

A bio-mimetic approach to marketing segmentation: Principles and comparative analysis

JACQUES-MARIE AURIFEILLE ¹

Abstract. – A regression algorithm is proposed for partitioning a population into clusters characterised by homogeneous models and predictions. This algorithm, Typren, is based on the hybridisation of a genetic algorithm with linear regression. An empirical illustration is provided, using real marketing data, which compares Typren with the fuzzy clustering approach, Glimmix, based on a regression mixture model. Typren provides better predictivity and better within-cluster homogeneity of predictions. However, the results are slightly less robust compared to Glimmix.

Introduction

A central tenet of marketing is that the firm must define its production according to consumer's needs. However, researching these needs generally leads to the identification of several groups of individuals, characterised by specific expectations. As a result, marketers must often consider whether to target one or other of these groups and their choice is based on the objectives of the firm, such as sales turnover or potential market volume for each group.

Segmenting market raises a double challenge:

- Clustering: the aim is to categorise the population in a way which maximises the homogeneity of consumers' profiles within groups while at the same time maximising the heterogeneity of consumers' profiles between groups. The variables used to describe these profiles will be thereafter known as 'predictors'.
- Predictivity: the groups must be also characterised by a degree of homogeneity in relation to the market objective sought, *e.g.* heavy *vs.* light users. The term 'purchase' will be used thereafter to refer to these market objectives towards which the segmentation is undertaken.

Several methods have been proposed to solve this dual problem. On the one hand, associative methods associate a precise level of purchase to typical predictor profiles. This association is atheoretical in that no attempt is made to estimate the parameters of a formal model in which the predictors would be the exogenous variables and purchase

¹ The FACIREM & GREGOI labs., University of La Reunion. E-mail: Aurifeil@univ-reunion.fr

Keywords: Segmentation, clusterwise regression, genetic algorithms.

level the endogenous one. Notable associative approaches include Learning Vector Quantization (Kohonen, 1987) or combinations of conjoint analysis and descriptive clustering (Green and Krieger, 1991). Regressive methods, on the other hand, seek groups sharing homogeneous models. The best known amongst those are Clusterwise Regression (Späth, 1979; 1981; 1982; 1985), Simulated Annealing (De Sarbo, Oliver and Rangaswamy, 1989) or Mixture Model approaches (Wedel and DeSarbo, 1995).

These two types of approaches reflect different marketing problems. Associative methods are best suited for identifying consumers. The selection of the usual predictors, including age, gender, occupation or income, enables the choice of a target segment based on communication or distribution costs. These also allow the later categorisation of consumers who did not belong to the initial sample and a 'scoring' process similar to that applied by many mail order organisations or banks. Conversely, regressive methods are best suited when the aim is to understand consumer behaviour in order to relate the purchase to some predictors. This should suggest more sophisticated and subtle persuasive communication than would associative methods. However, their reliance on internal, psychographic variables, makes the identification of the target and the establishment of direct contact more difficult, since the focus is on identifying the model of consumer behaviour rather than the consumers who undertake it. This 'dehumanisation' of the target is consistent with current trends towards the identification of phenomena as opposed to individuals. For example, at different times, a consumer may face different situations in relation to the same product. For each situation, his needs and behaviours may vary considerably. Thus, strategic decisions should be made that reflect situational characteristics as opposed to the consumer's basic identity. Hence, this present paper is dedicated to regressive segmentation methods alone.

The next section describes and discusses the main regressive segmentation algorithms. A connectionist approach, Typren, based on the hybridisation of a genetic algorithm with classical regression, is then presented. The final part of this paper provides an empirical study, based on real data, in order to illustrate the working of Typren. It also provides a comparison of its performance with that of Glimmix, a fuzzy clustering method based on a regression mixture model which appears as the most advanced of the non-connectionist approaches (Wedel and Kamakura, 1997) and which is now available as a commercial software package.

Regressive approaches

Two main types of regressive approaches can be identified: two-step or simultaneous. Two-step processes begin either with a cluster analysis followed by a regression in each cluster, or by a regression (conjoint analysis, Green and Srinivasan, 1978) followed by a clustering in which the predictors are moderated by their predictivity. The latter approaches are more common, especially those based on estimates of predictors' weights at the individual consumer level (Kamakura, 1988; Green and Krieger, 1991). Whilst used relatively frequently, two step approaches do not deal simultaneously with the criterion of group homogeneity and that of predictivity: one is favoured at the expense of the other.

Moreover, the use of purely descriptive clustering introduces a high degree of indeterminacy in the solution. Our focus, therefore, will be on simultaneous approaches.

Simultaneous approaches are such that assignment to a cluster must take into account the predictivity criterion. In this paper, the most commonly used procedure (Automatic Interaction Detection (AID), (Assael and Roscoe, 1976) is discussed, before more formal optimisation methods (Clusterwise Regression, Simulated Annealing, Mixture Models) which are currently replacing it.

Aid

AID is a dichotomous categorisation technique which extracts, in a stepwise manner, the predictor discriminating the most clearly between heavy and light purchasers. As a rule, categorisations are dichotomous and predictors must therefore be transformed into binary variables. Each predictor is subjected to an analysis of variance (ANOVA) in order to determine the one for which the two values are related to the most contrasted purchase levels. This procedure is then repeated separately for each of the two sub-group thus obtained, examining only those remaining predictors.

Let Y be the continuous variable and N the number of individuals. Without segmentation, the prediction relating to individuals would be the mean of their purchases, thus providing a Sum of Squared Errors (SSE) equal to:

$$\sum_{i=1, N} (Y_i - \bar{Y})^2. \quad (1)$$

When the population is split into two clusters, c_1 and c_2 , numbering n_1 and n_2 , the SSE becomes:

$$\sum_{j=1, 2} \left(\sum_{i \in C_j} Y_{i,j}^2 - n_j \cdot \bar{Y}_{i,j}^2 \right) = \sum_{i=1, N} Y^2 - (n_1 \cdot \bar{Y}_{i,1}^2 + n_2 \cdot \bar{Y}_{i,2}^2). \quad (2)$$

Hence the dichotomy has induced a reduction in the error term of :

$$n_1 \cdot \bar{Y}_{i,1}^2 + n_2 \cdot \bar{Y}_{i,2}^2 - N \cdot \bar{Y}^2. \quad (3)$$

The selected variable is the one maximising this reduction.

AID analyses are popular with companies, as they do not require an *a priori* number of clusters. Moreover, by providing a nested view of the segments, they allow a concentric expansion of the target market. On the other hand, AID has a number of limitations. Its empirical validity is reduced by its requirement of binary variables and by the absence of a formal model. Whilst some algorithms may use variables with more than two levels (McLachaln and Johansson, 1981), the number of segments obtained at each stage remains a function of the number of levels and a rapid atomisation of the population in minutes segments occurs. Statistically, the dichotomisation may also hide significant interactions. For example, if two interactions of decreasing order $a * b$, $a * c$ exist but no interaction exists between b and c , extracting a then b may mask the $a * c$ interaction. For these reasons, the authors of AID recommend its use as a preliminary tool for identifying the importance of the predictors, before undertaking a complementary cluster analysis (Doyle and Hutchinson 1976; Doyle and Fenwick 1975). This is akin to the two steps approaches previously discussed, along with their shortcomings.

Simultaneous optimising approaches

These methods of segmentation consist in a simultaneous exploration of clusters and models, and therefore, in estimating two types of parameters:

- A matrix Alpha = $\{\alpha_{i,c}\}$ indicating the degree to which an individual i belongs to a segment c . For a clear categorisation, belonging is a binary variable. For fuzzy clustering, the variable is the probability that i belongs to c : $\sum \alpha_i = 1$.
- A matrix Beta = $\{\beta_{p,c}\}$ indicating the weight of predictor p in segment c .

The corresponding algorithms are heuristic comprising numerous iterations in which two sub-iterations alternate: one for estimating alphas and one for estimating betas. Given an individual i , an endogenous variable y , J exogenous variables x_j and a function g corresponding to the distribution of the y (e.g. identity function for normal distribution, logit for binomial distribution, log for Poisson distribution, inverse for Gamma distribution and so forth), the prediction error relative to i is as follows:

$$g(y_i) - \sum_{c=1}^C \sum_{j=1}^J (\alpha_{i,c} \cdot (\beta_{j,c} \cdot x_{i,j} + k_j)) \quad (4)$$

After transformation of this error by an ad hoc function (e.g. squared error), segmentation becomes a matter of minimising the function F :

$$F = \sum_{i=1}^N g \left((y_i) - \sum_{c=1}^C \sum_{j=1}^J (\alpha_{i,c} \cdot (\beta_{j,c} \cdot x_{i,j} + k_j)) \right)^2 \quad (5)$$

The first algorithms of simultaneous segmentation (Späth, 1979; 1981; 1985) are extensions of k -means clustering and provide strict partitioning. The basic procedure is as follows: having fixed the number of clusters and assigned arbitrarily individuals to clusters, one examines each individual in turn to assess whether his or her transfer to another cluster would reduce F , with g the identity function and F the squared error. The main limitation of the algorithms proposed by Späth (1979; 1981; 1982) is that they do not allow any temporary decay in the fit. They are, therefore, prone to local minima and highly dependent on the sequence according to which the transfers are undertaken. Other algorithms have thus been sought, based on bio-mimetic approaches such as Simulated Annealing, or alternatively, based on statistical bases derived from mixture model approaches.

The use of the heuristic 'Simulated Annealing' for solving segmentation problems has been illustrated by DeSarbo, Oliver and Rangaswamy (1989). This process differs from that advocated by Späth in two ways:

- The transfers of individuals are considered randomly as opposed to systematically,
- Some temporary decay of fit is tolerated.

The algorithm is as follows:

- (i) Initialisation: fix the number of clusters and attribute random parameters (regression coefficients and constant) close to zero for each clusters, then assign randomly individuals to clusters. Fix a maximum percentage of decay allowed in the global

optimum, D_0 . Fix a ‘cooling factor’, r , smaller than 1. Fix the number n of transfers of individuals for each iteration. Let $t = 0$.

- (ii) Let $t = t + 1$, $D_t = r \cdot D_r$. Let $s = 0$. Measure the global fit $F_{t,s}$.
- (iii) If $s < n$, let $s = s + 1$:
 - choose randomly a cluster c and an individual i not belonging to this cluster;
 - assign i to c and calculate the new global fit $F_{t,s}$. Randomly select a percentage of progression for the global fit, f , such that $f < D_t \cdot F_{t,s-1}$. If $F_{t,s} > f$, then take the individual i back to the initial cluster, otherwise go back to (iii).
- (iv) If the end criterion is achieved (maximum number of iterations, acceptable global fit, weak progression of the fit), then go to (v), otherwise go back to (ii).
- (v) End.

Theoretically, the simulated annealing approach should allow the achievement of the absolute optimum. In practice, however, one must often limit the number of transfers (as far as $n = 1$) and favour a very slow cooling factor in order to obtain results comparable with those derived with Späth’s algorithms (Aurifeille and Bergmann, 1995).

Mixture models (McLachlan and Basford, 1988) have been extended to the case of Clusterwise Regression by Wedel and DeSarbo (1995). A software now exists under the name of Glimmix (Wedel, 1998) which is available commercially. Using the previous formulas and describing α_c as the percentage of individuals from cluster c , and considering that y_i is distributed as the finite sum of the mixed normal univariate conditional densities, it is necessary to estimate α , σ^2 et β_{jc} so as to maximise a probability function H such that:

$$H = \sum_{i=1}^N \ln \left[\sum_{c=1}^C \alpha_c \cdot f_{i,c}(y_i | x_{i,j} \hat{\sigma}_c^2, \hat{\beta}_{j,c}) \right] \tag{6}$$

where:

$$f_{i,c}(y_i | x_{i,j} \hat{\sigma}_c^2, \hat{\beta}_{j,c}) = (2\pi\sigma_c^2)^{-1/2} \cdot \exp \left[\frac{-(y_i - x_i \hat{\beta}_c)}{2\sigma_c^2} \right] \tag{7}$$

with $0 \leq \alpha_c \leq 1$ and $\sum \alpha_c = 1$.

Once the estimations are made, it becomes possible to assign each i to a segment according to the *a posteriori* probability given by Bayes’ formula:

$$\hat{p}_{i,c} = \frac{\hat{\alpha}_c f_{i,c}(y_i | x_{i,j} \hat{\sigma}_c^2, \hat{\beta}_{j,c})}{\sum_{c=1}^C \hat{\alpha}_c f_{i,c}(y_i | x_{i,j} \hat{\sigma}_c^2, \hat{\beta}_{j,c})} \tag{8}$$

The estimation of α , σ^2 and β_{jc} is achieved y reformulating the probability function to take into account the constraints on α :

$$H = \sum_{i=1}^N \ln \left[\sum_{c=1}^C \alpha_c \cdot f_{i,c}(y_i | x_{i,j} \hat{\sigma}_c^2, \hat{\beta}_{j,c}) \right] - \mu \left(\sum_{c=1}^c \alpha_c - 1 \right) \tag{9}$$

After development and simplification, the parameters can be defined using $p_{i,c}$ as the weighted averages of the maximum probability equations:

$$\sum_{i=1}^N \hat{p}_{i,c} \frac{\partial \log f_{i,c}(y_i | x_{i,j}, \hat{\sigma}_c^2, \hat{\beta}_{j,c})}{\partial \sigma_c^2} = 0 \quad (10)$$

and

$$\sum_{i=1}^N \hat{p}_{i,c} \frac{\partial \log f_{i,c}((y_i | x_{i,j}, \hat{\sigma}_c^2, \hat{\beta}_{j,c}))}{\partial \beta_{j,c}} = 0. \quad (11)$$

The algorithm used (DeSarbo and Cron, 1988; Wedel and Kamakura, 1997) is known as EM (for Estimation-Maximisation) (Dempster, Laird and Rubin, 1977) and its convergence has been well established. Initially, it aims to estimate a mixture regression model when data is missing. Given a distribution hypothesis of the α and of the predictions y , the process involves alternating several times an estimation phase of the parameters p and of the corresponding α with a maximisation phase of the likelihood providing an estimation of the parameters β and σ^2 . The algorithm of the Clusterwise Regression consists in considering some values α as missing, estimating them by EM, reintroducing them in the likelihood function, maximising this function and repeating the estimation.

This procedure has been extended by Wedel and DeSarbo (1995) to several common cases: plurality of dependant variables, repeated measures, endogenous and exogenous variables, discrete or continuous. It also enables the integration of several likelihood functions, according to the assumptions made concerning the data distribution. The availability of a software package (Glimmix) makes this approach particularly attractive (Wedel 1998).

Despite these advantages, the process is based on statistical assumptions which may limit its effectiveness. These assumptions relate to the distribution of the predictors, the probabilities of cluster membership and the type of functions operating within each cluster. Moreover, this approach shares a drawback of Simulated Annealing since the homogeneity of the predictions for each cluster is not controlled for. Hence, one may achieve clusters characterised by homogeneous models but with heterogeneous predicted scores.

Finally, one may consider fuzzy clustering as a poor substitute for clear categorisation as the increase in global fit thus achieved is gained at the cost of lesser interpretable assignments. Often, fuzzy clustering is interpreted as if it described the probability of alternative assignments. For example, a consumer belonging for 70% to a cluster and 30% to another would follow the model of the first cluster in 70% of cases and that of the second cluster in the other cases. This interpretation, however, is inaccurate since the individual belongs simultaneously to both clusters, yet with differing degrees of intensity. It is therefore difficult to operationalise a fuzzy cluster analysis, be it at the level of the consumer or that of the models. For the sake of practicality, most practitioners assign the consumer to that cluster where the stronger probability of membership is apparent. Yet, if the aim of such clustering is to achieve a partition, a better optimality could result from seeking it directly. An intermediary solution is often chosen which limits the degree of acceptable

fuzz in the heuristic. However, this implies many trials before an acceptable compromise can be reached between predictivity and fuzz.

Overall, these restrictions lead us to propose an alternative process during which a partition is sought *via* the use of a bio-mimetic algorithm.

A bio-mimetic algorithm for regressive segmentation

Bio-mimetic approaches have been considered in the past to solve the problems associated with regressive segmentation (see previous section on LVQ and Simulated Annealing). This can be explained by the topological complexity of the problems as well as the restrictive nature of the statistical assumptions on which the non bio-mimetic approaches rely (distributions, independence, similarity of assumptions across segments).

Bio-mimetic approaches rely on many agents, equipped with simple rules, seeking solution in a parallel manner and exchanging information about their search *via* a dense network of connections (neural networks) or *via* an 'evolutionary' process of selection and reproduction (genetic algorithm). Although it is possible to explain a bio-mimetic heuristic in statistical terms, it does not rely on statistical assumptions and favours instead a trial and error approach based on large numbers of random trials.

Neural networks are appropriate when the shape of the model is unknown and they are therefore more suited to associative segmentation (*e.g.* LVQ). Other processes, such as genetic algorithms, can be used to estimate the parameters of a function explicitly formulated and are therefore more suited to the focus of this paper.

A genetic algorithm (Goldberg, 1991a) relies on several 'chromosomes', that is on several vectors comprising several complete sets of parameters to estimate. The coding of these values can be either binary or real. The latter is best suited for cases when the parameters to estimate are also real, require a high degree of precision, or are numerous (Renders, 1995, pp. 90-100; Goldberg, 1991a).

The effectiveness of genetic algorithms (GAs thereafter) stems from the parallel and random nature of their explorations. Unlike the single 'climber' used in gradient descents, GAs exploit information derived from multiple agents. Moreover, GAs can contract or expand their field of exploration on a continuous basis. Finally, GAs can afford the imposition of many types of constraints (*e.g.* with no derivative).

However, GA search is slow. This is why they are sometimes associated with a more deterministic heuristic, as indeed is proposed in this paper.

One particular characteristic of the Typren algorithm is its ability to optimise simultaneously the homogeneity of the models (coefficients) and the predictions for each segment. It aims to avoid the case of individuals following the same model and thus belonging to the same cluster, but exhibiting markedly different purchase levels. Whilst this problem is not explicitly addressed by regressive segmentation algorithms, it remains a managerial imperative for the marketing decision maker wanting to select a segment based on its purchase level.

The characteristics of the Typren algorithm are as follows:

(i) Initiation. Fix a number of chromosomes, K . Allocate to each chromosome $C^*(X + 1)$ random values very close to zero. Fix a renewal rate of the chromosome population: $0 < t \leq 1$.

(ii) For each chromosome $h = 1 \dots H$, assign each individual to the segment c of which the parameters $[x_{i,c}]$ minimise the prediction error:

$$c | \min \left(e_i = \hat{y}_i - \sum_{j=1}^J (\beta_{j,c} \cdot x_{i,c} + k_c) \right). \quad (12)$$

(iii) After all individuals have been assigned, estimate with stepwise linear regression the parameters β and k of each segment.

(iv) Measure the standard deviation of the chromosomes predictions:

$$E(h) = \sum_{c=1}^C \sum_{i \in c} (y_i - \bar{y}_c)^2. \quad (13)$$

(v) Repeat $t \cdot K$ times the following procedure:

– Draw two chromosomes randomly, with probability of the draw inversely proportional to their standard deviations.

– Split each chromosomes in two strands, on either sides of a randomly selected cutting point (whole number between 0 and $S \cdot X$).

– Create two new chromosomes by interpolating/extrapolating the values of the first strand (new chromosome 1) and the values of the second strand (new chromosome 2). The details of the interpolation/extrapolation used is described elsewhere (Aurifeille, 1998): random draw between 3 possible types of calculation: average weighted by standard deviations, simple permutation or values multiplied by a random coefficient between 0 and 2.

(vi) For each new chromosome: select at random several old chromosomes, except for that with the lowest dispersion, eliminate that with parameters closest to the new chromosome (Euclidean distance).

(vii) If the convergence criterion is achieved (e.g. limit number of iterations, stability of the values for the best chromosome) go to (viii), or else go to (ii).

(viii) End. The solution corresponds to the values of the chromosome with the lowest standard deviation.

According to the empirical work undertaken by De Jong (1975), the following options have been adopted:

- No mutations, since the crossover rule creates enough random disturbance.
- An elitist rule: the best chromosome is never replaced.
- Use of a crowding factor (a chromosome replacement process aimed at moderating the speed of convergence, Davis, 1991). As per Dejong (1975), we have observed that the best results are achieved with a crowding factor of 2.
- The rate of chromosome renewal (step v) is fixed, by default, to 60% in Typren.

The combination of GA with a classical linear regression analysis – step (ii) – aims to redress one of the major weakness of GAs by accelerating considerably the heuristic.

This combination also allows to control for multi-colinearity with the use of the stepwise regression in step (ii). These advantages are significant:

- The average number of significant exogenous variables is increased.
- The risk of obtaining an indeterminate solution is lessened, notably when compared with Glimmix solutions.
- The size of segments is increased. This advantage is not necessarily linked with lesser optimality, but results from the mechanism according to which a smaller segment is more likely to be multi-colinear, and therefore, subject to correction through the stepwise regression. Thus, the chromosomes containing these segments are less likely to reproduce.

In addition, the Typren process combines two criteria: the predictability of the models of each segment (step ii) and the homogeneity of the predictions (step IV to VI). The selection of the chromosomes is undertaken in relation to the homogeneity of the prediction, hence this principle should dominate. However, in practice, the preliminary assignment to clusters plays an important part, such that the procedure converges by both decreasing the prediction error and increasing the prediction homogeneity. This is illustrated more fully in the following empirical section as well as in appendix.

Empirical illustration of Typren and comparison with Glimmix

This section will first describe the data set used for this comparison, along with some global statistics. Then the criteria used for comparing the two processes are discussed, before a detailed examination of the results is provided.

Data used for the comparison

The comparison deals with real data, collected from 431 wine consumers for a study undertaken by Lockshin *et al.* (1997) who attempted to explain the yearly expenditures of wine based on three types of involvement, product-, brand- and purchase-related.

Involvement is a fundamental concept in consumer behaviour: at the nexus between motivations (the internal forces that drive behaviours) and product attributes, it reflects the situational nature of purchases. For example, a wine buyer will not be similarly involved depending on the intended consumption situation, be it an everyday meal or a special occasion. The time spent in performing the purchase, the sensitivity to the brand, to price or to the retail outlet are some of the variables affected by situation. Hence, a segmentation based on involvement profiles seems wiser than one based on consumers.

According to Laurent and Kapferer (1985), involvement depends on 5 factors: pleasure, social status afforded by the consumption, technical interest for the product characteristics (production process, storage...), risk of mistake, probability of mistake. Based on this, three domains of involvement can be identified: the product, the brand and the purchase act itself. As shown by Lockshin *et al.* (1997), a consumer can be involved in the product without being involved in the brand. Likewise, according to Mittal and Lee (1989) or Ohanian and Tashchian (1992), a consumer can be involved in the purchase act (search, negotiation) independently from the brand or the product concerned.

The data was collected for the purpose of examining whether the three levels of involvement could predict consumers' wine purchase. The analysis of the co-variance

structures with LISREL 8.1 (Linear Structural Relationship – Jöreskog and Sörbom, 1996), confirmed the weak correlation of the three principal components of involvement: in the product (PI), in the brand (BI) and in the purchase (PuI). Considering the interaction of second and third orders, seven predictors were defined for the whole population: PI, BI, PuI, PI * BI, PI * PuI, BI * PuI, PI * BI * PuI. Table 1 shows the result of the global correlation of predictors. The dependant variable is the consumer's yearly expenditure on wine.

TABLE 1. Global correlation of predictors.

	PI	BI	PuI	PI * BI	PI * PuI	BI * PuI	PI * BI * PuI
PI	1.0000						
BI	0.1564	1.0000					
PuI	0.3747	0.5392	1.0000				
PI * BI	-0.0481	-0.2272	-0.1159	1.0000			
PI * PuI	-0.1008	-0.1130	-0.2253	0.5777	1.0000		
BI * PuI	-0.0890	-0.4665	-0.3489	0.4283	0.3529	1.0000	
PI * BI * PuI	0.3451	0.3364	0.3491	-0.5420	-0.5011	-0.5580	1.0000

Method of comparison

The comparison of Typren and Glimmix was undertaken by running 10 trials of each procedure for each size of solution. For each attempt, a new set of initial parameters was generated. Solutions from 1 to 8 clusters were examined, hence a total of $2 \cdot 7 \cdot 10 = 140$ cluster analyses of 431 consumers.

Whilst Glimmix does not exclude partitioning, its author does not consider this a desirable application of this heuristic (Wedel, 1990) and the software, as a result, does not offer this option (Glimmix, Wedel, 1998). The comparison, therefore, involves two solutions with different structures: one fuzzy (Glimmix) and one partitioned (Typren). The study will use criteria that do not require a preliminary treatment of the solutions (squared correlation, sum of squared errors) and others that require a 'hardening' of the Glimmix solution (SSE hard, R2 hard, prediction variation for each cluster). 'Hardening' will involve the transformation of the fuzzy clustering in partitioned clusters by assigning the individuals to the segment where their probability of membership is greater.

As discussed in the previous section, whilst 'hardening' is common practice, its legitimacy is debatable. However, its main justification stems from the operational imperatives faced by marketing decision makers and from the fact that a good fuzzy cluster analysis should provide only limited multiple membership.

The last criterion used in this comparison is the robustness of the methods, or their ability to provide comparable solutions when noises are introduced in the data or in the solution used as the starting point for the analysis. Given that we are dealing with complex heuristics, the risk of convergence towards local minima is high. Indeed, the authors of Glimmix suggest that several successive analyses be undertaken in order to select the best one. However, the probabilistic nature of Glimmix does suggest that a good

robustness can be achieved. In the case of Typren, the parallel nature of the GAs and their lesser sensitivity to past explorations suggest that even more stable solutions could be expected. In this paper, the robustness is tested in relation to assignments. A perfectly robust process would result in consumers pairs who would either always or never be in the same cluster, across all 10 solutions attempted. This would correspond to a maximum variation in the vector summarising the frequencies of co-assignment of all possible pairs of consumers.

Hence the 6 criteria retained for the comparison (with their names following in brackets):

- C1: Sum of the Squared distances (SSE).
- C2: Global R Square (R2).
- C3: SSE of 'hardened' solution (SSE hard).
- C4: R2 of hardened solutions (R2 hard).
- C5: Within-cluster variation in purchase levels as described in Typren step (iii) (Dispers.) Limiting this variation is a necessary condition to enable the selection of segment based on their attractive sales potential. This is measured by the sum of squared distances between the consumers' predicted purchase levels and the mean prediction within the cluster they belong to.
- C6: standard deviation of the frequency vector of co-assignment between all pairs of consumers (Robust).

Results and discussion

The theoretical convergence of the EM algorithm used in Glimmix has been demonstrated. In practice, however, the procedure is interrupted when the marginal gain in fit becomes very small. This criterion is not well established and so this paper used the maximum number of possible iterations for Glimmix 1.0 (of 250). The marginal gains of the last iterations were all included between 10^{-4} and 10^{-6} .

For Typren, the convergence is assured by the progressive homogenisation of the chromosomes. The systematic conservation of the best historical solution enhances this convergence and results in the maintenance of the best solution after a given number of iterations. Since Typren uses different criteria in the assignment phase (minimising the error) and in the selection/reproduction phase (minimisation of the within-cluster prediction variations), it would be expected that the fit would be biased in favour of the homogeneity in purchase levels, as this is the criterion used in the selection/reproduction phase. However, one observes empirically that the two criteria improve simultaneously. For example, in appendix A, the curves obtained during the 10 analyses for the most complex case (*i.e.* 8 clusters) are provided. The R2 are shown on the *x*-axis and the within-cluster variation of prediction is shown on the *y*-axis. Across 70 segmentation analyses undertaken, only 5 led to a situation where the last gains made in relation to homogeneity were conducive to a lesser R2. Even then, the decrease in R2 remained very small (less than 0.1%) and there were no obstacle to returning to a previous solution with a greater R2 if need be. It appears therefore that a strong relationship exists between the minimisation of the individual error (assignment criterion) and the within-cluster prediction homogenisation (solution selection criterion).

The observed convergence times were consistently in favour of Glimmix which took on average three times less long than for Typren. In practice, with a Pentium 300Mhz, times recorded spanned from 1 to 3 minutes for Glimmix and 3 to 10 minutes for Typren. The hybridisation of the GA does therefore translate into a very marked acceleration which enables it to become operational and to surpass Simulated Annealing or classical GAs.

The means and standard deviations of the 6 criteria are provided in Table 2. The first figure of each case corresponds to the mean, the second to the standard deviation. Note that the regression without segmentation of the population provided a sum of squared errors of 385.61 and a R2 of 0.103.

TABLE 2. Results.

Solution size	Method	SSE fuzzy	R2 fuzzy	SSE hard	R2 hard	Dispers	Robust.
2	Glimmix	245.47 1.61	36.67 2.92	285.28 11.28	34.36 2.9	429.53 0.86	3.01
	Typren	116.28 6.89	72.96 0.16	116.28 6.89	72.96 0.16	141.25 0.58	2.96
3	Glimmix	133.59 36.10	60.22 12.56	154.96 54.05	51.41 9.86	415.62 23.08	3.50
	Typren	61.17 5.36	85.90 0.316	61.17 5.36	85.90 0.32	102.79 0.7	2.89
4	Glimmix	72.07 9.54	81.90 3.77	79.75 16.24	68.18 4.9	397.65 22.15	3.17
	Typren	35.77 12.02	90.83 0.80	35.77 12.02	90.83 0.80	81.42 4	2.62
5	Glimmix	69.30 14.44	83.50 3.28	72.40 14.11	72.14 3.12	412.90 25.02	2.55
	Typren	23.12 2.78	94.76 0.47	23.12 2.78	94.76 0.47	66.41 3.2	2.55
6	Glimmix	59.22 23.76	85.94 5.66	61.59 24.38	76.26 6.43	422.53 19.25	2.17
	Typren	17.13 3.491	96.02 0.812	17.13 3.49	96.02 0.81	59.28 1.9	2.31
7	Glimmix	52.87 16.77	87.66 4.15	54.26 17.87	79.81 4.89	426.49 7.17	2.56
	Typren	11.07 2.95	96.90 0.74	11.07 2.95	96.90 0.74	50.28 2.37	1.89
8	Glimmix	46.65 27.69	88.98 6.54	47.98 28.10	82.10 7.5	426.92 4.12	2.10
	Typren	10.89 3.34	96.67 1.19	10.89 3.34	96.67 1.19	46.05 2.20	1.77

Several results emerge from Table 2. First, and as expected, the predictivity increases as the population becomes segmented. Second, the predictivity afforded by Typren is consistently better than that of Glimmix, be it for the sum of the squared errors or for the R². The difference is particularly important for small size solutions. The difference in SSE is large for all sizes, whereas that for R² decreases with the solution size. Third, as indicated by the within-cluster dispersion criterion, the hardened Glimmix solution exhibits the classical defect of regressive segmentation: consumers in the same cluster may share an homogenous model but exhibit heterogenous purchase levels. One fundamental condition of segment operationalisation is therefore missing. On the other hand, Typren provides highly homogenous within-cluster prediction. This homogeneity increases with the size of the partition and does not precludes from a better predictivity.

Fourth, the fuzzy solutions of Glimmix are better than the hardened ones but the difference is less than expected. This suggests that the software succeeds in limiting the degree of confusion between clusters. The ex-post hardening process, whilst considered as sub-optimal and theoretically questionable, could be a way to clarify the interpretation of the fuzzy solutions generated by Glimmix. Fifth, the Glimmix solutions are generally more stable than the Typren solutions. In other words, the probability of achieving the same partition when starting with initially different parameters is greater in the case of Glimmix. It is therefore necessary to seek ways of making Typren more robust, while at the same time protecting its better predictivity. For example, in the assignment phase, an alternative and less varying function than the squared error may be considered.

Conclusions

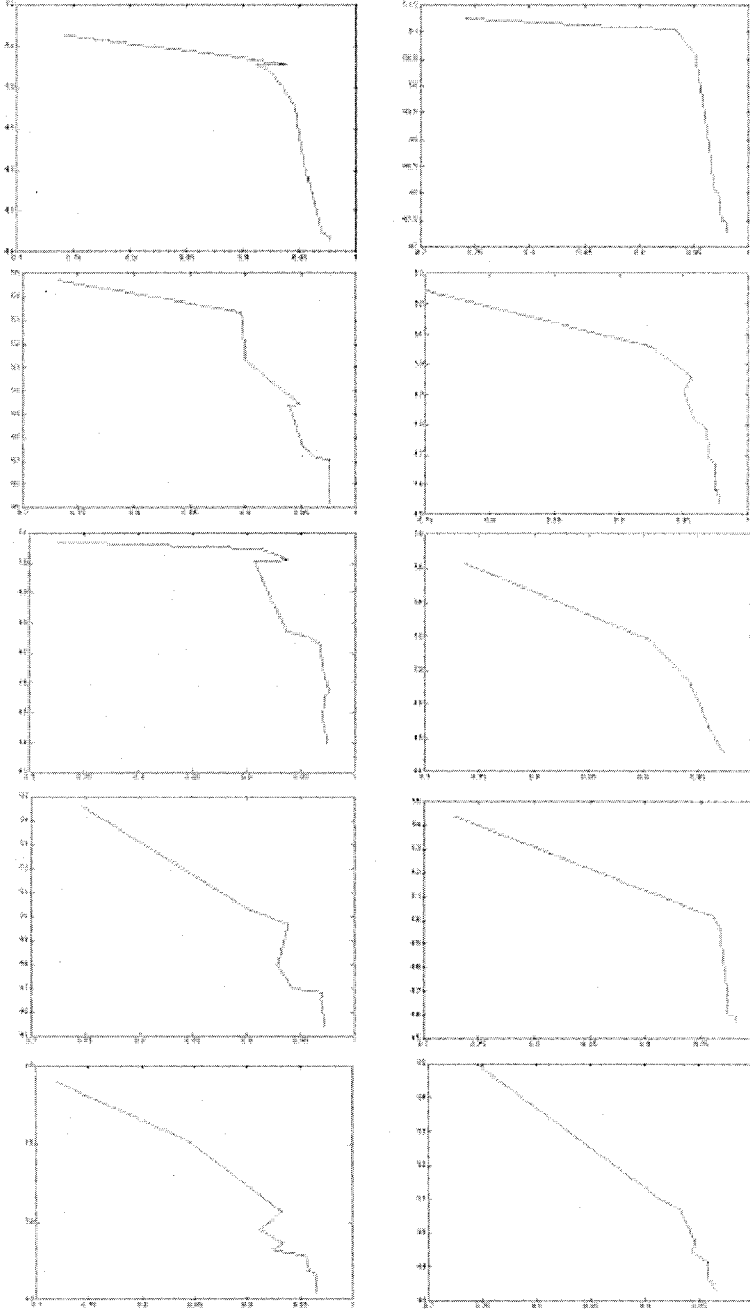
The focus has been on two points traditionally ignored about regressive segmentation methods:

- The need to achieve homogenous predictions for each segment. It is of little concern to the marketer that two consumers follow the same model if at the same time one consumes little and the other one lots.
- The remedial nature of fuzzy clustering which should only be undertaken when one is sure that a clear partition does not provide satisfactory results. Fuzzy clustering is not only difficult to understand but also generally inaccurately operationalised by artificially hardening the clusters as if they were alternatives when in fact their effects are joint.

We have shown that a bio-mimetic algorithm could resolve both of these issues and that the time required to do so was greatly reduced by using an hybrid of GA and Linear Regression. This result is, however, only indicative since it rests on a single data set and a limited type of research problems (predictors and prediction are both continuous variables). While this type of problem is common and often used as reference for research on Clusterwise Regression (Späth, 1979; 1981; 1982) and on Simulated Annealing (DeSarbo, Oliver and Rangaswamy, 1989), it still only represent a sub-part of the type of problems that Glimmix can potentially address.

The extension of Typren to other forms of regression is achievable. We are currently working on this as well as on improving its robustness without affecting its advantages in terms of predictability.

Appendix A: Change in R2 (x-axis) and within-cluster dispersions (y-axis) for the 10 eight-cluster solutions.



References

- Assael H., Roscoe A.M. (1976) Approaches to Market Segmentation Analysis, *Journal of Marketing* **40**, pp. 67-76.
- Aurifeille J.M. (1998) A biomimetic Clusterwise Regression Algorithm for Consumer Segmentation, in *Bio-mimetic Approaches in Management Science*, Aurifeille J.M., Deissenberg Ch. (Eds), Kluwer, Dordrecht, pp. 163-178.
- Aurifeille J.M., Bergmann J. (1995) Segmentation de marchés : une méthode fondée sur l'optimisation contrainte d'un réseau de neurones, *Actes des 27èmes Journées Internationales de Statistique*, A.S.U. S.S.F., Paris, pp. 42-47.
- Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood Estimation from incomplete data via the EM Algorithm, *Journal of the Royal Statistical Society* **39**, pp. 1-22.
- Davis L. (1991) *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- De Jong K.A. (1975) *An analysis of the behaviour of a class of genetic adaptive systems*. Dissertation Abstracts International **36** (10), 5140B.
- DeSarbo W.S., Cron W.L. (1988) A Maximum Likelihood Methodology for Clusterwise Linear Regression, *Journal of Classification* **5**, pp. 249-282.
- DeSarbo W.S., Howard D.J., Jedidi K. (1991) MULTICLUS : A New Method for Simultaneously Performing Multidimensional Scaling and Cluster Analysis, *Psychometrika* **56**, pp. 121-136.
- DeSarbo W.S., Oliver R.L., Rangaswamy A. (1989) A Simulated Annealing Methodology for Clusterwise Linear Regression, *Psychometrika* **54** (4), pp. 707-736.
- DeSarbo W.S., Wedel M., Vriens M., Ramaswamy V. (1992) Latent Class Metric Conjoint Analysis, *Marketing Letters* **3**, pp. 273-288.
- Doyle P., Fenwick I. (1975) The Pitfalls of AID Analysis, *Journal of Marketing Research* **12**, pp. 408-413.
- Doyle P., Hutchinson P. (1976) The Identification of Target Markets, *Decision Science* **7**, pp. 152-161.
- Goldberg D.E. (1991a) *Genetic Algorithms*, Addison-Wesley, New York.
- Goldberg D.E. (1991b) Real-coded Genetic Algorithms, virtual alphabets and blocking, *Complex Systems* **5**, pp. 139-167.
- Green P.E. and Krieger A.M. (1991) Segmenting Markets with Conjoint Analysis. *Journal of Marketing* **55**, pp. 20-31.
- Green P.E., Srinivasan V. (1978) Conjoint Analysis in Consumer Research: Issues and Outlook, *Journal of Consumer Research*, pp.103-123.
- Jöreskog K., Sörbom D. (1996) *Lisrel8*, Scientific Software International : Chicago.
- Kamakura W.A. (1988) A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments, *Journal of Marketing research* **25**, pp. 157-167.
- Kohonen T. (1987) *Self-Organization and Associative Memory*, Springer, Berlin.
- Laurent G., Kapferer J-N. (1985) Measuring Consumer Involvement Profiles, *Journal of Marketing Research* **12**, pp. 41-53.
- Lockshin L.S., MacIntosh G., Spawton A. (1997) Using product, brand and purchasing involvement for retail segmentation, *Journal of Retailing and Consumer Services* **4**, pp. 171-183.
- McLachlan D.L., Basford K.E. (1988) *Mixture Models*, Marcel Dekker, New York.
- McLachlan D.L., Johansson J.K. (1981) Market Segmentation with Multivariate AID, *Journal of Marketing* **45**, pp. 74-84.

- Mittal B., Lee M. (1989) A Causal Model of Consumer Involvement, *Journal of Economic Psychology* **10**, pp. 363-389.
- Ohanian R., Tashchian A. (1992) Consumers' Shopping Effort and Evaluation of Store Image Attributes: The role of Purchasing Involvement and Recreational Shopping Interest, *Journal of Applied Business Research* **8** (6), pp. 40-49.
- Renders J.-M. (1995) *Algorithmes génétiques et réseaux de neurones*, Hermès, Paris.
- Späth H. (1979) Algorithm 39: Clusterwise Linear Regression, *Computing* **22**, pp. 367-373.
- Späth H. (1981) Correction to Algorithm 39: Clusterwise Linear Regression, *Computing* **26**, p. 275.
- Späth H. (1982) Algorithm 48: A Fast Algorithm for Clusterwise Linear Regression, *Computing* **29**, pp. 175-181.
- Späth H. (1985) *Cluster Dissection and Analysis*, Wiley, New York.
- Wedel M. (1998) *GLIMMIX: Simultaneous estimation of latent classes and generalized models within each latent class*, User's Manual, version 1.0, ProGAMMA: Groningen.
- Wedel M. (1990) *Clusterwise regression and market segmentation: developments and application*, Doctorial Thesis, Holland (Groningen): Rijksuniversiteit.
- Wedel M., DeSarbo W.S. (1995) A mixture likelihood approach for generalized linear models, *Journal of Classification* **12**, pp. 21-55.
- Wedel M., Kamakura W.A. (1997) *Market Segmentation: Conceptual and Methodological Foundations*, Dordrecht: Kluwer.