Advanced analysis of branch and bound algorithms

Turkensteen, Marcel

Chapter 5

Balancing the Fit and Logistics Costs of Market Segments

5.1 Introduction

The key to success in the customer-oriented philosophy is the ability to understand customer needs (Kotler and Armstrong, 2006). These needs usually vary widely between consumers, a phenomenon which is called *consumer heterogeneity* (Allenby *et al.*, 1998). Since it is in general too expensive to serve each consumer separately, companies identify groups of customers, *segments*, with approximately the same relevant characteristics (Wedel and Kamakura, 2002). The process of designing segments is called *segmentation*; its purpose is to construct groups of consumers that are similar within groups and dissimilar between groups. Companies select their target segments and develop strategies to fulfill the needs of the customers in these segments (Smith, 1956), so that advantages over competitors are obtained. The requirement is that the marketing mixes, that the company deploys, fit individual consumer's preferences as much as possible.

Most segmentation studies maximize the fit of the segmentation. However, when consumers in the same segment are located far apart, the logistics costs for the organization can become very high. Steenkamp and Ter Hofstede (2002) report that "geographically dispersed consumer segments will often not allow profitable entry strategies to be pursued." This is in particular true "in industries where distribution costs constitute a large part of the total costs, such as in retailing and in industries dealing with perishable products" (Steenkamp and

Ter Hofstede, 2002). The importance of logistics costs is growing as a result of the trend of globalization (Steenkamp and Ter Hofstede, 2002). Although studies observe that transnational segments exist (Yavas *et al.*, 1992; Ter Hofstede *et al.*, 1999), many companies still target nations or groups of nations as separate segments; see for example Salah and Pervel Kathanis (1994).

When logistics costs are kept low, the resulting segmentations may fit consumer preferences worse than a standard segmentation strategy. However, this effect need not be large. In the segmentation study Ter Hofstede *et al.* (2002), it is found that, geographically close segments still represent consumer heterogeneity well.

These arguments form a motivation for constructing segments which may be less similar in their characteristics, but which are more similar in geographical location. We balance the maximization of the consumer benefits with the minimization of the logistics costs of supplying the products.

In most modern day segmentation studies, large populations of consumers are considered. Segments are identified, and an appropriate marketing mix is designed for each of them; see for example Helsen *et al.* (1993); Ter Hofstede *et al.* (1999); Bijmolt *et al.* (2004). The state-of-the-art tool for solving these problems is *mixture modeling* (Wedel and Kamakura, 1998). Mixture models use a separate statistical distribution for each segment, with which the preferences of individual consumers are modeled. Since the properties of individual consumers are represented in mixture models, consumer heterogeneity is well preserved. Similar types of segmentation approaches are *Bayesian modeling* (**?**) and *mixed models* (**?**). The results are not just averaged, but each individual observation in a segment is preserved in the statistical distribution. This conceptual advantage has recently given statistical methods, such as mixture modeling, the upper hand over other methods (Ter Hofstede *et al.*, 1999; Wedel and Kamakura, 1998).

We consider in this paper the special class of segmentations as described in Boone and Roehm (2002), where the company not only constructs marketing mixes for each chosen segment, but also assigns each individual consumer to a segment. Examples of this type of segmentation are also provided in Boone and Roehm (2002). The performance of mixture models depends strongly on the data. According to Wedel and Kamakura (1998), “mixture models are sensitive to poorly defined and separated segments, large numbers of local maxima, and

other violations of its underlying assumptions". Boone and Roehm (2002) find that the performance of mixture models is poor: many consumers are assigned incorrectly. This incorrect assignment leads to a reduction in the effectiveness of a segmentation strategy and squanders marketing resources. In this case, *hard non-overlapping* cluster methods are the most appropriate ones.

Hard cluster methods assign subjects deterministically to segments; non-overlapping methods assign each subject to exactly one segment (Punj and Stewart, 1983). Commonly used non-overlapping methods are *Ward's* algorithm (Ward, 1963) and the k*-means* algorithm (MacQueen, 1967). Cluster algorithms are usually heuristics; exact algorithms are hardly applicable to clustering problems, since they require so much solution time that only small instances can be solved (Du Merle *et al.*, 2000). Recently, there has been a major development in the application of *meta-heuristics* to segmentation problems, such as simulated annealing (Klein and Dubes, 1989; Brusco *et al.*, 2002), and artificial neural networks (Boone and Roehm, 2002). For an overview of hard non-overlapping clustering methods, we refer to Jain *et al.* (1999) and Mirkin and Muchnik (1998). The *NORMCLUS* cluster methods, presented in DeSarbo and Grisaffe (1998), enable the user to incorporate a broad set of possible constraints on segments, such as minimum size of segments and maximum acceptable difference between two subjects within a segment.

5.2 The logistics costs of segmentations

The segmentation decision and the decision on the design of the supply chain are usually taken independently, and until now, segmentation modeling and logistics modeling have been two disconnected fields of research. From Steenkamp and Ter Hofstede (2002), it can be concluded that large distances form the main cause of high logistics costs of a segmentation. This is a motivation for considering the logistics costs of segmentations.

When distances between consumers in the same segment are large, the logistics costs are high, in particular for perishable goods; see Steenkamp and Ter Hofstede (2002). Longer distances lead to longer lead times, more inventory, and more variability in the orders; see Nelson and Toledano (1979). The following figures illustrate the importance of logistics costs. In Davis (1990), it is found

that, on average, the cost of the physical distribution makes up about 22% of the total cost of a product. Recently, the GMA (Grocery Manufacturers Association) has found that transportation costs accounts for 62% of the logistics costs in the American grocery sector, and that transportation costs have been rising steadily (GMA, 2005). In this section, we identify the logistics costs associated with the choice for a segmentation.

We assume that for each segment a separate facility is established. This assumption appears to be realistic for our case study on retailing; the European Council on Applied Sciences and Engineering (Euro-CASE) has conducted research on supply chains in European retailing (Euro-CASE, 2001). The report observes that stores used to be supplied from a retailer-controlled Regional Distribution Centers (RDCs). Currently, there is a trend towards European Distribution Centers (EDCs), which are opened to establish Pan-European supply networks. This means that one central EDC supplies stores all over Europe, That is, if it is not too costly to transport the commodities over large distances. For some products, the logistics costs of usual transnational segmentations are too high. If Pan-European EDCs are established, the distances from those Pan-European facilities to individual stores are too large, leading to high transportation costs. RDCs are then usually maintained.

We introduce a measure for quantifying the logistics costs, namely the *Distance to a Central Facility (DCF) measure*. The logistics costs are estimated with the sum of the distances to a central facility in each segment. When the DCF measure is used, the underlying assumption is that a depot for each segment is built in a central position, and that each subject is served from this central facility. Consumers within each segment are then continuously replenished from the central location. The DCF measure is rooted in the *center of gravity* model, introduced in Francis and White (1974). The model locates all demand points on a two-dimensional plane; the weight of a demand point is the expected demand in that point. The virtual facility is then located in the center of gravity of all demand points. Note that the purpose of our model is not to establish optimal locations for facilities, but to estimate the logistics costs. In our measure, the logistics costs are proportional to the sum of the distances from all consumers to the location of the central facility. This is graphically depicted in Figure 5.1.

The following small example illustrates the DCF measure. Assume that there

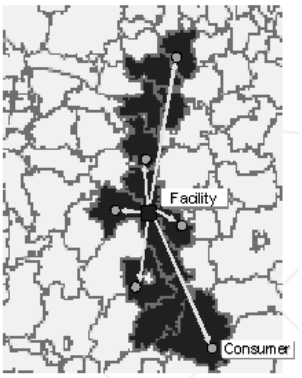

Figure 5.1. Example of DCF measure of logistics costs

are three consumers in a segment, all with equal demands. Assume that the consumers are located in a two-dimensional plane on the coordinates $(0, 0)$, $(2, 0)$ and $(1, 3)$, respectively. The central position is $(1, 1)$, so the the DCF measure is the sum of the distance between the consumer demand points and the central facility, being $\sqrt{2} + \sqrt{2} + 2 \approx 4.828$.

The following extensions can be made to the logistics costs model; see Krarup and Pruzan (1989).

- We assume that one central facility supplies all consumers in the corresponding segment. If more than one facility is needed, then we obtain an capacitated p-median problem, where p denotes the number of facilities; Mirchandani and Francis (1989).

- The DCF logistics costs measure places the facility on a central location. Other considerations for the location of the facility, listed in Kotler *et al.* (1993), are not taken into account.

- The DCF measure only consists of the transportation costs. The costs of fixed facilities, such as warehouses, are not included, because they are more or less independent of the chosen segmentation.

- We assume in our case study from Section 5.5 that the demand of consumers are more or less equal. If, however, reliable demand estimations are known, the facility is shifted towards consumers with larger demands.

An example of a reliable demand estimation is the All Commodity Value (ACV): the value of all commodities bought in a region.

5.3 The Budget Constraint Approach

We have seen that, due to high logistics costs, many companies resort to a segmentation strategy which takes countries or regions as segments (Salah and Pervel Kathanis, 1994; Steenkamp and Ter Hofstede, 2002). Such a traditional segmentation ignores the similarity of consumers across boundaries and the dissimilarity of consumers within boundaries found in, among others, Ter Hofstede *et al.* (2002) and Yavas *et al.* (1992). Is it possible to serve transnational segments well against reasonable logistics costs?

Ideally, the segmentation strategy with the highest expected profit level is selected, where profit is defined as the expected revenues of the segmentation minus the costs. However, direct profit maximization is not possible, because it is not known how much the sales increase when the fit of consumer preferences increases. Although it is well known that an increase in the fit of consumer preferences has a positive effect on sales (Simonson, 2005; Kumar and Petersen, 2005), none of the studies actually quantify this relationship.

We therefore choose to restrict the logistics costs. Within a given *budget* B, we try to find the best segmentation with the DCF logistics costs at most B. An alternative is the restriction on the maximum distance between two subjects in the same segment; see Section 5.8.

DeSarbo and Grisaffe (1998) present the NORMCLUS framework for solving clustering problems, with which several requirements on segments can be imposed. For example, to prevent an organization from serving small segments at a loss, a minimum can be imposed on the segment size, and the maximum distance of each consumer in a segment to a fixed location can be restricted. The objective function is flexible, so that it can accommodate the simultaneous optimization of multiple criteria. However, when the NORMCLUS framework is used to simultaneously maximize the fit and minimize the logistics costs of a segmentation, two problems arise. Firstly, the logistics costs constraint is imposed on the entire segmentation. This is a new and very specific type of constraint, requiring geographical information of each customer and the computation of

central locations of every solution considered in the search process. The second problem is that the weight of the logistics costs compared to the fit of the segmentation is hard to determine in advance. It is difficult, or even impossible, to know in advance which logistics costs budget leads to a segmentation with the highest profit. It is then better to vary the budget and compute a number of alternative segmentations.

To make the trade-off between lower logistics costs and an increase in the fit of consumer preferences, we introduce the *Budget Constraint Approach*. The approach constructs segmentations for different values of B, resulting in a large set of candidate segmentations. From this set, we then choose a few viable alternatives. The process is summarized in Figure 5.2.

The Budget Constraint Approach comprises two major steps:

Step 1: Construction of candidate segmentations;

Step 2: Reduction of the number of alternatives.

The set of candidate segmentations is formed as follows. Two input parameters jointly influence the quality of the solution values, namely the number of clusters K and the budget B. The Budget Constraint Approach solves a sequence of clustering problems for which the value of B is gradually increased. The solution obtained for a given value of B is used as starting solution for the next clustering problem in the sequence. The use of starting solutions from previous steps is only possible when the budget is gradually increased: a solution satisfying a budget B automatically satisfies every budget $B^* > B$. As a consequence of this procedure, no new starting solutions need to be computed. It is important to make an appropriate choice of the step size between two subsequent budgets. If the increase in B is too large, we take the risk that promising segmentation solutions are overlooked; if the increase in B is too gradual, the procedure becomes very time consuming. If the number of clusters is high, it is easier to construct clusters with relatively small B values and at the same time achieve a high level of homogeneity within the segments; see Section 5.4. Therefore, the clustering problem is not only solved for the number of segments obtained by a standard clustering method, but also for slightly bigger values of K.

The following clustering problem is solved for each selected combination of K and B. Define $x_{ik} = 1$ if subject i belongs to cluster k, and $x_{ik} = 0$ otherwise.

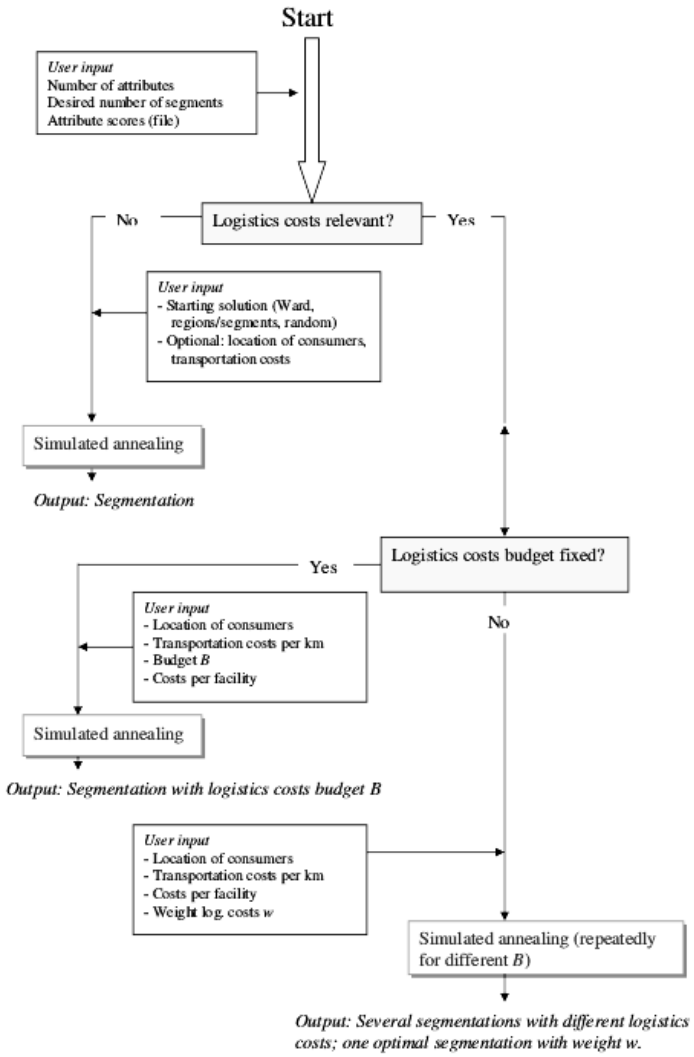

Figure 5.2. Flowchart of the Budget Constraint Approach

$$\begin{aligned}
\min \quad & \textstyle\sum_{i=1}^{N}\sum_{k=1}^{K} x_{ik}\|f_i - z_k\|^2 && && (5.3.1)\\
s.t. \quad & \textstyle\sum_{k=1}^{K} x_{ik} = 1 && i = 1,\ldots,N && (5.3.2)\\
& \textstyle\sum_{k=1}^{K}\sum_{i\in C_k} d_{ig_k} x_{ik} \le B && && \\
& x_{ik} \in \{0,1\} && i = 1,\ldots,N && \\
& && k = 1,\ldots,K && (5.3.3)
\end{aligned}$$

Here, f_i is the vector of attribute scores of subject i, and z_k is the vector of attribute scores of an average subject of cluster k. Moreover, d_{ij} denotes the geographical distance between two subjects i and j, and d_{ig_k} denotes the distance from subject i to the center of gravity of segment k. C_k denotes the set of subjects i that belong to cluster k in a given solution. The objective function, (5.3.1), is the commonly used *Minimum Sum-of-Squares Criterion (MSSC)* (Du Merle *et al.*, 2000; Milligan and Cooper, 1987), which minimizes the sum of the squared distances to an average subject of each segment. Constraint (5.3.2) requires that each consumer is assigned to exactly one segment, and constraint (5.3.3) requires that the total logistics costs do not exceed B.

Due to the constraints on the clusters, the problems belong to the class of so-called *constrained clustering problems*; see Everitt *et al.* (2001); DeSarbo and Grisaffe (1998). These problems may be more difficult to solve than unconstrained clustering problems, firstly, because it is hard to find a good starting solution. For example, Ward's algorithm is often not able to construct a starting solution at all, even when B is relatively large. Secondly, many moves become unavailable for small values of B, because those moves lead to clusters with logistics costs larger than B. If only improving moves are allowed, such as in k-means, then it is very likely that bad local optima are found for low values of B.

We solve the Budget Constraint clustering problems with *simulated annealing*. Wedel and Kamakura (1998, p.54) report that simulated annealing is an "especially appealing approach to overcome the local optimum problem". Annealing is the process in physics occurring when a substance is heated such that particles arrange themselves randomly. The temperature is lowered slowly, so that the particles stabilize and the desired structure emerges; see Van Laarhoven

and Aarts (1987). Simulated annealing in physics is the simulation of this annealing process. It evaluates a sequence of states; the step from one state to another is called a *move*. Kirkpatrick *et al.* (1983) describe the analogy between the annealing of a substance and optimization: initially the temperature of the cooling schedule is high, and many random perturbations of the solution at hand are allowed, even if they lead to moves towards worse solutions than the one at hand. When a promising subset of solutions is established, the temperature is lowered, meaning that a smaller number of perturbations is allowed. Finally, the freezing temperature is reached and a good or optimal solution is returned. Simulated annealing is applied to solve clustering problems in Klein and Dubes (1989) and Brusco *et al.* (2002), and also for many other problems; see Henderson *et al.* (2003) and the references therein.

The simulated annealing algorithm for solving clustering problems is described below. The algorithm proceeds through a sequence of cluster solutions. It randomly changes the cluster membership of one subject in the current solution. If the new cluster solution has a lower MSSC score, then the algorithm proceeds to this solution. Otherwise, if the MSSC score of the new solution is worse than the score of the current solution, the move to the new solution is made with a probability that depends on the magnitude of the deterioration and on a parameter value of the algorithm, called the *temperature*. Typically, the temperature is initially high, so that many deteriorating moves may be made. Later on in the process, the temperature is decreased. The simulated annealing algorithm described below is similar to the algorithm described in Brusco *et al.* (2002).

The algorithm makes moves from a current solution to solutions in the neighborhood in which the cluster membership of a single subject is altered. However, only some of these solutions in the neighborhood of a current solution satisfy the budget restriction. The algorithm computes the logistics costs of a new solution S and compares it to the user-defined budget B. The algorithm can also accommodate other requirements on segments, such as a minimum segment size.

After having obtained a set of cluster solutions for various values of K and B, we enter the second step of the approach: the reduction of the number of alternatives. We try to find solutions which combine a good score on the fit of the data with a small number of clusters and low logistics costs. The relative importance of each of these factors determines which segmentation is the best

Algorithm 2 Simulated annealing approach for budget-constrained clustering

INPUT

S	Segmentation;
$c(S)$	MSSC cost of a segmentation S;
$l(S)$	logistics costs of a segmentation S;
B	logistics costs budget;
K	number of segments.

PARAMETERS

T_0	initial temperature;
α	cooling parameter;
$STABLE$	number of tries before stability at temperature T is achieved;
T_f	temperature at which the algorithm is frozen.

MAIN ALGORITHM

Generate initial cluster solution S^* such that $l(S^*) \leq B$;
$T := T_0$;
repeat
  count := 0;
  repeat
    randomly select cluster solution $S := findMove(S^*)$;
    $\delta = c(S^*) - c(S)$;
    if $\delta \geq 0$
      $S^* := S$;
    else
      rnd := random number from $U(0, 1)$;
      if $rnd < \exp\{\frac{\delta}{T}\}$
        $S^* := S$;
    count := count + 1;
  until count = STABLE;
  $T := \alpha \times T$;
until $T < T_f$;

FUNCTION findMove(S^*)
feasible := **false**;
while not feasible
  S := perturb S^* by randomly changing the membership of a randomly chosen subject n;
  if $l(S) \leq B$ feasible := **true**;
return S.

OUTPUT

(Almost) optimal cluster solution S^* with cost value $c(S^*)$.

one for the organization. Unfortunately, it is very difficult to translate the fit of the clustering solution into the corresponding expected profit when a segmentation is based on preferences and perceptions. This is true for the case study from Section 5.5. As a consequence, we cannot attach fixed weights to the objectives and choose an 'optimal' segmentation.

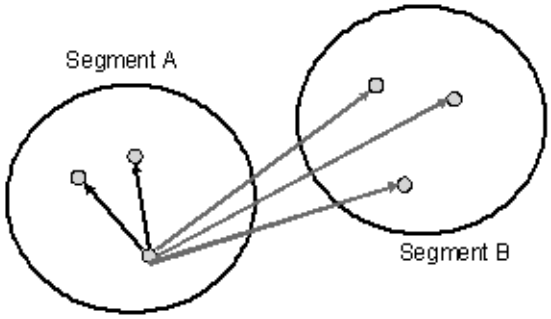

Figure 5.3. The silhouette width of a subject

We evaluate the alternatives on two criteria: the silhouette width measure of fit, and the logistics costs. The *silhouette width* of subject i divides the average distance to other subjects in the segment to which i belongs to the average distance of subject i to subjects in other segments; see Rousseeuw (1987). The distance in this context is between the attribute scores of two subjects, and not the geographical distance. The silhouette width ranges between -1, the worst possible score, and +1, the best possible score.

The silhouette width is computed as follows. Assume that subject i belongs to segment A. Then, let $a(i)$ be the average distance of i to all other subjects in A. Moreover, let $d(i, B)$ denote the average distance of i to all subjects in cluster B, and let $b(i) = \min_{B \neq A} d(i, B)$. Then the silhouette width $s(i)$ of subject i is:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5.3.4}$$

The *silhouette width of a segmentation* is the average silhouette width over all consumers. Other measures, such as the value of the maximum likelihood of mixture models and the MSSC score, are either non-increasing or non-decreasing with the number of clusters, i.e., the larger the values of K, the better the score

of the cluster solution. The value of the silhouette width, however, does not necessarily improve with increasing values of K. Rousseeuw (1987) claims that the silhouette width attains its highest value for the number of clusters that describes the data in the best way, i.e., the largest similarity within segments and the smallest dissimilarity between segments. Some refined measures of the classification likelihood, such as the so-called AIC3, BIC, and CAIC, share this favorable property; see for example **?**.

An important decision in segmentation studies is the choice of the number of segments. An increase in the number of segments generally leads to better fitting segments, but this is at the expense of higher costs: new marketing strategies must be developed and segment facilities need to be set up. Methods for determining the number of segments are described in Milligan and Cooper (1987). A usual method is *visual inspection*: the number of segments is chosen in such a way that increasing the number of segments does not improve the the quality of the cluster solution substantially. In our approach, the best number of segments needs to be determined for each considered logistics costs budget. Instead of visual inspection, we choose the value of K in such a way that the resulting segmentation achieves the highest silhouette width. We assume that if the silhouette width increases when a new segment is added, the profit gained, as a result of the increase in the fit of consumer preferences, is sufficient to outweigh the additional costs of new segments.

After determining the best number of segments for each logistics costs budget, we vary the relative weight w of the logistics costs compared to the silhouette width. The final score of an alternative S is denoted by $F(S; w)$ and defined as:

$$F(S; w) := (1 - silhouette\ width) + w \times logistics\ costs. \tag{5.3.5}$$

The segmentation S^* is called *efficient* if, for some w, the value of $F(S^*; w)$ is lower than the values of $F(S; w)$ for all other considered segmentations $S \neq S^*$. The Budget Constraint Approach returns the set of efficient segmentations. The user can then specify his or her own importance of the logistics costs, and make the trade-off between an acceptable logistics costs level and good fit.

Table 5.1. Random instances by Milligan (1985)

Name	Type of clustering problem
TYPE 1	Error free data
TYPE 2	20% outliers
TYPE 3	40% outliers
TYPE 4	Error perturbed coordinates (low level)
TYPE 5	Error perturbed coordinates (high level)
TYPE 6	Added random noise dimension
TYPE 7	Two added random noise dimensions
TYPE 9	Standardized coordinates
TYPE 11	Random noise data without cluster structure

5.4 Experiments with randomly generated cluster instances

In this section, we evaluate the simulated annealing approach of Section 5.3 on randomly generated cluster instances. We also examine for which types of clustering problems the Budget Constraint Approach is able to obtain well-fitting segmentations with moderate logistics costs.

We use the randomly generated artificial instances from Milligan (1985). The attribute scores are drawn from normal distributions, but many instances have nasty properties, or 'errors'. The included errors are outliers, additional noise variables, unequal cluster sizes and erroneous observations; see Table 5.1. These properties, which are also encountered in practical situations, make cluster algorithms less accurate. The instances from Milligan (1985) have been used in comparative studies of algorithms, such as Boone and Roehm (2002) and Milligan and Cooper (1987).

The true clustering solutions of the instances of Milligan (1985) are known, so that the solutions obtained by any algorithm can be compared to the true cluster solutions. The Adjusted Rand Index (ARI), introduced in Hubert and Arabie (1985), measures how similar a given cluster solution is compared to the 'true' cluster solution. More precisely, it determines the fraction of pairs of subjects which are in the same cluster in both the 'true' and the given clustering. A value of 1 indicates that the algorithm has returned the true cluster solution; a value of 0 means that the cluster solutions at hand has the quality of an average random classification. Milligan and Cooper (1986) report that the Adjusted Rand Index is the most accurate criterion for measuring solution quality.

The Adjusted Rand Index (ARI) is defined as follows. Suppose there is a set of objects O and there are two partitions $L = L_1, \ldots, L_R$ and $M =$

$M_1, \ldots, M_C$ of O such that $\bigcap L_i \cap L_j = M_i \cap M_j = \emptyset$ for each pair of subsets $L_i \neq L_j$ of L and $M_i \neq M_j$ of M. So $O = L_1 \cup \ldots \cup L_R = M_1 \cup \ldots \cup M_R$. Assume that $|O| = n$. Construct the following $R \times C$ matrix, in which element n_{ij} denotes the number of elements which are both in class L_i and in class M_j. Furthermore, let $n_{i.}$ and $n_{.j}$ denote the row and column totals of this matrix, i.e., the number of elements in class L_i and M_j, respectively. Then:

$$ARI = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}] - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}]/\binom{n}{2}} \tag{5.4.1}$$

In Boone and Roehm (2002), experiments are performed on the instance types 1, 2, 4, 6, and 7 from Table 5.1 with k-means, artificial neural networks and mixture models. The k-means method (with Ward's algorithm to generate initial solutions) and artificial neural networks perform approximately equally well. Mixture models achieve lower average scores on all tested types of instances. Since k-means turns out to be an effective cluster method in Boone and Roehm (2002), we compare simulated annealing to this method.

Table 5.2. Average Adjusted Rand Index of simulated annealing and k-means

	SA Ward	k-means Ward	SA Random	k-means Random
Type 1	0.999	0.999	0.936	0.512
Type 2	0.714	0.618	0.698	0.399
Type 3	0.530	0.519	0.521	0.434
Type 4	0.988	0.987	0.920	0.487
Type 5	0.896	0.861	0.845	0.418
Type 6	0.711	0.694	0.620	0.400
Type 9	0.986	0.986	0.936	0.791

In Table 5.2, the adjusted Rand indices of simulated annealing (SA) and k-means are shown, both with randomly generated starting solutions and with starting solutions obtained with Ward's algorithm. When Ward's algorithm is used to generate starting solutions, simulated annealing achieves only marginally better results than k-means for all types of instances. The role of Ward's algorithm is very important here: it constructs good starting solutions and even retrieves the true clustering solutions for many test instances. Our simulated annealing algorithm requires longer solution times to obtain the improvements: 0.1 to 4.5

Table 5.3. Results for random cluster instances with increasing budget

		Logistics costs level			
Instance	Comment	Measure	Low	Intermediate	High
Type 1	Normal,	ARI	0.033	0.636	0.937
	random	Log. costs	2211	3467	3643
Type 3	40% outliers	ARI	0.004	0.189	0.536
		Log. costs	4310	5736	6830
Type 4	Coordinates	ARI	0.033	0.626	0.941
	perturbed	Log. costs	2210	3457	3625
Type 5	Coordinates	ARI	0.032	0.559	0.848
	perturbed	Log. costs	2207	3446	3611
Type 6	Added noise	ARI	0.023	0.449	0.628
	dimension	Log. costs	2218	3454	3611
Type 9	Standardized	ARI	0.028	0.900	0.954
	data	Log. costs	2268	3640	4567

seconds, where the average is 2.225 seconds and the median 2.418 seconds. On the other hand, the solution times of k-means are usually within 0.1 second. However, when the starting solutions are randomly generated, simulated annealing achieves clearly better solutions than k-means. These findings confirm the conclusion in Wedel and Kamakura (1998) that simulated annealing is less dependent on the starting solution. We also find that, if the value of B is small, Ward's algorithm is often not able to construct a feasible solution.

To include the logistics costs in the experiments, we disperse the locations of the subjects in a 100 by 100 plane and assume that the logistics costs are proportional to the distances in this plane. An initial solution is obtained by grouping closely located subjects into segments. The logistics costs budget is then gradually increased until an unconstrained segmentation is obtained.

The Budget Constraint Approach is particularly useful when good segmentations with relatively low logistics costs exist. Is this true for the randomly generated instances? Of each type reported in Table 5.3 and for each logistics costs level, 48 cluster instances are solved. The fit decreases when the logistics costs budget is cut down. However, for the standardized random data of type 9, the solution with intermediate logistics costs is almost as good as the unconstrained solution, but with much lower logistics costs.

In the experiments on random instances, we use the Adjusted Rand Index to compare the classifications obtained by the tested algorithms with a 'true'

classification. Such a true classification is not available in the case study that we discuss in Section 5.6. Since the Adjusted Rand Index cannot be used, we suggest to use the silhouette width instead. The silhouette width (SW) and the Adjusted Rand Index (ARI) appear to be positively and linearly related; see Figure 5.4. The linear regression explaining the silhouette width with the Adjusted Rand Index achieves an R^2 of 0.861, meaning that 86% of the variation in the Adjusted Rand Index values can be explained by changes in the silhouette width. The silhouette width appears to provide a worse estimation if the Adjusted Rand Index is either very low or very high. Nevertheless, the result indicates that, when the true clustering is not available and the Adjusted Rand Index cannot be used, the silhouette width by Rousseeuw (1987) is a reasonable alternative.

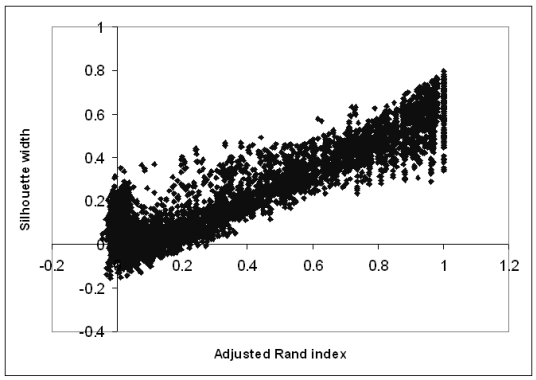

Figure 5.4. Adjusted Rand Index and silhouette width: a comparison

We explore two sources of variation in the results: the *spatial contiguity* and the *number of clusters*. *Spatial contiguity* is the degree to which geographically close subjects have a mutual influence on each other's attribute scores; see, for example, Ter Hofstede *et al.* (2002) and the references therein. We include spatial contiguity in our models as follows. When spatial contiguity is present in the data, subjects in the same segment are likely to be located close to each other. A seed point is determined for each segment. From this seed point, a path through the 100 by 100 plane is formed by adding random numbers drawn from $U[-a, a]$ to the current x and y-coordinates on which the subjects in the segment are placed, where the parameter $a \in \mathbb{N}$ determines the level of spatial contiguity. When the path reaches the boundary of the plane, the orientation is reversed. If the value of a is small, subjects in the same segment are located close to each

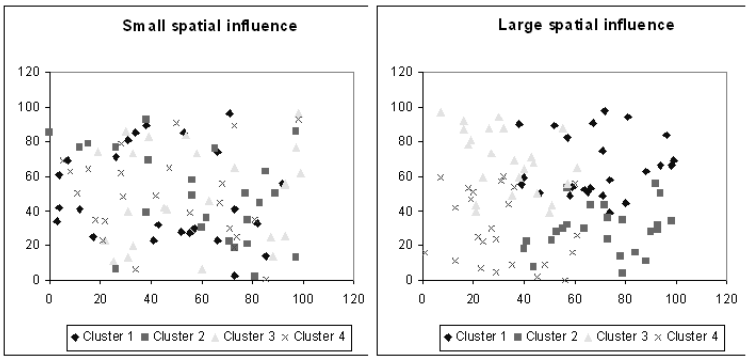

Figure 5.5. Subjects with small and large degrees of spatial contiguity

other. On the other hand, if the value of a is large, the subjects in the same segment are more or less randomly dispersed in the plane. Figure 5.5 shows typical planes with small and large simulated spatial contiguities. In case of large spatial contiguity, the subjects in the same cluster are located close to each other in the plane.

Figure 5.6 presents the effects of varying spatial contiguity. When the budget B is small, the fit of the cluster solution remains relatively high when the consumer preferences show a high degree of spatial contiguity.

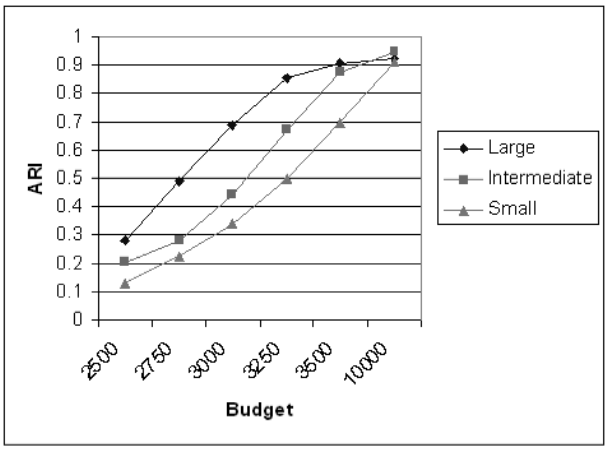

Figure 5.6. Average cluster solution quality for large, medium and small degrees of spatial contiguity

The second factor taken into account is the *number of clusters*. The number

of clusters in the data set of Milligan (1985) varies between 2 and 5. Figure 5.7 shows the quality of the solutions at different values of the budget B. The solution value, measured with the Adjusted Rand Index (ARI), decreases slightly with B when the number of segments is relatively large, but it declines strongly when $K = 2$. Obviously, when there are only a few clusters, remote subjects are forced into a cluster with different other subjects in order to keep logistics costs low.

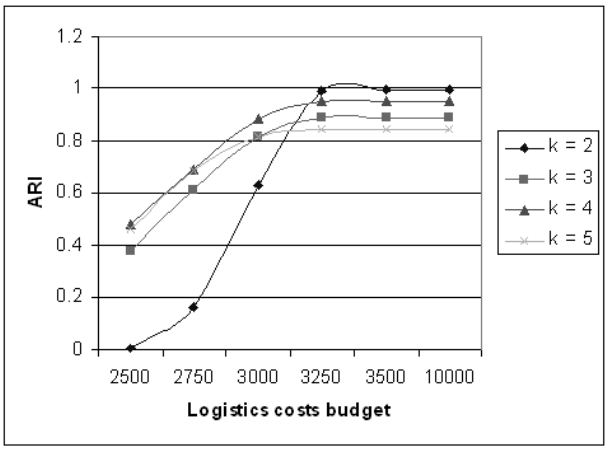

Figure 5.7. Average cluster solution quality and the number of clusters K

To summarize, the results of our experiments indicate that simulated annealing is a suitable method for solving clustering problems, in particular when a good starting solution is not available. Moreover, the Budget Constraint Approach presented in the previous section makes the trade-off between logistics costs and fit of segmentations. This trade-off is worthwhile when a decrease in logistics costs is associated with a slow decrease in fit of consumer preferences. This appears to occur most clearly for the random instances with standardized data from Milligan and Cooper (1986), when the number of segments should not be too small, and when spatial contiguity is present in the data.

5.5 Case study: European meat outlet segmentation

In this section, we describe a segmentation study on European meat outlets. The case study is used as an example of segmentation studies encountered by in-

ternational retail chains which adjust their stores to regional differences. The stores are tailored to the specific characteristics of the region in which the store is located. For each homogeneous group of regions, a separate retail formula is developed.

In order to determine how similar shops in different regions are, an international market research agency has executed a large survey on consumer behavior; see Ter Hofstede *et al.* (2002). Stores can obtain a distinctive position through the development of a particular *store image*. Important store image attributes are, for example, pricing, assortment, service, atmosphere, and quality. However, consumers in different regions may place a different relative importance on these attributes (Ter Hofstede *et al.*, 2002).

The results are obtained by sending mail questionnaires to members of a script panel in seven countries within the European Union, namely Germany, the Netherlands, Belgium, France, Spain, Portugal, and Italy. Store image measures are obtained, where for each respondent data are obtained for the store most frequently visited. The attributes are: price, quality, service, atmosphere, distance, and variety in meat. These six attributes are used to predict the general opinion about the store. For each of those six attributes and for the general opinion about the store, the respondent gives an opinion on a scale from 1 to 7. A separate regression is then carried out for each region in the data set, and the resulting regression parameters are used to estimate the relative importance of each attribute. If the number of respondents in a region is too small for a viable regression, the region borrows its characteristics from neighboring regions. The subjects of the segmentation are 123 so-called NUTS-2 regions, depicted in Figure 5.8. In order to compute the geographical distances in the case study, we use central points of regions, the *centroids*.

In Ter Hofstede *et al.* (2002), several mixture modeling approaches are implemented and compared. It is found that the so-called *spatial contiguity* model achieves the best fit of the data. In this model, the probability p_{ik} of region i belonging to segment k is not only influenced by the neighboring regions, but also by regions further away. The influence diminishes as the distance between a pair of regions is larger.

Logistics costs can play a key part in store segmentations. If the locations of shops within a selected segment are very dispersed, then supplying these shops

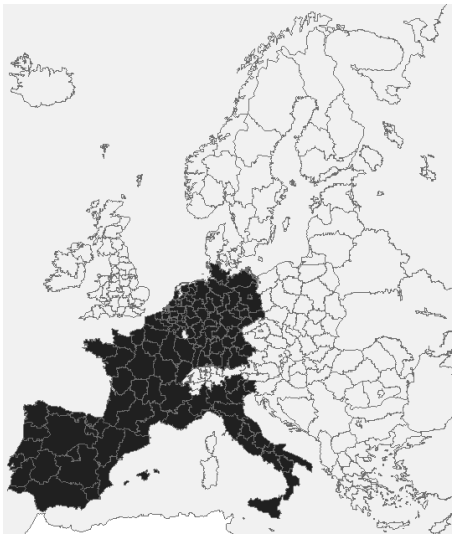

Figure 5.8. European regions in the data set

becomes expensive. Ter Hofstede *et al.* (2002) obtain clusters of regions that may be connected, but still stretch out over large sections of Europe, so with high logistics costs.

The mixture modeling approach is a suitable tool to determine what types of consumers are living in a region, and then offer multiple marketing mixes in each region in order to serve these segments. On the other hand, when only one retail formula can be offered in each region and an undifferentiated marketing approach within each region is chosen, Boone and Roehm (2002) find that hard clustering approaches, which assign a subject to a single segment, are more appropriate.

5.6 Computational experiments on the case study data

In this section, we construct segmentations for the case study discussed in the previous section. We use the Budget Constraint Approach from Section 5.3. We compare the resulting strategies with a mixture modeling approach, an unconstrained hard clustering approach, and the countries-as-segments segmentation strategy. Recall that the segmentation based are perceptions on price, quality, service, atmosphere, distance, and variety in meat.

The segmentations are evaluated with the following criteria:

Customer heterogeneity. The fit of customer heterogeneity, formed by the homogeneity of consumers within and the heterogeneity between segments, is measured with the Minimum Sum-of-Squares Criterium and the silhouette width; see Section 5.3. The MSSC is the optimization criterion for our hard cluster algorithm. Obviously, these methods are likely to achieve higher scores than mixture models, which maximize the classification likelihood. When a likelihood measure is taken instead, mixture models perform best; see e.g. Ter Hofstede *et al.* (1999). The silhouette width is a more neutral measure, similar to the Adjusted Rand Index from Section 5.4.

Number of segments. If possible, the number of segments in a segmentation should be kept small. For a large number of segments, many facilities need to be set up and new marketing strategies must be designed.

Unique characteristics of segments. For every segmentation solution, the attribute scores of all segments are evaluated. A segmentation performs well if each segment achieves a high score on one or more unique attributes (Ter Hofstede *et al.*, 1999). These unique attributes can be used to target the segments. For example, it is relatively easy to develop a marketing strategy for a segment that is very sensitive to price changes, or for a segment that appreciates quality highly.

Logistics costs. The logistics costs of each segmentation are estimated with the Distance to Central Facility (DCF) measure from Section 5.2. The sum of the distances from each region to its central segment facility is computed.

The first three criteria are also used in Ter Hofstede *et al.* (1999),where k-means and mixture modeling are compared.

Our first benchmark segmentation takes the countries as segments. Belgium and the Netherlands are joined into one segment, as well as Spain and Portugal. France, Germany and Italy are served as separate segments. The MSSC score of this segmentation is 22.96; the silhouette width is -0.0009. This means that much improvement in the solution quality can be achieved by transferring regions to

other segments. On the other hand, the logistics costs, measured as the sum of the distances of each region to its fictitious central segment facility, amount to 26813 km.

The second method is the standard hard clustering approach, consisting of a hierarchical method in the first stage and a non-hierarchical method in the second stage. Ward's method (Ward, 1963) is used to obtain a good initial solution and the number of clusters, namely $K = 4$. The initial solution is improved using simulated annealing (Kirkpatrick *et al.*, 1983), which we choose instead of the usual k-means approach. We obtain the solution depicted in Figure 5.9; the value of the MSSC is 12.16, and the DCF measure of logistics costs is 71552 km. The silhouette width score is 0.409. The longest distance between two regions in a single segment is 2326 km between the centroids of the regions Mecklenburg-Vorpommern in Northeast Germany and Algarve. This is the longest possible distance between any pair of regions in the data set.

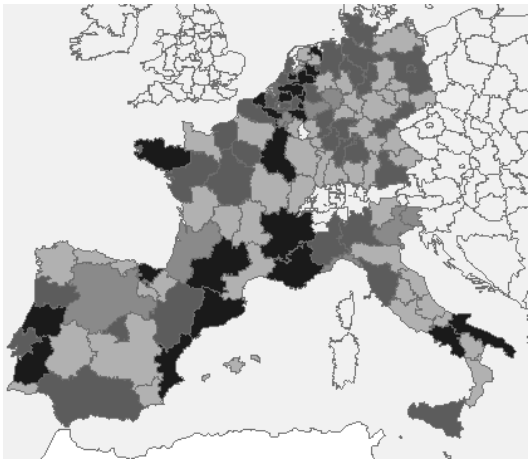

Figure 5.9. Solution obtained with unconstrained hard clustering

The *mixture modeling* is performed in Latent Gold (Vermunt and Magidson, 2003). The segments are estimated with normal distributions with different averages, but equal standard deviations. The number of segments in mixture modeling is usually determined with the AIC3 criterion; see Andrews and Currim (2003). In our case, the best fitting segmentation is obtained for $K = 4$. The posterior probabilities are rounded off, meaning that each region i is assigned to a retail formula k for which the probability p_{ik} is the highest. This

is a heuristic method, but the highest posterior probabilities are close to 1 for most regions. The resulting segmentation, depicted in Figure 5.10, has an MSSC score of 23.56, and the total distance is 68622 km. The silhouette width score is -0.058, which is even worse than the score of the countries-as-segments strategy. The mixture model leads to a very different segmentation from the unconstrained cluster solution and it achieves a very poor score on our fit criteria. This finding confirms the results from Boone and Roehm (2002).

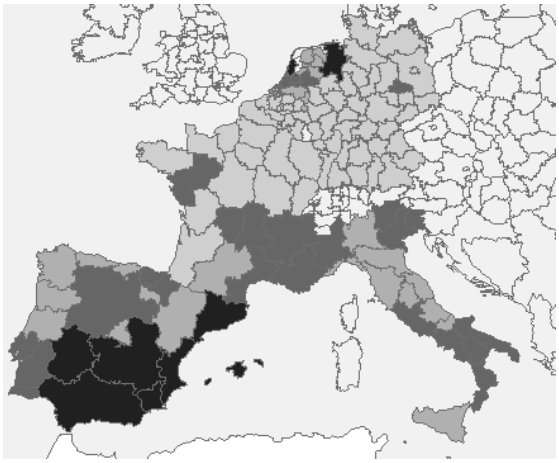

Figure 5.10. Solution obtained with mixture modelling

Next, we apply the Budget Constraint Approach from Section 5.3. The silhouette widths for segmentations with several values of B and K are shown in the surface plot from Figure 5.11 and in Table 5.4. The results indicate that the fit remains on a high level for smaller B vales when the value of K is large, namely $K = 6$ and $K = 7$.

In this segmentation study, the four segmentation strategies reported in Table 5.5 are efficient; see Section 5.3. If the logistics costs are the main determinant profits, then the countries-as-segments strategies is optimal. When the importance weights of the logistics costs w is increased, the segmentation with relatively low importance of the logistics costs consists of seven segments. This is option 1, and it is presented in Figure 5.12. The MSSC score is 9.52, the silhouette width is 0.433 and the logistics costs amount to 64219.

When w is increased, option 2 becomes the best alternative, with $K = 6$ and $B = 55000$. Option 2 combines large parts of Spain and Portugal into a

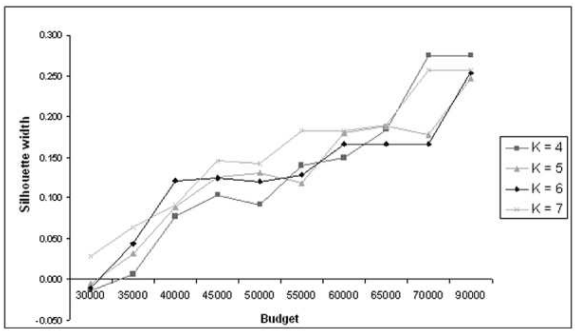

Figure 5.11. Fit of segmentations for various K and B

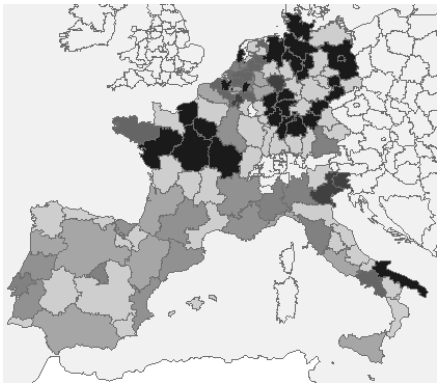

Figure 5.12. Segmentation option 1

Table 5.4. Silhouette width of segmentations for various values of K and B

	Number of clusters				
B	4	5	6	7	Best K
30000	0.006	0.028	-0.007	0.028	$K = 5$
35000	0.047	0.017	0.008	0.087	$K = 7$
40000	0.067	0.097	0.169	0.146	$K = 6$
45000	0.144	0.177	0.175	0.179	$K = 7$
50000	0.173	0.232	0.142	0.253	$K = 7$
55000	0.183	0.194	0.341	0.253	$K = 6$
60000	0.074	0.194	0.341	0.351	$K = 7$
65000	0.231	0.310	0.341	0.433	$K = 7$
70000	0.393	0.432	0.391	0.403	$K = 5$
Unconstr.	0.409	0.432	0.391	0.403	$K = 5$

Table 5.5. Segmentation alternatives based on silhouette widths

Option #	K	B	MSSC score	Silh. width	Log. cost	Relative weight log. costs
Option 1	7	65000	9.53	0.4329	64219	$w < 9.35$
Option 2	6	55000	11.40	0.3411	54399	$9.35 \leq w \leq 11.53$
Option 3	6	40000	14.47	0.029	39438	$11.53 \leq w \leq 14.11$
Country/segment	5	-	22.96	-0.009	26813	$w > 14.11$

single segment. The regions in this area are sufficiently similar to be combined into separate segments without losing much fit. The largest distance between two regions in the same segment is 1429 kilometers, the logistics costs amount to 54399, the MSSC score is 11.40, and the silhouette width is 0.341.

Option 3 also takes Southern Italy into a separate segment. In France, Germany and the low countries, the deviation from the traditional countries-as-segments is larger. The MSSC score is 14.47 and the silhouette width is 0.029. The logistics costs level is 39438.

Figure 5.15 summarizes the scores of the segmentations on the fit and the logistics costs. The fit of the segmentation is measured with the silhouette width and the logistics costs are measured with the distance to central location-measure. A fictitious ideal segmentation is located close to the right bottom corner of the figure.

Segmentations are also evaluated on the degree to which a segment has unique high scores on one or more characteristics. In Table 5.6, the segments averages of all segmentations are reported. In the hard unconstrained segmentation and in option 1, most segments are clearly reachable by one or more character-

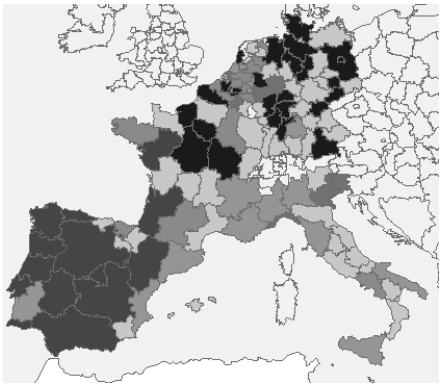

Figure 5.13. Segmentation option 2

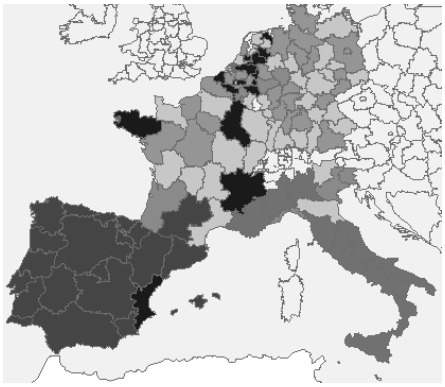

Figure 5.14. Segmentation option 3

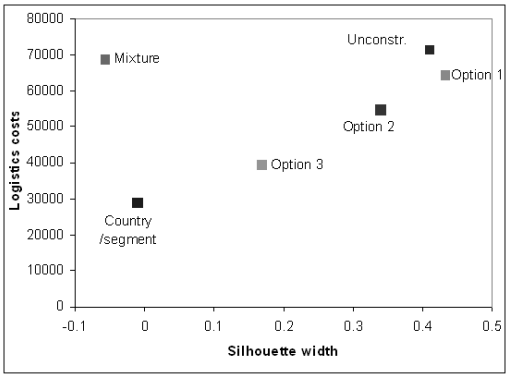

Figure 5.15. Comparison of logistics costs and fit of segmentation strategies

Table 5.6. Segment averages of different segmentations

	Attributes					
Segment	Price	Quality	Service	Atmosphere	Variety	Distance
			Countries as segments			
1	0.0360	0.1218	0.2963	0.1739	0.0850	0.2341
2	0.1047	0.1622	0.5319	0.0492	0.0747	0.1932
3	-0.0809	0.0191	0.2612	0.4240	0.1221	0.1142
4	-0.0272	0.4687	0.1724	0.3298	-0.0252	0.0301
5	-0.3024	0.1602	0.0485	0.5313	-0.0251	0.3451
			Hard unconstrained segmentation (option 1)			
1	0.0431	0.0404	0.0461	0.0516	0.0445	0.0427
2	-0.0006	0.1066	-0.3268	0.7601	0.0350	0.2379
3	0.0985	0.2080	0.1883	0.2614	0.1198	0.2431
4	-0.0040	0.0195	0.5395	0.0935	0.0605	0.1830
			Mixture model segmentation			
1	0.0598	0.1043	0.1668	0.1890	0.0638	0.1187
2	0.0401	0.0635	0.1296	0.1516	0.0689	0.1280
3	-0.0091	0.0518	0.1653	0.1007	0.0509	0.1816
4	0.0960	0.1531	0.1479	0.2123	0.1002	0.1713
			Option 1: segmentation with high logistics costs			
1	0.0362	0.0362	0.0362	0.0363	0.0362	0.0362
2	-0.0460	0.1777	-0.0605	0.3615	0.0905	0.3664
3	0.0279	-0.0969	0.6283	0.2003	0.0561	0.1439
4	-0.0998	0.0726	0.2607	0.3301	0.0228	0.1677
5	0.0238	0.1453	0.5069	-0.0661	0.0529	0.2563
6	0.0410	0.0390	-0.4103	0.9893	-0.0498	0.1195
7	0.1817	0.2047	0.2284	0.2175	0.1630	0.1725
			Option 2: segmentation with intermediate logistics costs			
1	0.0318	0.0396	0.0383	0.0414	0.0376	0.0416
2	-0.0309	-0.0785	0.4075	0.3493	0.0424	0.1842
3	0.0664	0.1475	0.5400	-0.0289	0.0751	0.1649
4	0.0022	0.0782	-0.3320	0.8355	-0.0348	0.2152
5	-0.0115	0.1523	0.0207	0.2154	0.0959	0.1821
6	0.1630	0.2213	0.1931	0.2171	0.1496	0.2408
			Option 3: segmentation with low logistics costs			
1	0.0410	0.0398	0.0437	0.0512	0.0440	0.0473
2	-0.0360	0.0138	0.4967	0.1043	0.0428	0.2359
3	0.1592	0.2207	0.2040	0.2209	0.1533	0.1956
4	0.0133	0.0382	0.1669	0.1740	0.0224	0.1228
5	0.0009	0.0822	0.1656	0.1384	0.0498	0.1194
6	-0.0040	0.0884	-0.3404	0.7703	0.0346	0.2587

istics. For example, segment 6 in option 1 values 'atmosphere' very high, and segment 4 from the unconstrained segmentation is very sensitive to service. In option 3, the segments 4 and 5 achieve similar scores, but segment 4 is located in Southern Italy, and segment 5 in Spain and Portugal. The difference in location justifies different retail formulas. The differences between the characteristics of the segments are smaller for mixture modeling than for hard clustering. This result implies that the targeting of segments is difficult when the mixture modeling segmentation strategy is chosen.

The Budget Constraint Approach provides the user with intermediate solutions with moderate logistics costs and a reasonable fit of consumer preferences, in addition to the countries-as-segments segmentation and the unconstrained segmentation. The consequence is that transnational segments can be served with an international segmentation with moderate logistics costs.

5.7 Reliability of simulated annealing

In Section 5.3, a simulated annealing approach for clustering is presented, and in Section 5.4 and 5.6, the approach is used to solve market segmentation problems with promising results. However, simulated annealing is a *randomized* search process: a deteriorating move to a new solution is made with probability $p(t)$, depending on the current temperature t, and the selection of the next trial move is done randomly. The final outcome depends on the realization of random variables, meaning that different cluster solutions may be obtained in different runs. In this section, we measure the variability in the quality of the resulting cluster solutions. If the variability is large, multiple runs should be carried out in order to rule out the possibility that, after an unlucky run, a low quality cluster solution is accepted. In statistics of quality, *repeatability* is the variability in the performance of a system due to internal factors of that system (Montgomery, 1997). We define repeatability here to be the degree under which an algorithm produces the same outcome of a cluster instance, given a set of fixed parameter values. An algorithm with a high degree of repeatability produces roughly the same solution in each run.

In the following experiment, the seed of the random distribution is varied; 200 different randomly generated seeds are taken. The seed determines the ac-

ceptance probabilities in the successive steps of the algorithm. The parameters of the simulated annealing and the starting solutions are fixed; the starting solution is the countries-as-segments solution from Section 5.6. The distribution of the MSSC scores of the resulting 200 simulated annealing solutions are depicted in Figure 5.16. After 200 runs, seven different simulated annealing solutions have been returned. Among them is an outlier with an MSSC score of 9.804 occurring in one run; we have no explanation for this outlier. The MSSC scores of all other solutions are situated in the interval $[9.18, 9.36]$. The remarkable outlier in the MSSC scores is not present in the the scores of the silhouette width.

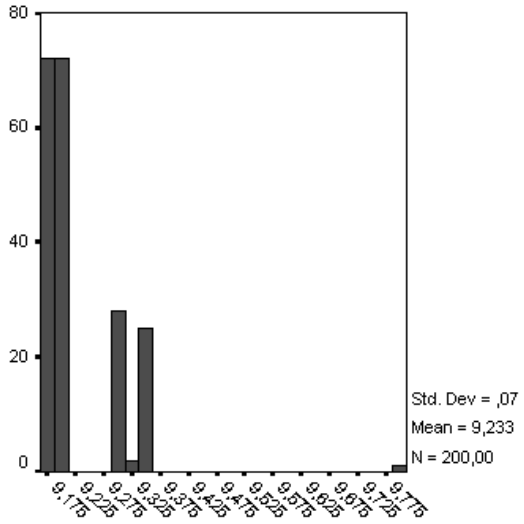

Figure 5.16. Histogram of MSSC scores

Table 5.7. Statistics of multiple simulated annealing

	MSSC	Silh. width	Log. costs
Mean	9.233	0.4558	70026
St. dev.	0.073	0.0045	417
Minimum	9.181	0.4454	68819
Maximum	9.804	0.4602	70902

Table 5.7 presents the most relevant statistics on the MSSC score, the silhouette width, and the logistics costs of the segmentation. The results show that

the variability between the solutions is small. Our results here indicate that simulated annealing is reliable when applied to the data set of Section 5.6, and a single run is probably sufficient.

5.8 An Alternative Approach: Restricting the Maximum Diameter within Segments

In the Budget Constraint Approach, the logistics costs of each candidate segmentation are computed. This can be a very time-consuming activity, in particular for large data sets. As a quick alternative, we present the *Diameter Constraint Approach*.

Low logistics costs are achieved when all consumers are easily and quickly reachable from a central facility, and the distance between each pair of consumers is small. An intuitive idea is to require segments to be connected; see e.g. Pawitan and Huang (2003) on Irish precipitation data. However, connected segments are no guarantee for low logistics costs. Instead, we choose to restrict the maximum distance between each pair of subjects in the same segment. This makes sense if there is a positive relationship between the logistics costs of a segment and the largest distance between two consumers a segment. More formally, a pair of subjects is allowed to be in the same segment if the geographical distance between them is at most D. We call D the *diameter* of the segment.

Is in true in general that a smaller diameter leads to segmentations with lower logistics costs? Our experimental results on the randomly generated instances from Milligan (1985), where the locations of the consumers are located on a 100 by 100 plane, shows that it is indeed true. Figure 5.17 shows the effect of changing the value of D on the logistics costs, measured with the DCF-measure. The logistics costs of the unconstrained segmentation are set to 100%. It appears that there is a positive relationship, but when the value of D increases beyond a threshold value, in this figure approximately $D = 110$, the logistics costs seem to stabilize. The explanation for this stabilization is that the resulting segmentations for values of D between 110 and 140 are more or less the same.

The *Diameter Constraint Approach* works along similar lines as the Budget Constraint Approach: it starts with a low logistics costs segmentation with a small value of D. Then it gradually increases the value of D until its value is

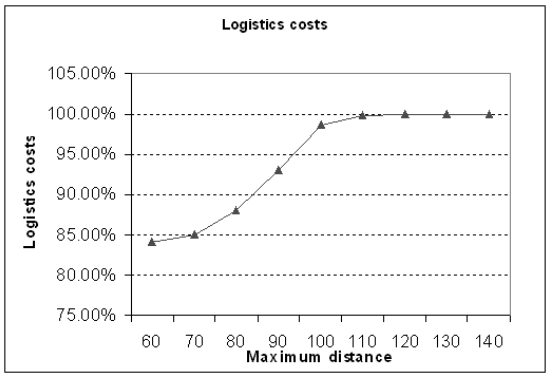

Figure 5.17. Logistics costs of random instances (unconstrained = 100%)

so large that no constraints are actually imposed on segmentations. As a consequence, we obtain segmentations with various diameters, and hence, with varying logistics costs.

The following clustering problem is obtained for selected combinations of K and D.

$$\begin{aligned}
\min \quad & \textstyle\sum_{i=1}^{N} \sum_{k=1}^{K} x_{ik} \|f_i - z_k\|^2 && && (5.8.1)\\
s.t. \quad & \textstyle\sum_{k=1}^{K} x_{ik} = 1 && i = 1, \ldots, N && (5.8.2)\\
& d_{ij} x_{ik} x_{jk} \leq D && i, j = 1, \ldots, N && \\
& && k = 1, \ldots, K && (5.8.3)\\
& x_{ik} \in \{0, 1\} && i = 1, \ldots, N && \\
& && k = 1, \ldots, K && (5.8.4)
\end{aligned}$$

Constraint (5.8.3) requires that the distance between two subjects in the same segment does not exceed D.

Consider the following small example with 15 consumers, and only one attribute on which the consumer scores are measured. Consumers 1 to 5 achieve a score of 2 on the attribute; 6 to 10 achieve a score of 6, and consumers 11 to 15 a score of 10. The consumers are randomly dispersed across a 20 by 20 plane; see Figure 5.18. The usual objective of segmentation is to construct segments which are homogeneous within and heterogeneous between segments. Clearly, the best segmentation divides the population into three segments consisting of the consumers 1 to 5, 6 to 10, and 11 to 15, respectively. Segment 1, indicated by

the triangles, and segment 3, indicated by the squares, have large geographical dispersions; see Figure 5.18.

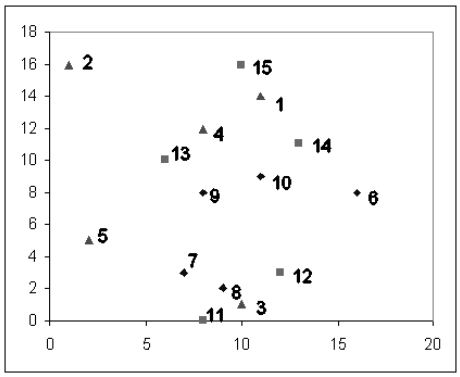

Figure 5.18. Location of consumers in the example

We reduce the logistics costs by restricting the diameter D of the segments to 15 and 10, respectively. When $D = 15$, consumer 4 moves from segment 1 to 2, and consumer 11 moves from segment 3 to 2, so the geographical outliers in segments 1 and 3 are reassigned. When D is decreased further to 10, three other consumers change segment membership; see Figure 5.19. The logistics costs of the segmentation decrease from 38.52 to 33.36 when $D = 15$ and to 27.96 when $D = 10$. On the other hand, when $D = 15$, two consumers are incorrectly classified, and when $D = 10$, the number of incorrectly classified consumers is five. This means that five consumers are served with marketing mixes which are ill-suited to their preferences. For this small example, a decrease in the maximum diameter leads to segmentations with lower logistics costs, but also to consumers being assigned to the wrong segment.

The choice of restricting the diameter D instead of B has two advantages. In the first place, it takes a much smaller amount of time to check whether each segment has diameter D than to check whether the logistics costs of a segmentation stay within the budget B. For example, in the case study presented in Section 5.5 ($N = 123$), the Diameter Constraint Approach needs only about two seconds to compute a good segmentation on a Pentium computer with speed 2 GHZ and 256 MB RAM under Windows 2000. On the other hand, the Budget Constraint Approach needs about 200 seconds on the same machine to complete the computation of a good segmentation for fixed values of K and B. The difference in

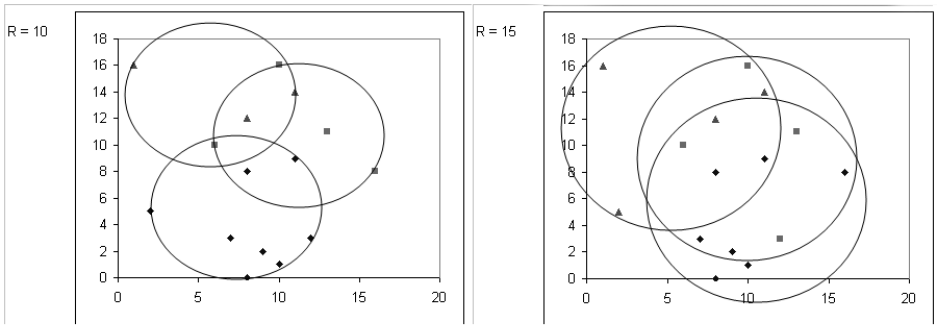

Figure 5.19. Segmentations with $D = 10$ (left) and $D = 15$ (right)

solution times becomes more evident for large data sets. Secondly, the diameter constraint can be implemented easily in the NORMCLUS of DeSarbo and Grisaffe (1998). This framework already contains constraints on the maximum travel distance between consumers.

A possible drawback of the Diameter Constraint Approach is that the diameter constraint applies to each individual segment. As a consequence, it is not possible to compensate a high logistics costs segment with a low logistics costs segment. This trade-off is clearly possible within the Budget Constraint Approach, so that the set of possible solutions is larger for the Budget Constraint Approach than for the Diameter Constraint Approach. At the same logistics costs level, we expect that the Budget Constraint Approach achieves a better fit score.

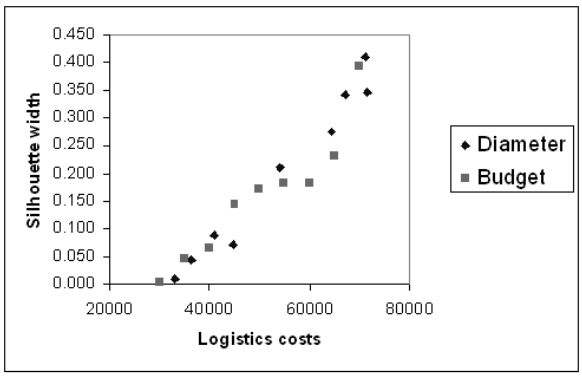

Figure 5.20. Comparison between BCA and DCA

Figure 5.20 shows the quality of the segmentations obtained with both approaches for $K = 4$; similar patterns are obtained for other K. The fit is measured with the silhouette width. The Diameter Constraint Approach achieves segmentations of approximately the same quality at fixed levels of the logistics costs, so it provides a reasonable approximation in case the Budget Constraint Approach is too time-consuming.

5.9 Solving Clustering Problems with Branch and Bound

Segmentation decisions are usually strategic decisions, and resources are fixed for long periods of time. For such decisions, it typically pays off to invest additional time to obtain optimal solutions. Here, we concentrate on exact methods for Minimum Sum-of-Squares Criterion (MSSC) clustering problems; see Section 5.3. Algorithms for other types of $\mathcal{NP}$-hard clustering problems are reviewed in Hansen and Jaumard (1997).

Most methods used to solve MSSC clustering problems are heuristics. BnB methods are presented in Koontz *et al.* (1975), Diehr (1985), and Du Merle *et al.* (2000). Exact methods are not so common for clustering problems, because until now, they have only been able to solve small instances to optimality. For example, the largest instance solved to optimality in Diehr (1985) contains 120 subjects, and in Du Merle *et al.* (2000) 150 subjects, requiring several hours of computing time. Exact methods for MSSC clustering are therefore only casually mentioned in the reviews by Jain *et al.* (1999) and Wedel and Kamakura (1998).

In Diehr (1985), a combinatorial BnB approach is presented that is based on the BnB algorithm from Koontz *et al.* (1975). In Du Merle *et al.* (2000), the MSSC clustering problem is formulated and solved as an Integer Programming problem. Variable Neighborhood Search (VNS), a meta-heuristic, is used to determine an initial solution. To the best of our knowledge, there is no direct comparison between both BnB algorithms, but the algorithm from Du Merle *et al.* (2000) solves harder and larger cluster instances to optimality.

The BnB algorithm from Diehr (1985) starts off by constructing an initial solution with a hill-climbing heuristic. It assigns subject 1 to cluster 1, leaving all other subjects unassigned. It then assigns subjects to clusters one by one in a prespecified order. The branching rule assigns the subject to the closest currently

available cluster. By 'closest' we mean that the center of the cluster is the closest to the attribute scores of the subject. Assume that there are N subjects to be clustered. When the BnB algorithm arrives at the n-th subject, it has already assigned and fixed the previous $n-1$ subjects to clusters. In Koontz *et al.* (1975), it is shown that the costs after n subjects have been assigned, form a lower bound for every cluster solution of that subproblem. In Diehr (1985), a tighter lower bound is suggested. When the BnB algorithm arrives at the n-th node, there are $N-n$ unassigned subjects left ($n = 1, \ldots, N-1$). In the terminology of Section 3.6, these subjects are called *offenders*, because the subjects are still unassigned. At subject n, a lower estimation of the minimum cost of assigning the remaining subjects to clusters is added to the lower bound.

It is too time-consuming to compute a new lower bound at each node of the search tree. Instead, the BnB algorithm determines optimal cluster solutions of subsets before the actual BnB process begins. For example, for the subsets $\{1, \ldots, 15\}, \{16, \ldots, 30\}, \ldots$, optimal solutions are computed. Diehr (1985) shows that the sum of the MSSC costs of optimal cluster solutions of subsets of subjects form a lower bound for the value of optimal cluster solutions. Small clustering problems of size 15 are solved rapidly. To obtain a lower bound, we first compute the cluster solutions of the subsets $\{N-14, \ldots, N\}, \{N-29, \ldots, N-15\}, \ldots$. At $n = N-30$, the solution values of the subsets $\{N-29, \ldots, N-15\}$ can be added to the current lower bound.

In this section, we take the simulated annealing solutions from Section 5.6, and try to improve these solutions using a BnB algorithm or to prove their optimality. Cluster problems with a restricted diameter D are also solved. We expect that the BnB algorithms are most successful when the search space is restricted, since the number of feasible solutions is then much smaller.

In Table 5.8, the basic BnB algorithm from Koontz *et al.* (1975) is compared to a version of the algorithm that employs simulated annealing to obtain an upper bound (UB), a version which employs the lower bound described above (LB), and a version that employs both (UB & LB). The basic algorithm has no additional upper or lower bound computations. The values in the table are the MSSC scores obtained when the corresponding the BnB algorithm is terminated after at most one hour of computing time. The instances are from the case study from Section 5.5, the number of clusters here is 6. In addition, we solve instances with

Table 5.8. Comparison of cluster solutions of BnB cluster algorithms with new upper and lower bounds

	Added to BnB algorithm				
D	Sim. ann.	Basic	UB	LB	UB & LB
800	19.679	22.668	19.679	22.668	19.679
1000	15.747	22.668	15.747	22.668	15.747
1200	13.851	22.668	13.851	22.668	13.851
1400	13.284	22.668	13.284	22.668	13.284
1600	11.881	22.668	11.881	22.668	11.881
1800	10.748	22.668	10.748	12.732	10.748
2000	10.236	13.002	10.236	12.459	10.236
2200	9.902	12.621	9.902	12.219	9.902
2400	9.292	12.188	9.292	11.828	9.292
2600	9.237	12.188	9.237	11.828	9.237

various values of D. The number of possibilities in the restricted cluster problems is smaller. Hence, we expect cluster problems with small values of D to be more easily solvable with the BnB algorithms.

It turns out that none of the BnB algorithms in Table 5.8 are able to solve one of the tested instances within an hour; the simulated annealing starting solution is not improved and optimality of the simulated annealing algorithm is not shown. It turns also out that for restricted problems, the BnB algorithms do not perform well. A possible reason is that the BnB algorithms have trouble finding good upper bounds. The lower bound improves the quality of the solutions obtained, but the results indicate that the lower bounds require the most urgent improvements.

In the iterative patching procedure of Chapter 2, upper bounds are constructed at each node of the search tree. Is it worthwhile to use simulated annealing in a similar iterative fashion, i.e., to compute upper bounds at multiple nodes of the BnB search tree? Iterative simulated annealing is probably not leading to tighter lower bounds. We have found that simulated annealing is relatively independent of starting solution, and that multiple optimization runs are not leading to solutions of different quality.

In this section, we find that the BnB algorithm described in Diehr (1985) is not able to improve the simulated annealing solution for instances of the case study from Section 5.5. This confirms the general agreement that exact algo-

rithms are not sufficiently developed to be usable in practice. For future research, it is interesting to include the approach by Du Merle *et al.* (2000) into the comparison.

5.10 Limitations and future research

The Budget Constraint Approach determines segmentations for various levels of the logistics costs budget B and the number of segments K. The logistics costs and the fit are then weighted, and a small number of candidate segmentations is obtained. This approach has the following limitations:

- The approach presented in this paper is tailored to 'hard' segmentations in which every consumer is assigned to a single segment, but it is doubtful whether it works well for general segmentations. The majority of segmentation studies is best solved with mixture models (Wedel and Kamakura, 1998). An interesting direction of future research is to develop a mixture modeling approach which limits the logistics costs, for example, by offering marketing mixes only to a closely located set of consumers. The NORMCLUS framework of DeSarbo and Grisaffe (1998) also offers great possibilities for constrained fuzzy clustering. Mixture models with restrictions on the set of feasible segmentations are also discussed in Law *et al.* (2004) and Shental *et al.* (2004).

- The Budget Constraint Approach returns a small set of candidate segmentations, but the choice for the most profitable one should still be made manually; we cannot compute directly which one generates the highest profit for the organization. The most profitable segmentation can be determined accurately if the model is able to convert the expected consumer benefits of a segmentation into revenues for the company.

- We assume that the number of segments does not influence the logistics costs level. This assumption appears to be plausible, since we find that the number of segments has very little influence on the transportation costs. If the number of segments is taken into account and weighted separately, efficient combinations of three variables need to be determined. This requires *multi-criteria analysis*.

- The DCF measure of the logistics costs is based on the assumption that each segment is served from a separate facility. When the geographical dispersion of consumers in the same segment is large, it may be the most profitable strategy to build one joint central facility, such as an EDC, for all segments. The cost savings of joining segment facilities are not taken into account.

- The performance of simulated annealing depends strongly on the choice of the parameters (Henderson *et al.*, 2003). This means that, in order to obtain high quality segmentations, these parameters should be readjusted for each new instance. An interesting direction of future research is to develop an *adaptive simulated annealing algorithm* (Henderson *et al.*, 2003), which automatically adapts the settings of the algorithm to the segmentation instance at hand. In our experiments, the initial temperature and the freezing temperature appear to be the most important parameters; they should both be set proportional to the approximate MSSC score of the cluster solution.

- An emerging topic in marketing is *store location* (Achabal *et al.*, 1982; Craig *et al.*, 1984; Clarke *et al.*, 1997). It turns out to be one of the key factors for the success of the stores (Reinartz and Kumar, 1999; Kumar and Karande, 2000; Mittal *et al.*, 2004). When a retail organization plans to serve the customers in a region with differentiated marketing mixes, it is not only important to know the characteristics of the different types of customers, but also their distances to possible store locations. The travel times can be seen as the logistics costs made by the consumers. An interesting direction of future research is the application of our approach to the problem of assigning similar customers to stores in a region, while minimizing the travel costs made by the customers. A similar application is described in Yorke (2001) where leisure facilities in Britain are considered. These facilities should be adapted to the desires of children in the neighborhood, but travel times to the locations should be taken into account simultaneously.

5.11 Conclusions

In the literature on segmentation, it is assumed that profit is maximized when the fit of consumer preferences is maximized. However, the costs of physical distribution make up, on average, about 20% of the total cost of a product (Davis, 1990). Steenkamp and Ter Hofstede (2002) report that logistics costs force organizations to maintain a countries-as-segments strategy in retailing and in case of perishable goods. We introduce new segmentation strategies, which make the trade-off between the consumer benefits of a segmentation and the logistics costs of serving the segments possible. The *Budget Constraint Approach* uses simulated annealing to construct segmentations. The simulated annealing algorithm finds good segmentation solutions, independent of the chosen starting solution (Wedel and Kamakura, 1998).

Research on random instances indicates that the Budget Constraint Approach is applied successfully if the selected number of segments is not too small. We have also applied the approach on a European meat outlet study, in which retail formulas are assigned to European regions. The approach is able to generate intermediate solutions for which both the logistics costs and the fit of the segmentation are reasonable.

The method presented in this paper is applicable to a small subclass of segmentations, namely those in which the company assigns the consumers to segments. Interesting directions of future research are to include the logistics costs into general segmentations for which mixture models appear to be the most suitable modeling method. Finally, it may be worthwhile to quantify the relationship between profit of the organization and the fit of consumer preferences. When the profits of a segmentation strategy are known, an effective cost-benefit analysis can be made.