and has not reached the level of practical applicability yet. EP and ES are very

similar, so that they could actually compete on the same types of application. The dominance of GAs should not be interpreted as superiority of this approach. It rather seems to be a good “infrastructure” that contributes to the trend for GAs.

Genetic algorithms have been the focus of studies in many areas of management. Table I outlines several of these areas and gives the associated literature references.

| Subject area  | Reference   |
|---|---|
| Financial services:   |   |
| risk management, trading strategy search,<br>credit scoring, portfolio optimization | Noble[7]  |
| stock trading   | Margarita[8,9], Burke[10], Johnstone[11],<br>Star[12], Schwartz[13] |
| speculative stock markets   | Nottola <i>et al.</i> [14]  |
| Facility layout and location planning   | Khuri and Batarekh[15]; Tam[16]<br>Huntley and Brown[17]            |
| Distribution  | Hughes[18]  |
| Vehicle routeing  | Baker and Schaffer[19];<br>Thangiah <i>et al.</i> [20]              |
| Transport problems  | Michalewicz[21]   |
| Forecasting company profit  | Thompson and Thompson[22]   |
| Organization  | Bruderer[23]  |
| Resource allocation   | Bean[24]  |
| Manufacturing:  |   |
| process control, modelling and forecasting  | Vancza and Markus[25]; Stevens[26];<br>Venugopal and Narendran[27]  |
| production planning   | Ward <i>et al.</i> [28]   |
| intelligent manufacturing   | Kulkarni and Parasei[29], Ward <i>et al.</i> [30]                   |
| packing problems  | Dowsland and Dowsland[31]   |
| job shop scheduling   | Biegel and Davern[32]   |
| Document clustering   | Gordon[33]  |
| Development of world economic model   | Johnstone[34]   |

**Table I.**  
Applications of genetic  
algorithms

---

Margarita[8] developed an application of neural networks and genetic algorithms to the analysis of the stock market dynamics. The author tried to find an optimal strategy for the stock trader network using genetic algorithms.

Evolutionary algorithms have also been hybridized with knowledge-based systems and neural networks for management applications. Evolution strategies have been applied in the area of health care – for scheduling patients in a hospital and allocating investments to various health-service programmes, as well as in strategic management and controlling – in the top-down calculation of budget models and general resource allocation problems.

Classifier systems have been applied to solve flow-shop scheduling[35]; in research and development – in an experimental study of product life cycles[36] and for developing learning models of consumer choice[37]; organization – in the study of the relationship between organizational structure of the firm and its capability to learn and adapt to changes[38]; and in financial services – for modelling trading behaviour in financial markets.

A great deal of information can be gained from the application of evolutionary biology to marketing. To this end, past research has demonstrated that the product evolutionary cycle (PEC) may be more helpful in describing the nature of competitive product-market development than the product life cycle (PLC). Tang and Holak[39] developed a conceptional framework linking marketing concepts to Darwin's evolution mechanism, specifically, a product species' sales growth (multiplication), core technology (inheritance), product extensions (variations), and failure or success (natural selection), by extending the previous PEC framework to a general differential equations system. A family of curved paths is then derived from the proposed dynamical system to explain how proliferation patterns of product species in divergence (cladogenesis), development (anagenesis), differentiation (adaptive radiation), mature (stasigenesis) and demise (extinction) emerge. Several managerial implications of the model can be drawn in terms of product-market development strategy. To our knowledge, this was the first work attempting explicitly to model the PEC framework mathematically.

A limitation of conjoint analysis is that in more realistic situations complete enumeration becomes infeasible and this has led to the specification of heuristics for identifying "good" products. Towards this end, Balakrishman and Jacob[40] investigated the use of genetic algorithms as a more natural way to address the problem of natural selection of products in the marketplace which results in only the "fittest surviving". These authors have also discussed the efficiency of various cross-over operators and compared their results with previously employed techniques such as Lagrangian relaxation.

### **Potential applications**

The two worked examples of application of GAs in marketing deal with, first, site location analysis and, second, optimization of product-market structures. Nevertheless, the range of potential applications of genetic algorithms in

marketing management is quite considerable as can be seen by the following list:

- (1) Consumer behaviour:
  - Learning models of consumer choice.
  - Schema theory – evoked set.
  - Consumer information processing.
  - Reproductive product trials.
  - Reference group influence, word-of-mouth networking and complaint behaviour.
- (2) Segmentation, targeting and positioning:
  - Optimization of product-market structures.
  - Family life cycle segmentation.
  - Analysis of key buying factors.
  - Product positioning.
- (3) Managing the marketing mix:
  - Optimization of the product life cycle.
  - Product design.
  - New product development – concept testing and combinatorial optimization.
  - Price optimization strategies.
  - Advertising strategy and media planning.
  - Sales management – customer profitability analysis, sales territory design and optimal sales force routing.
  - Distribution management – site location analysis, optimization of distribution channels and physical distribution strategies.
- (4) Strategic marketing:
  - Analysis of product parity, market parity and competitive advantages.
  - Competitive reaction strategies.
  - Optimization of marketing strategies.
  - Marketing effectiveness measurement.

#### **Example one: site location analysis**

Several approaches have been proposed in the market literature for the location of retail stores[41-49].

Breheny[47] has identified two main types of location techniques:

- (1) store-turnover forecasting techniques which are most useful where retail performance is a function of the location of outlets or where the market is outlet-dominated; and

- 
- (2) spatial marketing techniques which are most useful where the performance of an outlet is more dependent on being in the right general location, with the right socio-economic structure, than on local catchment area characteristics.

More “conventional” models are based generally on multivariate statistical analysis which in turn are founded on numeric or algebraic relationships between relevant variables. For example, Penny and Broom[50] developed a model in which store turnover is dependent functionally on variables such as household expenditure and population.

45

---

Simple plotting of stores in a specific area, along with competitors, might give a retailer a clearer idea of how the market is divided spatially and how the presence or absence of competitors affects the new outlet’s performance. This might, in turn, help identify “ideal market areas”[51] and, hence, gaps in the existing spatial market.

Disaggregate data on consumer expenditure patterns, together with data on the number, size and location of competing stores, would provide the basis for assessing the potential impacts of future alterations to the product mix and space allocation within existing stores or of the potential rationalization or expansion of the number of stores in a network[52]. A new generation of store location models, such as SLAM, has appeared, which facilitate comprehensive evaluation of both present and prospective retail outlets[48,49].

Of the various approaches proposed, the analogue methods and multiplicative competitive interactive (MCI) models seem to be the most popular. It has been shown by Nakanishi and Cooper[53] that the MCI model can be calibrated using least squares procedures. Furthermore, a generalized formulation to incorporate zero-one-type store attributes in the MCI model has been developed by Mahajan *et al.*[54].

In the competitive retailing environment, a multi-unit company may find that individual optimum locations may not result in an overall optimum network of stores in a market area. Achabal *et al.*[55] developed a multiple store location (MULTILOC) model which extends the multiplicative competitive interactive (MCI) model to the multistore location problem, utilizing a random search procedure combined with an interchange heuristic to identify optimal or near-optimal sets of locations.

One of the popular heuristics used to solve combinatorial problems such as the MULTILOC model is the interchange heuristic described by Teitz and Bart[56]. Similar implications emerge from a successful case study carried out by Rosenfield[57] who has also developed a formal optimizing model using mathematical programming techniques. Hogue and Fields[58] report on a proprietary decision support system which can optimize profits, sales or customer traffic.



### **Application of genetic algorithms to site location**

In the situation where a large number of sites already exist, say 50 or more, and there is a choice of several proposed sites, the problem of finding the optimal network is difficult. However, genetic algorithms can be used to solve various formulations of the site location problem. In particular, given an existing network of outlets and a set of possible new sites, a genetic algorithm can be used to solve the following:

- (1) Find and identify how many of the proposed sites to use, to augment and improve the existing network in terms of profitability, attractiveness to potential customers or any other suitable metric. The existing network is assumed to be static, i.e. remains unaltered, with no closure of sites.
- (2) As problem (1), but here the existing network is assumed to be dynamic, i.e. existing sites may be removed from the network if this gives an improvement in overall network performance.
- (3) Given an existing network of outlets, determine the best subset of these to market a particular product or service.

To illustrate the use of genetic algorithms in this area consider problem (2).

#### *Chromosome representation*

Each chromosome represents a possible network consisting of existing sites and possible new sites. If there are  $S_e$  existing sites and  $S_p$  possible new sites, the length of the chromosome will be  $S_e + S_p$ . The individual genes within each chromosome are represented by a binary alphabet; 0 indicates that a particular site is not used in the network, whereas a 1 would indicate that a site is used. The position within the chromosome is important as this indicates the site under consideration, e.g. bit position 4 (from left to right) represents site 4. To illustrate, if we have four existing outlets and three possible new sites, then the chromosome

1 1 1 1 0 1 0

represents a network where sites 1, 2, 3 and 4 are used (the existing network) and site 6 is used; sites 5 and 7 remain unused.

#### *Fitness function*

The most important problem-dependent aspect in applying genetic algorithms is finding a suitable function to determine the fitness or goodness of a chromosome in the genetic population. One possible basis for the fitness function involves a slight modification of the Penny and Broom spatial interaction model[50]:

$$\text{EXP}_{ij} = \beta_0 \left( \sum_{k=1}^{N_k} E_k H_{ik} \right) \cdot \frac{W_j e^{-\beta_1 T_{ij}}}{\sum_{m=1: m \neq j}^{N_s + N_c} W_m e^{-\beta_1 T_{im}}} \quad (1)$$

where;

$N_k$  is the number of household categories;  $N_s = S_e + S_p$ ;  $N_c$  is the number of competitors;  $i = 1, 2, \dots, N_z$  ( $N_z$  is the number of zones) and  $j = 1, 2, \dots, N_s$ ;  $\text{EXP}_{ij}$  is the expenditure from zone  $i$  to site  $j$ . Note that only sites in the chromosome under consideration set to a 1 are used in equation (1); i.e. site  $j$  is used only if position  $j$  in the chromosome is set to 1; similarly, in the denominator term, site  $m$  is used only if position  $m$  in the chromosome is set to 1 ( $m = 1, \dots, N_s$ ). The competitors  $m = N_s + 1, \dots, N_s + N_c$  are always included.

$\beta_0$  and  $\beta_1$  are parameters;

$E_k$  is the mean expenditure on the product/service by household category  $k$ ;

$H_{ik}$  is the number of households of category  $k$  located in zone  $i$ ;

$W_a$  is a subjective measure of attraction of existing/proposed sites ( $a = 1, 2, \dots, N_s$ ) and competitors ( $a = N_s + 1, \dots, N_s + N_c$ ).

$T_{ia}$  is the travel time from zone  $i$  to the existing/proposed sites ( $a = 1, 2, \dots, N_s$ ) and competitors ( $a = N_s + 1, \dots, N_s + N_c$ ).

The actual fitness function is then

$$\sum_i \sum_j \text{EXP}_{ij}. \quad (2)$$

The genetic algorithm will then attempt to maximize this expression, i.e. find a network of sites which maximizes customer expenditure.

#### Genetic operators

Each chromosome is represented by a binary string with an associated fitness. If two chromosomes are selected for reproduction by the selection procedure, the standard cross-over operators[4] can be used to generate offspring. For example, if the two chromosomes:

1 1 1 1 0 1 0 (sites 5 and 7 unused)  
1 0 1 1 0 0 1 (sites 2, 5 and 6 unused)

are selected for reproduction and one-point cross-over is used (e.g. at position 3) then the following offspring are produced:

1 1 1 1 0 0 1 (sites 5 and 6 unused)  
1 0 1 1 0 1 0 (sites 2, 5 and 7 unused).

Similarly, the standard mutation operator can be used.

### Simple illustration

To illustrate the application of genetic algorithms to the area of site location analysis consider the case where there are four existing sites ( $S_e = 4$ ) and three possible new sites ( $S_p = 3$ ), i.e.  $N_s = 7$ . In addition, assume there are three zones ( $i = 1, 2, 3$ ), four household categories ( $k = 1, 2, 3, 4$ ) and five competitor stores ( $N_c = 5$ ).

A population of five chromosomes will be considered,

$C_1$  1 0 1 0 0 1 1  
 $C_2$  0 0 1 1 1 1 0  
 $C_3$  1 0 1 0 0 0 1  
 $C_4$  1 1 1 1 0 0 0  
 $C_5$  1 1 1 0 1 0 1.

Fitness values are assigned to each of these chromosomes using the fitness function (2) which uses the variable settings given in Tables II and III. The attractiveness measure in Table III was taken to be a value on a scale of 1 (least attractive) to 100 (most attractive). The parameters  $\beta_1$  and  $\beta_2$  were set to 1.0 and 1/60 respectively.

Equation (1) can then be used to calculate the expenditure,  $EXP_{ij}$ , from zone  $i$  to site  $j$  for  $i = 1, 2, 3$  and  $j = 1, 2, \dots, 7$ . As an illustration, Table IV contains the

**Table II.**  
Fitness values of mean expenditure and number of households for each category

|                      | Household category ( $k$ ) |     |       |     |
|----------------------|----------------------------|-----|-------|-----|
|                      | 1                          | 2   | 3     | 4   |
| Mean expenditure     | 15                         | 25  | 10    | 50  |
| Number of households |                            |     |       |     |
| Zone 1               | 487                        | 181 | 451   | 402 |
| Zone 2               | 1,515                      | 797 | 1,017 | 473 |
| Zone 3               | 2,438                      | 448 | 531   | 60  |

**Table III.**  
Attractiveness measure of existing/potential sites and competitors, and travel times from the zones to the sites

|                                     | Existing/potential sites |    |    |    |    |    |    | Competitors |    |    |    |    |
|-------------------------------------|--------------------------|----|----|----|----|----|----|-------------|----|----|----|----|
|                                     | 1                        | 2  | 3  | 4  | 5  | 6  | 7  | 8           | 9  | 10 | 11 | 12 |
| Attractiveness measure, $W_a$       | 98                       | 37 | 91 | 81 | 66 | 1  | 42 | 10          | 16 | 15 | 49 | 77 |
| Travel time (in minutes) from zones |                          |    |    |    |    |    |    |             |    |    |    |    |
| Zone 1                              | 42                       | 76 | 52 | 40 | 64 | 90 | 8  | 43          | 80 | 47 | 5  | 60 |
| Zone 2                              | 24                       | 59 | 9  | 48 | 66 | 20 | 50 | 27          | 24 | 13 | 61 | 29 |
| Zone 3                              | 58                       | 24 | 44 | 13 | 14 | 60 | 48 | 28          | 50 | 51 | 50 | 28 |

resulting expenditure values for chromosome  $C_1$  using the values given in Tables II and III.

The fitness of chromosome  $C_1$  is found using equation (2), i.e. by summing the values in Table IV. In a similar fashion, the fitness values of chromosomes,  $C_2$ ,  $C_3$ ,  $C_4$ , and  $C_5$  are calculated using equations (1) and (2). The expenditure values,  $EXP_{ij}$ , from zone  $i$  to site  $j$  are calculated for each particular chromosome using equation (1). These expenditure values are then used, in equation (2), to give the fitness of the chromosome. It is important to note that the expenditure values from a particular zone to a particular site are dependent on the chromosome under consideration (a 0 in the chromosome indicating a site makes a zero contribution in equation (1) and equation (2)).

|       |               |                       |
|-------|---------------|-----------------------|
| $C_1$ | 1 0 1 0 0 1 1 | fitness = 128943.360  |
| $C_2$ | 0 0 1 1 1 1 0 | fitness = 131958.939  |
| $C_3$ | 1 0 1 0 0 0 1 | fitness = 128956.529  |
| $C_4$ | 1 1 1 1 0 0 0 | fitness = 138322.517  |
| $C_5$ | 1 1 1 0 1 0 1 | fitness = 136513.592. |

Chromosomes  $C_4$  and  $C_5$  are the most fit, so they would be more likely to be selected for cross-over and mutation. If one-point cross-over was used at a randomly chosen position, say position 4, the following offspring would be produced:

|       |                |
|-------|----------------|
| $C_4$ | 1 1 1 1 1 0 1  |
| $C_5$ | 1 1 1 0 0 0 0. |

Their respective fitness values, calculated using equations (1) and (2), are 142622.938 and 127819.997. We have, therefore, found a network of sites, represented by chromosome  $C_4$ , which is better theoretically than any network in the previous population.

The advantages in using GAs are:

- (1) they find an optimal or near-optimal site location in reasonable time;
- (2) the "goodness" of alternative site networks can be accessed easily;
- (3) poor networks can be identified and, therefore, avoided; and
- (4) it is relatively easy to solve different formulations of the site location problem.

However, the limitation in using GAs is that the "goodness" of any network is relative to the fitness function used. If an inappropriate fitness function is used, the resulting network may not be as profitable as possible.

| Zone | Site      |   |           |   |   |         |          |
|------|-----------|---|-----------|---|---|---------|----------|
|      | 1         | 2 | 3         | 4 | 5 | 6       | 7        |
| 1    | 10774.069 | — | 7964.717  | — | — | 38.166  | 7588.705 |
| 2    | 26206.036 | — | 33450.197 | — | — | 213.482 | 5837.120 |
| 3    | 13737.358 | — | 16819.599 | — | — | 109.116 | 6204.795 |

**Table IV.**  
Expenditure,  $EXP_{ij}$ ,  
from zone  $i$  to site  $j$  for  
chromosome  $C_1$

### **Example two: segmentation**

Though research on product-market mapping approaches continues to progress, there is evidence that tree structures rather than spatial ones may be more appropriate to the cognitive structure of many domains[59,60]. In the marketing literature, the growing number of techniques for competitive structure analysis often posit the presence of hierarchical choice structures in a market[61]. Recent evidence by Moore and Lehmann[62] emphasizes the importance of incorporating consumer heterogeneity in the context of estimating hierarchical structures from preference data. Hence, it seems potentially useful to develop techniques for deriving product-market trees. Product-market tree representations are consistent with typical hierarchical portrayals of market structures (or, more precisely, product structures) used commonly in product-partitioning approaches in the marketing literature[61,63].

Ramaswamy and DeSarbo[64] have proposed a new methodology for deriving hierarchical product-market structures from disaggregate purchase data. A hierarchical product-market tree was estimated from scanner panel purchase data in a maximum likelihood framework. The derived product-market representation portrays both products and market segments as terminal nodes in a hierarchical tree where the “closer” a product is to a particular segment the more it revealed preference for that product. The hierarchical representation of products and segments as well as the composition of the market segments are derived simultaneously. Their new methodology called SCULPTURE (stochastic ultrametric purchase tree) has been developed by formulating a stochastic choice model. Unlike typical hierarchical clustering approaches, SCULPTURE estimates both the competitive structure and the segmentation scheme simultaneously. This approach is similar to those of Grover and Srinivisan[65] and Kamakura and Russell[66], where the structure of the latent choice probabilities provides a basis for understanding the structure of the product market. There are various approaches for deriving product-market structures from choice data and all the procedures[67-70] require binary data.

### *Application of genetic algorithms*

In this section we consider the problem of deriving product-market structures from scanner panel purchase data. In the technique briefly described later, the structure is obtained by maximizing the likelihood. This will normally be a multi-modal function in many dimensions, so that iterative techniques from mathematical optimization theory may not always be appropriate. This optimization problem may be solved using a genetic algorithm and we suggest a representation and a method that might be employed. One advantage of optimizing in this way is that the possibility of obtaining poor local maxima is avoided. Instead, the GA samples over the whole solution space but gives an

exponentially increasing number of trials to short low-order combinations of genes with high fitness.

As already mentioned, Ramaswamy and DeSarbo[64] suggest a new technique for using panel data to determine a hierarchical tree representation where terminal nodes are used for both products and market segments. The smaller the “distance” between a product node and a market segment node, the higher the segment’s preference for that product. The raw data for the method are the values of  $x_{hj}$ , which represent the number of choices of product  $j$  for household  $h$ . The objective is to determine the hierarchical tree and the values of  $a_s$  which are the proportions of households in the sample belonging to segment  $s$ . The tree is determined uniquely from constrained values  $d_{ij}$  representing “distance” between terminal nodes  $i$  and  $j$ . The likelihood is a function  $L(\mathbf{a}, \mathbf{x}, \mathbf{d})$  given by:

$$L(\mathbf{a}, \mathbf{x}, \mathbf{d}) = \prod_h \frac{\left( \sum_j x_{hj} \right)!}{\prod_j x_{hj}!} \sum_s a_s \prod_j \frac{\text{EXP}(-d_{sj} x_{hj})}{\left( \sum_k \text{EXP}(-d_{sk} x_{hk}) \right)^{x_{hj}}}$$

The constraints on the distances so that they correspond to a hierarchical tree can be expressed using the non-negative integer valued function:

$$P(\mathbf{d}) = \sum_{r < s} \sum_{j < k} u_{rsjk}^2,$$

where  $u_{rsjk}$  represents the difference between the largest and second largest of the set of segment-product distances  $d_{rj}$ ,  $d_{rk}$ ,  $d_{sj}$ ,  $d_{sk}$ .

The determination of the best tree and mixing proportions is obtained by maximizing the likelihood function  $L$  subject to the constraint  $P = 0$ . One way that genetic algorithms could solve this problem is described as follows. Note that alternative representations may be more efficient but this would be a matter for experimentation. The genes of the chromosome would be the real values of  $a_s$  and the integer values of the  $d_{ij}$ ; children would be formed by one- or two-point cross-over. The fitness function would be the quantity  $L/(1 + P)$ , thereby combining the maximization of  $L$  with the minimization of  $P$ . A near optimal solution, however, would not actually correspond to hierarchical tree if  $P > 0$ . In such cases, the vector  $\mathbf{d}$  can then be modified to force the constraints and then further optimization can be applied to the vector  $\mathbf{a}$  to achieve a near optimal solution with  $P = 0$ .

As explained earlier, the genetic algorithm would be better than mathematical optimization techniques in circumstances when there are a large number of local maxima for  $L$  satisfying the constraints. However, such techniques could be applied if necessary to enhance the near optimal solution of the genetic algorithm. We would expect the genetic algorithm to perform more slowly than the mathematical optimization technique.

### **Conclusions**

We have given a brief description of a simple genetic algorithm and have shown how it can be applied to problems in marketing. The aim has been to bring to the attention of marketing managers this relatively unknown technique for solving optimization problems. In particular we have shown, using worked examples, how GAs can be used to aid site location decisions and help in market segmentation analysis. Also, several other areas of potential application have been identified and, provided that a suitable cost or fitness function can be constructed to represent the goodness of a particular solution in each of these areas, GAs should be able to give useful results. An important point is that, since genetic algorithms only use an encoding of the solution and its associated fitness value and do not utilize auxiliary information such as derivatives or assumptions about continuity in the solution process, they can be applied efficiently to a wide range of marketing optimization problems with only minor changes to parts of the algorithm.

### **Directions for further research**

The work outlined in this article is ongoing. In particular, the authors are involved currently in the following three areas of research:

- (1) Having conducted a primary investigation into using genetic algorithms for site location decision problems and market segmentation analysis, work is planned for an in-depth study of the benefits of using GAs in these two areas. Issues that arise include, for example, whether alternative fitness functions improve the quality of the solutions; the need for a comparison of results with previous real-life site location decisions; whether to use real market data or artificially generated data to test the algorithms.
- (2) Use of related techniques in solving marketing problems: currently, the authors are involved in applying an alternative optimization technique known as Tabu Search to problems in marketing. Other areas in need of research include classifier systems, simulated annealing, neural networks and expert systems. One of the authors has conducted a substantial amount of research into using expert systems for site location analysis. Consequently, comparisons between using the expert systems model and the genetic algorithm model are needed, as well as direct comparisons with the other techniques mentioned.
- (3) Tap new problems in marketing: from the list of potential applications of genetic algorithms in marketing (see Table I), we have selected a number of fields of experimentation to which we will devote our priorities. These new developments will take place in the following areas:
  - segmentation strategies for consumer and industrial markets;
  - optimization of product life cycles;

- describing advertising strategy and measuring its effectiveness; and
- optimization of describing channels and physical distribution strategies.

## References

1. Holland, J.H., *Adaption in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, USA, 1975.
2. Davis, L.D., *Genetic Algorithms and Simulated Annealing*, Pitman, London, 1987.
3. Davis, L.D., *Handbook of Genetic Algorithms*, Van Nostrand, Reinhold, New York, NY, 1991.
4. Goldberg, D.E., *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
5. De Jong, K.A., "An analysis of the behavior of a class of genetic adaptive systems", doctoral dissertation, University of Michigan, *Dissertation Abstracts International*, Vol. 36 No. 10, 1975, p. 5140B.
6. Nissen, V., "Evolutionary algorithms in management science", *Papers on Economics and Evolution*, No. 9303, European Study Group for Evolutionary Economics, 1993.
7. Noble, A., "Using genetic algorithms in financial services", in *Proceedings of the Two-Day International Conference on Forecasting and Optimization in Financial Services*, 3-4 October, IBC Ltd. Technical Services, London, UK, 1990.
8. Margarita, S., "Neural networks, genetic algorithms and stocktrading", in Kohonen, T., Makisara, K., Simmula, O. and Kangas, J. (Eds), *Artificial Neural Networks. Proceedings of the 1991 International Conference in Artificial Neural Networks*, Elsevier Science Publishers B.V. (North-Holland), Finland, 1991, pp. 1763-6.
9. Margarita, S., "Genetic neural networks for financial markets: some results", in Newmann, B. (Ed.), *Proceedings 10th European Conference on Artificial Intelligence*, John Wiley & Sons, Ltd, London, 1992, pp. 211-13.
10. Burke, G., "Good trading a matter of breeding?", *Futures: The Magazine of Commodities & Options*, Vol. 22 No. 5, May 1993, pp. 26-9.
11. Johnstone, B., "Research and innovation: remaking markets", *Far Eastern Economic Review*, Vol. 156 No. 12, March 1993, p. 50.
12. Star, M.G., "Artificial intelligence moves to forefront", *Pensions & Investments*, Vol. 21 No. 1, January 1993, pp. 14-16.
13. Schwartz, E.I., "Where neural networks are already at work", *Business Week*, Vol. 3291, November 1992, pp. 136-7.
14. Nottola, C., Leroy, F. and Davalo, F., "Dynamics of artificial markets", in Varela, F.J. and Bourgine, P. (Eds), *Towards a Practice of Autonomous Systems, Proceedings of the First European Conference on Artificial Life*, MIT Press, Cambridge, MA, 1992, pp. 185-94.
15. Khuri, S. and Batarekh, A., "Genetic algorithms and discrete optimization", in *Abstracts of the "Operations Research 1990"*, International Conference on Operations Research, Vienna, Austria, 28-31 August 1990, p. A 147.
16. Tam, K.Y., "Genetic algorithms, function optimisation, and facility layout design", *European Journal of Operations Research*, Vol. 63 No. 2, 1992, pp. 322-46.
17. Huntley, C.L. and Brown, D.E., "A parallel heuristic for quadratic assignment problems", *Computers & Operations Research*, Vol. 18 No. 3, 1991, pp. 275-89.
18. Hughes, M., "Improving products and processes – nature's way", *Industrial Management & Data Systems*, Vol. 90 No. 6, 1990, pp. 22-5.



19. Baker, E.K. and Schaffer, J.R., "Solution improvement heuristics for the vehicle-routing problems with time window constraints", *American Journal of Mathematics and Management Sciences*, Vol. 6 No. 3-4, 1991, pp. 261-300.
20. Thangiah, S.R., Nygard, K.E. and Juell, P.L., "GIDEON: a genetic algorithms system for vehicle routing with time windows", in *Proceedings of the IEEE Conference on Artificial Intelligence Applications*, Miami, FL, 1991, pp. 322-8.
21. Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Program*, Springer, Berlin, 1992.
22. Thompson, B. and Thompson, B., "Evolving knowledge from data", *Computer Language*, Vol. 11 No. 3, 1986, pp. 23-6.
23. Bruderer, E., "How organizational learning guides environmental selection", School of Business Administration, The University of Michigan, unpublished working paper, 1992.
24. Bean, J.C., "Genetics and random keys for sequencing and optimization", Department of Industrial and Operations Engineering, University of Michigan, unpublished Technical Report, No. 92-43, 1992.
25. Vancza, J. and Markus, A., "Genetic algorithms in process planning", *Computers in Industry*, Vol. 17 No. 2, November 1991, pp. 181-94.
26. Stevens, T., "Fuzzy logic makes sense", *Industry Week*, Vol. 242 No. 5, March 1993, pp. 36-42.
27. Venugopal, V. and Narendran, T.T., "A genetic algorithm approach to the machine-component grouping problem with multiple objectives", *Computers & Industrial Engineering*, Vol. 22 No. 4, October 1992, pp. 469-80.
28. Ward, T.L., Ralston, P.A.S. and Davies, J.A., "Fuzzy logic control of aggregate production planning", *Computers & Industrial Engineering*, Vol. 23 Nos 1-4, November 1992, pp. 137-40.
29. Kulkarni, J. and Parasei, H.R., "Information resource matrix for production and intelligent manufacturing using genetic algorithm techniques", *Computers & Industrial Engineering*, Vol. 23 Nos 1-4, November 1992, pp. 483-5.
30. Ward, T.L., Ralston, P.A.S. and Stoll, K.E., "Intelligent control of machines and processes", *Computers & Industrial Engineering*, Vol. 19 Nos 1-4, 1990, pp. 205-9.
31. Dowsland, K.A. and Dowsland, W.B., "Packing problems", *European Journal of Operational Research*, Vol. 56 No. 1, January 1992, pp. 2-14.
32. Biegel, J.E. and Davern, J.J., "Genetic algorithms and job shop scheduling", *Computers & Industrial Engineering*, Vol. 19 Nos 1-4, 1990, pp. 81-91.
33. Gordon, M.D., "User-based document clustering by redescribing subject descriptions with genetic algorithms", *Journal of the ASIS*, Vol. 42 No. 5, June 1991, pp. 311-22.
34. Johnstone, B., "Modelling the future", *Far Eastern Economic Review*, Vol. 153 No. 32, August 1991, p. 70.
35. Badami, V.S. and Parks, C.M., "A classifier-based approach to flow shop scheduling", *Computers & Industrial Engineering*, Vol. 21 No. 1-4, 1991, pp. 329-33.
36. Lane, D.A., "Artificial worlds and economics", Santa Fe Institute, unpublished working paper, 92-09-048, 1992.
37. Greene, D.P. and Smith, S.F., "A genetic system for learning models of consumer choice", in Grefenstette, J.J. (Ed.), *Genetic Algorithms and Their Applications, Proceedings of the Second International Conference on Genetic Algorithms*, MIT, Cambridge, 28-31 July, Lawrence Erlbaum, Hillsdale, NY, 1987, pp. 217-23.
38. Marengo, L., "Structure, competence and learning in an adaptive model of the firm", *Papers on Economics and Evolution*, No. 9203, European Study Group for Evolutionary Economics, 1992.

39. Tang, Y.E. and Holak, S.L., "A dynamic model of the product evolutionary cycle: applying a biological mechanism to dynamics of product-markets", *TIMS Annual Conference*, London Business School, England, 13-14 July 1992.
40. Balakrishnan, P.V.S. and Jacob, V.S., "A genetic algorithm-based decision support system for optional product design", *TIMS Annual Conference*, London Business School, England, 15-17 July 1992.
41. Lilien, G.L., Kotler, P. and Moorthy, K.S., *Marketing Models*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1992.
42. Moutinho, L., Curry, B. and Davies, F., "Comparative computer approaches to multi-outlet retail site location decisions", *The Service Industries Journal*, Vol. 13 No. 4, 1993, pp. 201-20.
43. Curry, B. and Moutinho, L., "Expert systems for site location decisions", *Logistics Information Management*, Vol. 4 No. 4, 1991, pp. 19-27.
44. Curry, B. and Moutinho, L., "Expert systems and marketing strategy: an application to site location decisions", *Journal of Marketing Channels*, Vol. 1 No. 1, 1991, pp. 23-37.
45. Curry, B. and Moutinho, L., "Computer models for site location decisions", *International Journal of Retail & Distribution Management*, Vol. 20 No. 4, July/August 1992, pp. 12-17.
46. Stern, L.W. and El-Ansary, A.I., *Marketing Channels*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.
47. Breheny, M.J., "Practical methods of retail location analysis: a review", in Wrigley, N. (Ed.), *Store Choice, Store Location and Market Analysis*, Routledge, London, 1988, Chapter 2, pp. 39-86.
48. Simkin, L.P., "SLAM: store location assessment model – theory and practice", *OMEGA*, Vol. 7 No. 1, 1989.
49. Simkin, L.P., "Evaluating a store location", *International Journal of Retail & Distribution Management*, Vol. 18 No. 4, 1990.
50. Penny, N.J. and Broom, D., "The Tesco approach to store location", in Wrigley, N. (Ed.), *Store Choice, Store Location and Market Analysis*, Routledge, London, 1988, pp. 106-19.
51. Jones, K. and Mock, D., "Evaluating retail trading performances", in Davies, R.L. and Rogers, D. (Eds), *Store Location and Store Assessment Research*, John Wiley, Chichester, UK, 1984, pp. 333-60.
52. Beaumont, J.R., "Store location analysis: problems and progress", in Wrigley, N. (Ed.), *Store Choice, Store Location and Market Analysis*, Routledge, London, 1988, Chapter 3, pp. 87-105.
53. Nakanishi, M. and Cooper, L.G., "Parameter estimate for multiplicative competitive interactive model – least squares approach", *Journal of Marketing Research*, Vol. 11, August 1974, pp. 303-11.
54. Mahajan, V., Jain, A.K. and Ratchford, B.T., "Use of binary attributes in the multiplicative competitive interactive choice model", *Journal of Consumer Research*, Vol. 5, December 1978, pp. 210-15.
55. Achabal, D.D., Gorr, W.L. and Mahajan, V., "MULTILOC: a multiple store location decision model", *Journal of Retailing*, Vol. 58 No. 2, 1982, pp. 5-25.
56. Teitz, M.B. and Bart, P., "Heuristic methods for estimating the generalized vertex median of a weighted graph", *Operations Research*, Vol. 16, 1968, pp. 955-61.
57. Rosenfield, D.B., "The retailer facility location problem: a case study", *Journal of Business Logistics*, Vol. 8, 1987, pp. 95-114.
58. Hogue, J.T. and Fields, M., "Where should we put the store?", *Information Strategy: The Executive's Journal*, Vol. 5, Summer 1989, pp. 16-19.
59. Pruzansky, S., Tversky, A. and Carroll, J.D., "Spatial versus tree representation of proximity data", *Psychometrika*, Vol. 47, March 1982, pp. 3-24.

60. Tversky, A., "Features of similarity", *Psychological Review*, Vol. 84, July 1977, pp. 327-52.
61. Rao, V. and Sabavala, D.J., "Inference of hierarchical choice processes from panel data", *Journal of Consumer Research*, Vol. 8, June 1981, pp. 85-96.
62. Moore, W.L. and Lehmann, D.R., "A paired comparison nested logit model of individual preference structures", *Journal of Marketing Research*, Vol. 26, November 1989, pp. 420-8.
63. Grover, R. and Dillon, W.R., "A confirmatory methodology for hierarchical market structure analysis", *Marketing Science*, Vol. 4, Autumn, 1985, pp. 312-35.
64. Ramaswamy, V. and DeSarbo, W.S., "SCULPTURE: a new methodology for deriving and analyzing hierarchical product-market structures from panel data", *Journal of Marketing Research*, Vol. 27, November 1990, pp. 418-27.
65. Grover, R. and Srinivisan, V., "A simultaneous approach to market segmentation and market structuring", *Journal of Marketing Research*, Vol. 27, May 1987, pp. 139-53.
66. Kamakura, W.A. and Russell, G.J., "A probabilistic choice model for market segmentation and elasticity structure", *Journal of Marketing Research*, Vol. 26, November 1989, pp. 379-90.
67. DeSarbo, W.S. and Cho, J., "A stochastic multidimensional scaling vector threshold model for the spatial representation of binary choice data", *Psychometrika*, Vol. 54, March 1989, pp. 105-30.
68. DeSarbo, W.S. and Hoffman, D., "Constructing MDS joint spaces from binary choice data: a new multidimensional unfolding threshold model for marketing research", *Journal of Marketing Research*, Vol. 27, February 1987, pp. 40-54.
69. Holbrook, M.B., Moore, W.L. and Winer, R.S., "Constructing joint spaces from pick-any data: a new tool for consumer analysis", *Journal of Consumer Research*, Vol. 9, June 1982, pp. 99-105.
70. Kamakura, W.A. and Srivastava, R., "An ideal-point probabilistic choice model for heterogeneous preferences", *Marketing Science*, Vol. 5, Summer 1985, pp. 199-218.