PERGAMON

Computers & Operations Research 27 (2000) 1227–1247

computers & operations research

www.elsevier.com/locate/dsw

Segmentation-based competitive analysis with MULTICLUS and topology representing networks

Thomas Reutterer[a,*], Martin Natter[b]

[a]*Department of Retailing and Marketing, Vienna University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria*

[b]*Department of Production Management, Vienna University of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria*

Abstract

Two neural network approaches, Kohonen's self-organizing (feature) map (SOM) and the topology representing network (TRN) of Martinetz and Schulten are employed in the context of competitive market structuring and segmentation analysis. In an empirical study using brands preferences derived from household panel data, we compare the SOM and TRN approach to MULTICLUS, a parametric latent vector multi-dimensional scaling (MDS) model approach which also simultaneously solves the market structuring and segmentation problem. Our empirical analysis shows several benefits and shortcomings of the three methodologies under investigation. As compared to MULTICLUS, we find that the non-parametric neural network approaches show a higher robustness against any kind of data preprocessing and a higher stability of partitioning results. As compared to SOM, we find advantages of TRN which uses a more flexible concept of adjacency structure. In TRN, no rigid grid of units must be prespecified. A further advantage of TRN lies in the possibility to exploit the information of the neighborhood graph for adjacent prototypes which supports ex-post decisions about the segment configuration at both the micro and the macro level. However, SOM and TRN also have some drawbacks as compared to MULTICLUS. The network approaches are, for instance, not directly accessible to inferential statistics. Our empirical study indicates that especially TRN may represent a useful expansion of the marketing analyst's tool box.

Scope and purpose

Determination of competitive market structure among rival brands and market segmentation represent well-known concepts in strategic marketing planning. During the last decade, approaches that combine the two interrelated tasks into one single model have been introduced into marketing literature. Most of them respect consumer heterogeneity by including 'fixed' parameters (e.g., demographic or past purchase behavior

* Corresponding author. Tel.: + 43-1-31336-4619; fax: + 43-1-31336-717.

E-mail addresses: thomas.reutterer@wu-wien.ac.at (T. Reutterer), martin.natter@wu-wien.ac.at (M. Natter)

0305-0548/00/$ - see front matter
PII: S0305-0548(99)00147-1

variables) for each individual or by assuming consumer parameters to be distributed according to a (mixture of) probability distribution(s). However, the key to the success of these statistical modeling approaches seems to lie in the proper choice of parametric model assumptions and/or heterogeneity distributions. Due to its non-parametric nature, the neuro-computing methodology presented in this article imposes less rigorous assumptions on data properties and derives segment-specific patterns of competitive relationships between brands in a purely data-driven way.

1. Introduction

Competitive market structure (*CMS*) analysis refers to the task of deriving a configuration of brands in a product class on the basis of their competitive relationships. It is widely accepted by marketing scientists to operationalize the degree of inter-brand competition as a measure of substitution as perceived by consumers [1,2]. However, once the data analyst wishes to introduce heterogeneity across consumers (e.g., in terms of preference and/or consideration set) into the model, CMS turns out to be a *segment-specific concept*, which imposes the issue of deciding about the appropriate level of data aggregation and making CMS and market segmentation analysis to be dependent on each other.

The paper proceeds as follows: First, a brief outline of contemporary approaches to simultaneous CMS/segmentation analysis is provided. Following, the adaptive self-organizing (feature) map (SOM) methodology according to Kohonen [3] and the more recently introduced topology representing network (TRN) model by Martinetz and Schulten [4] are adopted to this task and applied in an empirical study using household-level panel data. Finally, validity issues are discussed and both SOM and TRN results are opposed to those emerging from an unfolding mixture approach.

2. Combined CMS and segmentation analysis

It is quite well understood in marketing literature that utilization of brand choice probabilities as segmentation basis turns CMS and market segmentation out to be "reverse sides of the same analysis" [5]. In fact, the basic difference of these two sides refers to the kind of how an observed data matrix is processed. According to data theoretical terminology, the shape of a data matrix is determined by the number of ways (dimensions) and the number of modes (sets of different entities that represent the ways of the matrix, e.g., consumers, brands or brand attributes). From this point of view, the two concepts may be considered as formally identical data reduction problems. As illustrated in Table 1, the only difference concerns the focused *mode for data reduction*:

Conventional approaches to CMS analysis reduce the brands mode only. They typically result in either a "non-spatial" arrangement of ultra-metric trees, overlapping or fuzzy cluster structures (for a review cf. [1]) or a "spatial" representation of brands configurations in a geometric space. Spatial models can be further subdivided into "compositional" methods involving reduction of high-dimensional attribute spaces via principal components, discriminant or correspondence analysis and "decompositional" approaches [2]. The latter are usually based on multi-dimensional scaling

Table 1
Synopsis of models for CMS/segmentation analysis

Mode of data reduction	Representation type	
	Discrete ("non-spatial")	Geometric ("spatial")
Brands	Hierarchical (tree models)/non-hierarchical classification	Compositional/decompositional positioning analysis
	'*Combined*' *approaches* (*e.g.*, *STUNMIX*)	
Subjects (consumers)	A posteriori market segmentation	Preference scaling models

(MDS) of respondents' proximity or dominance statements about rival brands. Furthermore, certain unfolding techniques of preference data embed the consumers mode as ideal vectors or points in a "joint space" of consumers and brands configurations.

A posteriori *market segmentation*, on the other hand, refers to a compression of the consumers mode, which is usually achieved via clustering or latent class techniques [6].

In contrast, approaches for *combined CMS/segmentation analysis* reduce the consumers and brands mode simultaneously in one single model. Proposals towards this direction are presented, e.g., by Hruschka [7] using fuzzy clustering methods, Grover and Srinivasan [5] or Kamakura and Russell [8] in a latent-class framework or Wedel and Steenkamp [9] using a cluster-wise regression procedure. Another promising stream of modeling efforts equipped with the option of introducing the consumer mode into CMS analysis is represented by models for multimode factor analysis. Cooper et al. [10] illustrated in a study that constrained versions of three-way factor analysis are also able to uncover asymmetric competitive relationship patterns between rival brands.

2.1. Multiclus

A particularly interesting class of models for simultaneous CMS/segmentation analysis employ "latent class multidimensional scaling" (LCMDS) techniques [11,12]. For illustration of current LCMDS methodology, consider the MULTICLUS model of DeSarbo et al. [13]:

MULTICLUS is designed to process profile or dominance data and simultaneously performs MDS and cluster analysis in such a way that a D-dimensional joint space of brand coordinates and cluster vectors is estimated for predetermined numbers of C clusters and D dimensions. In the MULTICLUS model, the observed profile/dominance values $\{\Delta_{km}\}$ of consumers $k = 1, \ldots, K$ for brands $m = 1, \ldots, M$ are interpreted as realizations of an $1 \times M$ random vector $\mathbf{\Delta}_k = (\Delta_{k1}, \ldots, \Delta_{kM})$, which probability density function is modeled as a finite mixture of conditional distributions:

$$g(\mathbf{\Delta}_k; \lambda, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}) = \sum_{c=1}^{C} \lambda_c f_{kc}(\mathbf{\Delta}_k | \mathbf{a}_c, \mathbf{B}, \mathbf{\Sigma}_c) \tag{1}$$

with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{C-1})$ independent mixing proportions (such that $0 \leqslant \lambda_c \leqslant 1$ and $\lambda_c = 1 - \sum_{c=1}^{C} \lambda_c$). The matrices $\mathbf{A} = \{a_{cd}\}$ and $\mathbf{B} = \{b_{md}\}$ resemble the $1 \times D$ row vectors of cluster and product coordinates, respectively, where $\mathbf{a}_c$ contains the vector terminus coordinates for cluster c. Finally, $\mathbf{\Sigma}$ is an $C \times M \times M$ array that stacks together the cluster-specific variance-covariance matrices $\mathbf{\Sigma}_c$.

In their article, DeSarbo et al. [13] specify the distribution of each f_{kc} as conditional multivariate normal. For a given sample of consumers' brand evaluations, dimensionality D, and number of clusters C, the estimation of model parameters λ, $\mathbf{\Sigma}$, $\mathbf{A}$, and $\mathbf{B}$ is achieved by maximizing the log likelihood function:

$$\ln L = \sum_{k=1}^{K} \ln\left[\sum_{c=1}^{C} \lambda_c f_{kc}(\mathbf{\Delta}_k | \mathbf{\Sigma}_c, \mathbf{a}_c, \mathbf{B}) \right]. \tag{2}$$

For this purpose, an EM algorithm is used (cf. Dempster et al. [14]). The MULTICLUS model structure respects consumer heterogeneity with regard to attribute perceptions or brand preferences via maximum likelihood estimates for segment-level ideal vectors in a joint space of brand coordinates.

More recently, Wedel and DeSarbo [15] presented a generalization of this and parallel developed approaches of stochastic MDS unfolding mixture models (STUNMIX). This generalized STUNMIX methodology is formulated in the framework of the exponential family of distributions, encompassing the normal, binomial, Poisson, gamma, inverse Gaussian, and other distributions within the class of the exponential family. Thus, a wide range of data, such as ratings, choices, frequencies duration times are allowed to be (simultaneously) analyzed by the same methodological approach. Furthermore, various combinations of different types of preference models, re-parameterization options and simultaneous segment description are available. The authors demonstrate their generalized STUNMIX approach using synthetic data generated from distributions of the exponential family and real-world data.

However, even if current state-of-the-art STUNMIX methodology is flexible enough for detecting a wide range of (hypothesized) underlying mixture distributions and seems to be relatively robust to moderate misspecifications (see the experimental findings reported by Wedel and DeSarbo [15]) the suitable representation of consumer segments and brand positions finally turns out to be dependent on the proper choice of some functional and distributional form of the parameters to be estimated. More flexibility is conceded by an approach recently introduced by Allenby, Arora and Ginter [16]. In this model, heterogeneity is interpreted as a continuous concept and 'segments' are modeled at the individual level using the Gibbs sampling technique in the framework of a model of normal mixture components.

In the following section, we introduce a non-parametric alternative to the STUNMIX model family. The proposed methodology inspired by neuro-computing technology imposes no specific assumptions on input data requirements (such as data type or underlying distributions). Thus, it may serve as a promising and flexible tool for multimode data compression especially for cases where the hypothesized distributional assumptions are seriously violated and/or the analyst is not able (e.g., due to a lack of any well-grounded theory) or willing to impose any reasonable pre-specification of the distribution type. In the following, we present the SOM model and the more

recently introduced TRN methodology, which may be viewed as a more flexible extension of the standard SOM procedure.

2.2. SOM

Like other partitioning techniques, the SOM methodology is based on an unsupervised learning scheme. Similar to principal components analysis or MDS, an SOM network constructs a low-dimensional mapping in order to detect the inherent structure of high-dimensional input data in a visually easily inspectable manner. However, SOMs do not produce geometric configurations of brand positions like those estimated by (LC)MDS (STUNMIX) models. SOMs, in contrast, are facing the dual problem of constructing a classification of input data space and (as an additional feature making SOMs different from more conventional clustering methods) simultaneously to this class centers or prototypes are mapped onto a pre-specified grid of units so that the topological structure of the input data is represented as accurately as possible. As a result they arrive at a non-linear projection of the original input data space onto a discrete map of topologically ordered units. The projection is usually done adaptively, i.e., disaggregated input vectors — each of them reflecting, e.g., the observed pattern of purchase behavior of one individual respondent — are presented in a sequential order, which in the neural network literature is frequently referred to as "network training" [3].

The SOM network architecture first introduced by Kohonen [3,17] consists of the following components: An m-dimensional input layer representing features (in our empirical study: brand preference values) of input vectors $x_k = [x_{k1}, x_{k2}, \ldots, x_{km}]^T$ out of a set of $k = (1, \ldots, K)$ training vectors (e.g., a sample of consumers or households), a two-dimensional competitive layer organized as a grid of units u_{ij}, where i represents the row index and j the column index of a unit position in the layer, and an m-dimensional weights vector for each SOM unit $u_{ij}: w_{ij} = [\mu_{ij1}, \mu_{ij2}, \ldots, \mu_{ijm}]^T$.

The standard SOM algorithm and its variations are well documented in the relevant literature (cf. Kohonen [3,17] or Nour and Madey [18] for a review). The implementation used in our study can be briefly described by the following iterative procedure (for further details see [19]):

1. Set counter $t := 0$ and initialize weight vectors $w_{ij}(t)$ to random values within $[-0.1; +0.1]$.
2. Randomly choose an input vector $x_k(t)$ (here: brand choice probabilities of household k) and determine the "winning" or "best matching" unit $u_{c_{ij}}(t)$ with associated activity $w_{c_{ij}}(t)$ according to:

$$||x_k(t) - w_{c_{ij}}(t)|| = \min_{ij} \{||x_k(t) - w_{ij}(t)||\}, \tag{3}$$

   where the index c_{ij} denotes the location of the winner's position in the two-dimensional layer.
3. Update the weights vectors as follows:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t) h_{c_{ij}}(t)[x_k(t) - w_{ij}(t)]. \tag{4}$$

4. If $t \geqslant T$ stop, else increment t and go to step 2.

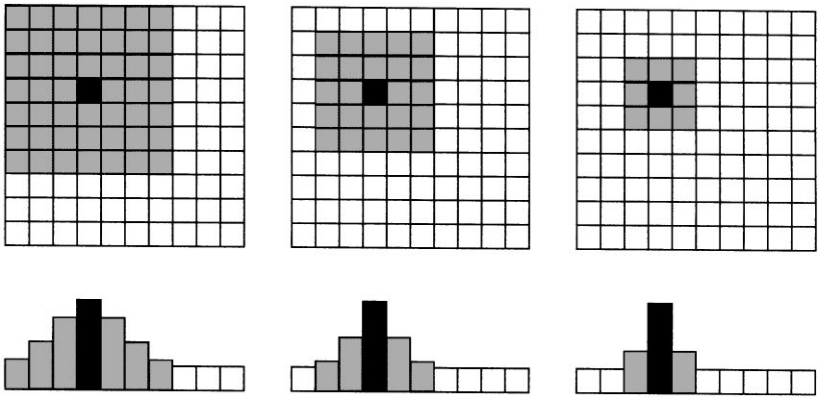

Fig. 1. Neighborhood adjustment for three different kernel radii.

Here t is the discrete time step and T the predefined maximum number of iterations. Both the learning rate $0 \leqslant \alpha(t) \leqslant 1$ and the neighborhood parameter $h_{c_{ij}}(t)$ are monotonically decreasing functions with time. In the SOM algorithm available from the first author under MATLAB the functional form $\alpha(t) = \alpha(0)(c/(t + c))$ with constant $c = T/100$ and $\alpha(0) = 0.1$ is used; T is set to 20 epochs (i.e., runs through the input data set). Unlike the discrete-valued neighborhood chosen by Mazanec [20] in a previous work, here a smoother kernel-based neighborhood update process is controlled by the Gaussian function

$$h_{c_{ij}}(t) = \exp\left(- \frac{[(i - c_i)^2 + (j - c_j)^2]}{2\sigma(t)^2} \right), \tag{5}$$

where i (j) is the row (column) index of an SOM unit u_{ij} and (c_i, c_j) the corresponding indices of unit $u_{c_{ij}}$ (the winner in step 2 of the above algorithm). Hence, the numerator in Gaussian (5) measures the topological distance between any SOM unit u_{ij} and the winner $u_{c_{ij}}$. The kernel width parameter $\sigma(t)$, which is responsible for the degree of neighborhood adjustment (see Fig. 1) also follows a shrinking function of time.

The adoptions imposed to the weight vectors by the combination of steps 2 and 3 are vital for the ultimate objective of the SOM learning procedure: In step 3, the input vector x_k presented to the SOM layer at iteration t is mapped onto unit $u_{c_{ij}}$, i.e., the "winning" unit determined in step 2. This is achieved by allowing a "maximum" adjustment of the corresponding weights vector for the amount of $\Delta w_{c_{ij}}(t + 1, t) = a(t)[x_k - w_{c_{ij}}(t)]$. Since the learning rate $a(t)$ is forced to decrease with time, a classification of the input data space would be achieved via stochastic approximation of the minimum (inner group) variance partition (cf. Bock [21]).

However, as the formulation of the updating rule in step 3 suggests, there is another feature of SOM learning, which is responsible for the topological ordering of cluster prototypes onto the discrete map of units: Not only the "winning" or minimum distance unit $u_{c_{ij}}$ to the vector x_k presented at iteration t but also those units adjacent to the "winner" according to the present width of the neighborhood kernel $h_{c_{ij}}(t)$ around the "winner" are activated to gradually learn from

the same input vector. Fig. 1 illustrates this neighborhood adjustment effect for three different kernel radii.

While the (black-colored) "winning" unit $u_{c_{ij}}$ for training vector x_k receives a maximum update by the respective learning rate $\alpha(t)$ weighted difference $x_k - w_{c_{ij}}(t)$, the adjacent units (gray-shaded) only benefit from the adjustment to a much smaller degree as allowed by the $h_{c_{ij}}(t)$ values for neighbors of first, second, third, etc., order. As the kernel width monotonically shrinks during training via decreasing $\sigma(t)$, topological ordering of the prototypes takes place during the initial training cycles and convergence towards centroids will be accomplished when the neighborhood kernel is sufficiently reduced.

Since we will use SOM formats of very small size (comparable to the applications reported, e.g., in [17]), $\sigma(0)$ is initially set to the diameter of the map and the decreasing functional form is chosen identical with those for the learning rate but a smoother decay rate of $c = T/10$.

Note, that in the case of keeping $\sigma = 0$ throughout the training (i.e., zero-neighborhood) the process is strictly equivalent to classical vector quantization techniques such as an on-line version of the K-means algorithm [22]. Furthermore, there also exists a link to supervised neural network models. As Mulier and Cherkassky [23] show for the batch version of the model (where the whole training sample is utilized in each iteration), SOMs can also be interpreted as an iterative kernel smoothing process to the class centroids and therefore may be used as an approximation device. In fact, the SOM model performs non-parametric regression of the u_{ij} coordinates on the input data space.

However, the predetermined 'grid' of SOM units imposes a rather rigid adjacency structure on the neighborhood update process during training which may not match with the topological structure of the input data space [4].

2.3. TRN

Martinetz and Schulten [4] adequately address this problem within the framework of their topology representing network (TRN) approach. Building up on a previous version of topology-sensitive vector quantization by Martinetz et al. [24] they start with a set of unordered units u_i, $i = (1, \ldots, N)$, and associated weights w_i. In contrast to the SOM model, adjacency of units is a dichotomous and dynamic concept in TRN analysis. The connectivity structure between units is stored in an $N \times N$ connection strength matrix $\mathbf{C}$, with elements $C_{ij} > 0$ denoting that unit u_i and unit u_j are connected and $C_{ij} = 0$ that there is no connection between them. In addition, each connection is evaluated by its age t_{ij} that indicates the number of iterations t the connections already exists without being confirmed. If the age of a connection exceeds a certain maximum lifetime T, the connection gets removed. This means that connections established in an early stage will be unlearned, if they are not regularly enough refreshed during the TRN training cycles. However, once removed the respective connection can again be re-established at a later stage of the adaptation procedure.

The TRN algorithm as applied in our study follows a similar iterative procedure than the above SOM algorithm and can be briefly described as follows (cf. [4,24]): First, the weights w_i are initialized with small random values and all connections C_{ij} are set to zero (i.e., no connections are active). At each iteration step of the adaptation process an input vector x_k is randomly chosen from the sample space and the "winning" unit u_{i_0} is determined equivalent

to the minimum distance rule applied in step 2 of the SOM procedure. However, according to the "neural gas" algorithm [24] employed in TRN analysis, for each prototype w_i the number r_i of units j with

$$\|x_k - w_j\| < \|x_k - w_i\|, \quad \forall j \tag{6}$$

is determined. Thus, at each iteration step a sequence $r = 0, 1, \ldots, N-1$ is constructed, with r_i indicating the rank number r associated with each of the prototypes (rank 0 is assigned to the "winning" prototype w_{i_0}, rank 1 to the "co-winner" w_{i_1}, rank 2 to the third-closest prototype to the current input pattern $x_k(t)$, etc.). Next, weight vectors updating is performed equivalent to step 3 of the SOM procedure:

$$w_i(t+1) = w_i(t) + \alpha(t) h_{i_0}(t)[x_k(t) - w_i(t)], \tag{7}$$

where $\alpha(t)$ again is a learning rate and the neighborhood function $h_{i_0}(t) = \exp(-r_i/\lambda(t))$ assures that only the "winning" prototype w_{i_0} gets fully updated. Adjacent prototypes participate in the updating process only gradually depending on their rank-ordered distance index from the "winner" as determined by (6). Both $\alpha(t)$ and the neighborhood width $\lambda(t)$ are monotonically decreasing functions of the iteration index. Since this kind of specification of neighborhood updating around the "winner" abandons the idea of a fixed grid of adjacent units, the TRN neighborhood concept is much more flexible than the rigid SOM adjacency updating.

Moreover, in the TRN model the adjacency structure between each pair of prototypes is constructed adaptively during training using information about the "winner/co-winner" relationships: In addition to the weights adjustments, at each iteration the connection strength $C_{i_0 j_1}$ between the current "winner" and "co-winner" prototype (i.e., the two prototypes closest to the input pattern x_k) is tested. The connection strength is activated by setting $C_{i_0 j_1} = 1$ and age $t_{i_0 j_1} = 0$. If the connection already exists (i.e., $C_{i_0 j_1} = 1$) it is only refreshed by re-setting the age counter to zero. The age of all other already existing connections of u_{i_0} is increased by one. Finally, connections of unit u_{i_0} which exceed lifetime T are removed ($C_{i_0 j} = 0$).

Martinetz and Schulten [4] showed that this adaptive procedure of dynamically learning and unlearning of adjacency structures between rivaling prototypes succeeds in preserving rather complicated topological data structures. They also quoted recommendations for parameterizations of the learning and neighborhood update parameters as well as suitable functional forms of time-dependent variables, which we adopted in the following.

3. Empirical study

In this section, we report the results obtained from empirical applications of the three models discussed (MULTICLUS, SOM, TRN). We proceed as follows: First, the issue of model identification and determination of the number of clusters (dimensions) is discussed and the derived solutions are characterized for each approach. Second, the resulting configurations are compared to each other in terms of variance accounted for and stability. Finally, we assess the predictive validity of the methodologies.

3.1. Data description

Input data are derived from purchase histories of 781 non-brand-loyal panel households for nine major brands accounting for 90% total market share in the margarine product class. There are six national brands and three private labels (#2, #8, #9). The market leading brand (#1) which accounts for about 31% of the total market volume, is a typical general purpose margarine. Brands #3 and #4 with a market share of about 11% each, are usually used for cooking and baking, whereas brands #5, #6, #7 and #9 are fat reduced or diet margarines.

The database for the combined CMS/segmentation problem consists of households' "zero-order" brand choice probabilities. Thus, each household k is described by a nine-component vector x_k. Element x_{km} measures the relative purchase frequency of brand $m = 1, \ldots, 9$ in a two years period, where $\sum_m x_{km} = 1$. Note that consideration set heterogeneity across households (the average set contains about three brands) results in a large number of zero entries.

3.2. MULTICLUS results

As recommended by DeSarbo et al. [13], input data for the MULTICLUS procedure need to be preprocessed prior to analysis. In fact, usage of the described raw data without utilizing any precautions leads to degenerate solutions (i.e., empty clusters with zero memberships) for all possible combinations of numbers of clusters/dimensions ($T \leqslant C = 1, \ldots, 15$) tested here. Row standardization, a frequently applied preprocessing step, is not appropriate for the data set at hand since brand choice probabilities (expressed as relative choice proportions) are constrained to sum up to unity and result in a singular covariance matrix. One common way out of the difficulties with unequal or correlated error variances is to apply a logistic transformation to the data:[1] $x' = \ln(x/(1-x))$. Furthermore, in order to account for different levels of market shares the logit transformed data are column standardized. As our main interest is to draw inferences about competitive relationship patterns among brands this additional preprocessing step is suitable to avoid configurations that simply reflect different choice shares of the brands.

Consequently, the results described below are derived from logit transformed and column (i.e., brand mode) standardized data. The MULTICLUS program allows for several options concerning estimation of the (cluster specific) covariance matrices $\mathbf{\Sigma}_c$. Here, estimation indicated that the best results in terms of Bayesian information criterion (BIC) and consistent Akaike information criterion (CAIC) measures are obtained for the case of one common diagonal covariance matrix $\mathbf{\Sigma}$ for all groups.

Of all the combinations of numbers of clusters/dimensions ($T \leqslant C = 1, \ldots, 15$) analyzed, lowest information criteria values are identified for a nine cluster/six-dimensional solution with CAIC = 17253.016 and BIC = 17253.004. This configuration accounts for 57.2% of variance. Unfortunately, the high dimensionality prohibits one single visualization of the market structure that is easily accessible to consistent interpretation. Therefore, dimensional projections have to

[1] We thank one of the anonymous referees for pointing this out (for further details on logit transformation and associated modeling techniques widely used in marketing research see [25,26]).

be analyzed two by two which may lead to contradictory conclusions on the underlying competitive structure. In general, MULTICLUS estimates showed a considerably high variance (each cluster/dimension combination has been replicated five times), especially with increasing numbers of clusters and dimensions. Note, that this symptom of extremely high parameterized solutions occurs even though the usual model-specific precautions in terms of suitable data transformation and preprocessing have been met. To reproduce a compressed version of the structure inherent to the data, MULTICLUS is quite parameter consuming. Of course, this is also an effect of sample size (here: 781 respondents); e.g., running the same procedure with random subsamples of only 10% of the data regularly yields three cluster solutions and two or three dimensions (with VAFs of around 30–40%).

For interpretation purposes we restrict ourselves to the first three two-dimensional sub-plots of the nine cluster/six-dimensional solution for the total data set as represented in Figs. 2–4. The configuration depicted in Fig. 2 clearly identifies clusters of households with distinctive preference patterns (relative segment sizes are indicated in brackets): The sub-markets portrayed by vector 9 (12.5%) as well as vectors 3 (6.5%) and 5 (6.9%) are characterized by store loyal customers since the private labels (brands #2, #8, and #9) are projected closest to them. Customers represented by segment vectors 2 (11.1%), 1 (11.4%) and partly by segment 4 (16.3%) show high relative preferences for cooking&baking brands. The latter segment, however, is clearly dominated by the market leading general purpose brand #1 with an average purchase probability of about 63%. Apart from a strong position in segment 1 brand #1 also yields a considerable share of choices in segment 6 (12.8%), where it's major competitor is the fat-reduced margarine brand #5. The diet

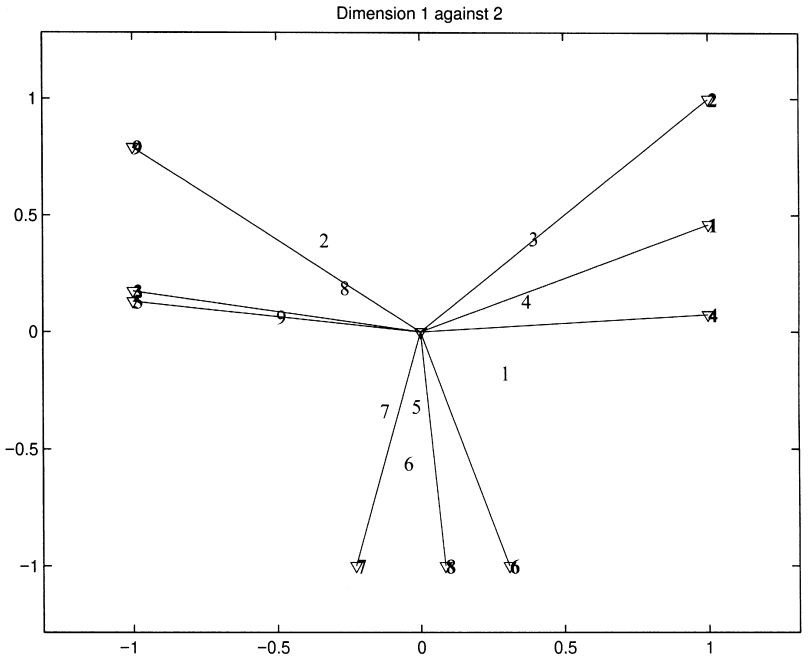

Fig. 2. MULTICLUS configuration (brands coordinates and cluster vectors in dimension 1 versus 2).

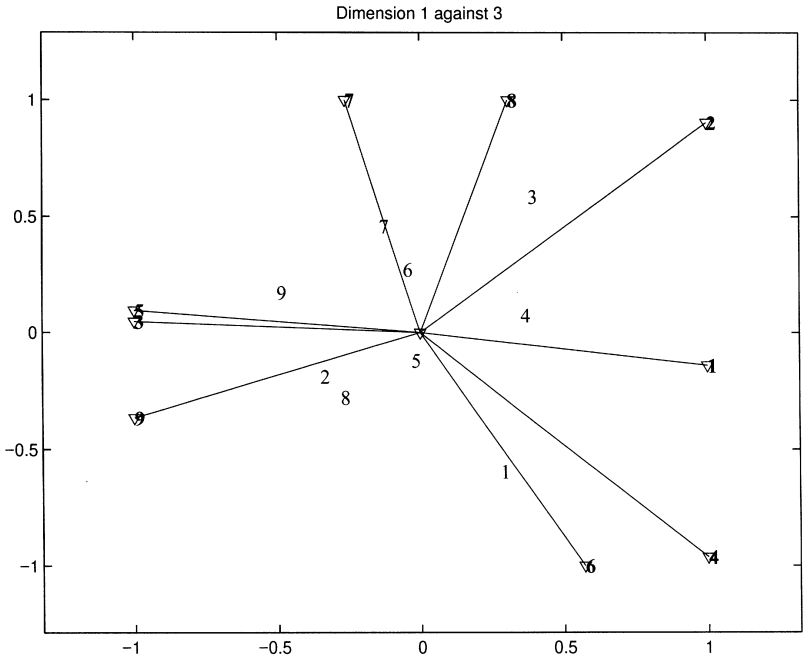

Fig. 3. MULTICLUS configuration (brands coordinates and cluster vectors in dimension 1 versus 3).

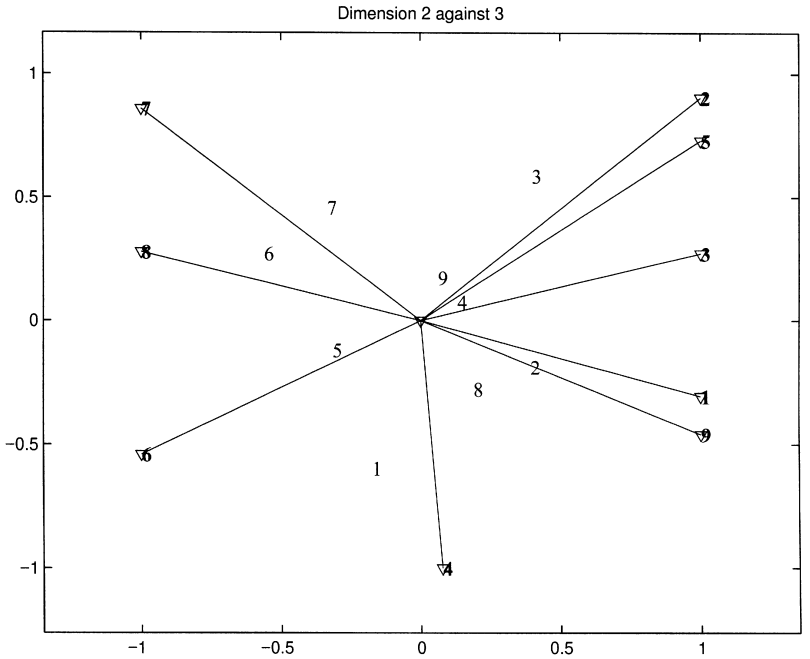

Fig. 4. MULTICLUS configuration (brands coordinates and cluster vectors in dimension 2 versus 3).

brands are mostly preferred by members of segments 6–8 with relative sizes of 12.8, 12 and 10.4% of households, respectively.

A similar portrayal of the market structure is provided in Fig. 3. However, projection of segment vector 6 provides a slightly different aspect of the market. Another variation is depicted by Fig. 4. As one common aspect of all the sub-plots (also of those not given here) three groups of brands that are plotted closer to each other are emerging, namely store brands (#2, #8, and #9), cooking&baking brands (#3 and #4), and diet or fat-reduced margarine brands (#5, #6, and #7); note also the outstanding position of the market leading brand #1. The smaller distances between the members of those 'natural' groups of brands are reflecting mutually higher degrees of competition and the positions of segment vectors relative to the brand positions are completing the picture by reflecting preference heterogeneity across households.

3.3. SOM results

While MULTICLUS configurations can be easily monitored by using likelihood based heuristics, approaches like SOM or TRN do not provide direct access to such information criteria for model selection. Instead, quantization error related criteria are consulted here. After SOM training, a final classification is determined by identifying the "best-matching" unit according to rule (3) during a "recall run" through the input data set. Analog to other techniques of exploratory data analysis, the suitability of SOM formats are determined heuristically, e.g., by examining "goodness-of-fit" measures against a sequence of shrinking SOM layers.

Two such measures are listed in Table 2: As a *heterogeneity measure*, the "mean-squared errors" (MSE) inform about relative distances between (raw, i.e. non-transformed and unstandardized) data points and corresponding prototype vectors w_{ij}. Since more and more dissimilar preference patterns are resembled together, total heterogeneity of cluster solutions should increase with decreasing numbers of prototypes.

The *simplicity measure* summarizes deviations of final prototypes from the pre-specified grid of units and is computed as "mean-squared inter-prototype distances" (MSIPD) for adjacent SOM units, i.e., distances between each prototype vector w_{ij} and it's immediately adjacent units w_{ij+1}.

Table 2
Heterogeneity, simplicity, VAF and replication values of SOMs

SOM layer	MSE	MSIPD	VAF	CVAF	A.Rand (total)	A.Rand (split half)
5×5	0.100	0.063	0.749	0.626	0.750	0.672
4×5	0.096	0.088	0.724	0.643	0.771	0.648
4×4	0.103	0.103	0.714	0.618	0.765	0.620
3×4	0.106	0.159	0.696	0.627	0.738	0.621
3×3	0.135	0.168	0.574	0.544	0.735	0.630
2×3	0.150	0.307	0.540	0.528	0.703	0.605
2×2	0.181	0.374	0.441	0.440	0.962	0.757
1×3	0.213	0.308	0.370	0.370	0.956	0.707
1×2	0.257	0.294	0.256	0.256	0.990	0.918

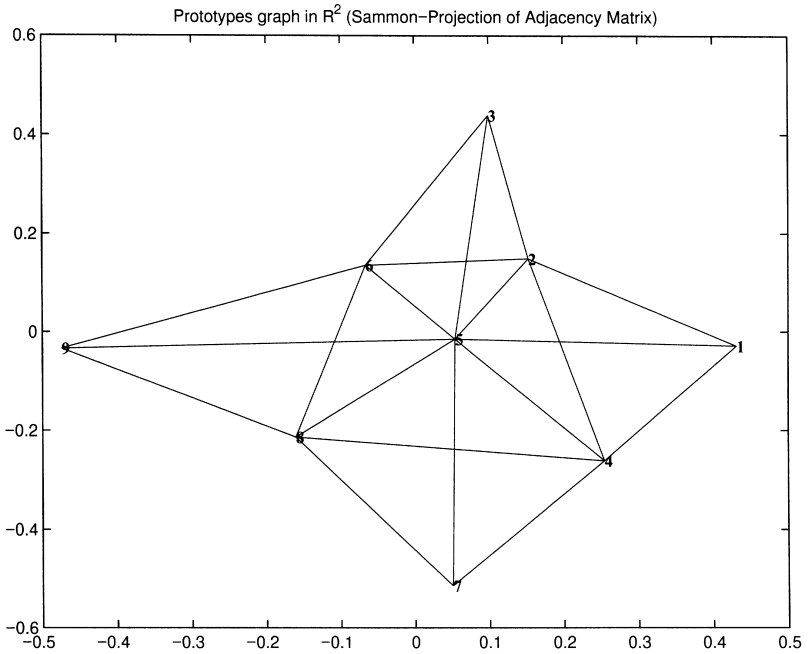

Fig. 5. Two-dimensional mapping of the SOM prototypes.

For comparison across different grid formats, the sum of distances is divided by the total number of possible neighborhood relations. The small simplicity values in Table 2 for layers with 9 and more units indicate a fairly "good" representation of the topological relationships between the prototypes or representing points. This property of final SOM configurations is also confirmed by the fairly good preservation of the SOM grid imposed adjacency structure as visualized in Fig. 5 by a two-dimensional Sammon [27] projection of the prototypes for a 3×3 SOM layer.

However, the topological quality of those maps is achieved at the cost of the centroid property of prototype vectors. Table 2 provides evidence for this *trade-off relationship* by comparing the "variance-accounted-for" (VAF) statistics of the partitions (within groups divided by total variance) with a "corrected VAF" (CVAF) measure adjusted for deviations of prototypes from respective class means: The spread between these two measures of total data recovery increases with improved topological quality of the map.

The issue of *cluster validity* is addressed in two ways here: First, as a measure of partition agreement the Hubert and Arabie [28] adjusted Rand-Index ('A.Rand') is computed between each pair of 30 replications of SOM training for the total sample. The considerably high averages reported in Table 2 suggest that SOMs seem to be insensitive to random initialization effects (the unadjusted Rand values reach values close to unity). Second, for each of the indicated SOM layers separate replication analyzes as proposed by Milligan [29] are conducted for 30 split-half random samples. Of course, mean adjusted Rand values do not reach those for the total sample. However, they achieve levels (unadjusted Rand values are all larger than 0.87) that indicate strong support for the assumption of SOM results stability.

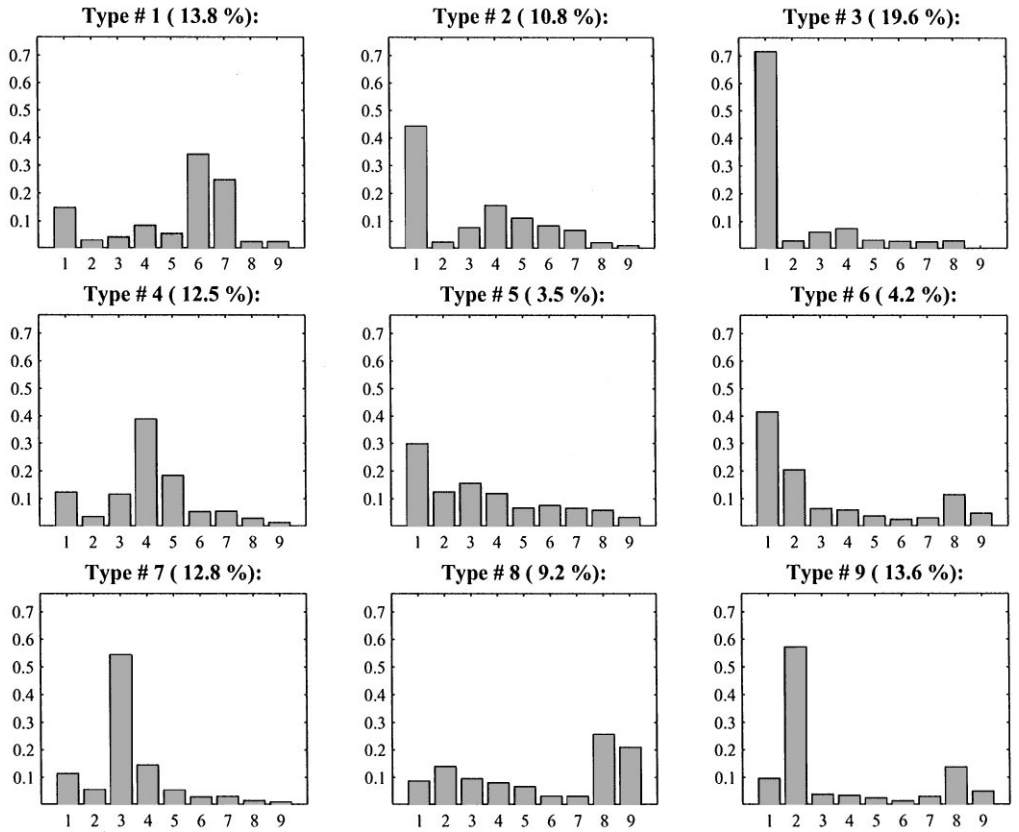

Fig. 6. SOM prototype profiles and relative segment sizes.

In contrast to the geometric MULTICLUS model, SOMs provide only ordinal adjacency information about various brand preferences along the discrete map of prototypes, which denote competitive relations between brands on a market segment level. According to Table 2, MSIPD scores "level off" when reducing an 3×3 format to an 2×3 SOM layer. Therefore, we turn to the 3×3 SOM solution as depicted in Fig. 6 for interpretation purposes.

Since each of the nine prototypes represents one segment of households (relative segment sizes are indicated in brackets), it is the combination of weight vector values (indicated in Fig. 6 by the magnitudes of the columns) that denotes the degree of competition intensity among rival brands in a specific sub-market: Segments with distinctive brand choice patterns (i.e., different types of CMS) are positioned in different directions or corners of the competitive map; e.g., households with extraordinarily strong brand preference for the general-purpose brand #1 are resembled in segment 3 (19.6 per cent). In the map's opposite corner, purchase behavior of members of segments 7 and 4 seem to be brand usage dominated (brands #3 and #4 are traditionally used for cooking and baking). Store loyal sub-markets for private labels (brands #2, #8, #9) are shifted to the more distant units positions 9 and 8, while segment 1 is characterized by above-average preferences

towards the diet brands #6 and #7. Other segments (of smaller size) represent "mixtures" of their adjacent prototypes.

Comparable to the MULTICLUS joint space, the outstanding effect of the market leading brand #1 can also be encountered here. Except for the nearly "brand-loyal" market segment 3, each of it's adjacent segments (of smaller relative sizes) shows purchase patterns with a notion of preference for different "co-brands" (e.g., the private labels #2 and #8 in segment 6, or the cooking&baking brands #3 and #4 in segments 5 or 2, respectively). Finally, low weights values of a brand throughout the map indicate a "fuzzy" preference position that might call for repositioning actions.

3.4. TRN results

For selection of a suitable TRN cluster solution, we can proceed analog to the above demonstrated screening of final SOM configurations with only one exception: The simplicity measure (MSIPD) used for SOM model selection does not provide any information for TRN results. This is due to the fact that the TRN model starts with a set of unordered units (i.e., there is no predetermined grid of units) and the adjacency structure between prototypes is constructed dynamically during the adaptation process.

Therefore, the analyst can restrict his investigations for model selection to the MSE (quantization error) and VAF measures. As indicated in Table 3 for a decreasing sequence of number of prototypes, MSE slightly levels off when constructing a partition on the basis of eight instead of nine TRN prototypes (also the associated VAF measures get considerably reduced as compared to finer partitions). Therefore, the properties of a nine prototype solution are commented below. Note, that the differences between VAF and corresponding measures corrected for deviations of

Table 3
Heterogeneity, VAF and replication values of TRN

Proto-types	MSE	VAF	CVAF	A.Rand (total)	A.Rand (split half)
16	0.069	0.776	0.765	0.948	0.633
15	0.071	0.779	0.765	0.874	0.645
14	0.073	0.779	0.762	0.857	0.640
13	0.077	0.770	0.745	0.815	0.647
12	0.080	0.765	0.742	0.814	0.664
11	0.082	0.759	0.737	0.824	0.684
10	0.088	0.727	0.716	0.954	0.654
9	0.091	0.729	0.715	0.890	0.662
8	0.109	0.673	0.667	0.893	0.636
7	0.128	0.614	0.612	0.859	0.635
6	0.136	0.582	0.575	0.944	0.661
5	0.156	0.574	0.570	0.828	0.586
4	0.181	0.457	0.456	0.946	0.784
3	0.213	0.375	0.375	0.904	0.752
2	0.257	0.255	0.255	0.919	0.942

Table 4
Statistical neighborhood of prototypes

Winner/ co-winner	Percentage of data points
4–6	27.0
1–8	13.4
2–6	6.4
2–5	6.1
5–6	5.9
3–6	5.0
3–9	4.2
1–7	3.7
6–9	2.9
3–7	2.8
6–8	2.0
6–7	1.9
—	—

prototypes from class centroids (CVAF) are much smaller here as compared to SOM results. This indicates that TRN prototypes are better representants of respective cluster means than the SOM weights are. Furthermore, the results generated by the TRN algorithm are very stable in terms of average adjusted Rand values for the training sample. The same applies to the split-half random sampling validity measures provided by the last column of Table 3.

After completion of TRN training, the adjacency structure between prototypes is determined by the final entries of the connection strength matrix. In addition, TRN delivers information about the frequencies of winner–co-winner relations between prototypes, which were responsible for establishing the connectivity structure during training and may be considered as indicators of the relative strengths of statistical neighborhood. Table 4 shows an extract of the decreasing order of percentages of data points involved in a winner–co-winner relationship in a recall run.

In terms of statistical neighborhood prototypes 4 and 6 are most often encountered in a winner–co-winner relationship which represents about 27% of the data, followed by the pair of prototypes 1, 8, etc. Trimming the adjacency graph at the value of 3.7 results in a configuration of eight pairwise prototype connections which form two substructures as portrayed in a two-dimensional Sammon projection in Fig. 7.

In analogy to the interpretation of SOM configurations again the combination of the prototype vector values given in Fig. 8 serves for detecting competitive intensity between the brands in the respective segments. For example, according to the profiles obtained for the above mentioned two strongly connected prototypes 4 and 6 both of them indicate segments of households with strong preferences for the market leading brand #1.

However, in contrast to SOM configurations after defining substructures of internally more adjacent prototypes the marketing analyst is now in a position to investigate competitive patterns between the rivaling brands both on a micro- and a more aggregated macro-segment level. Here, substructure 1 accounts for about 30% of the total market and is formed by the three adjacent

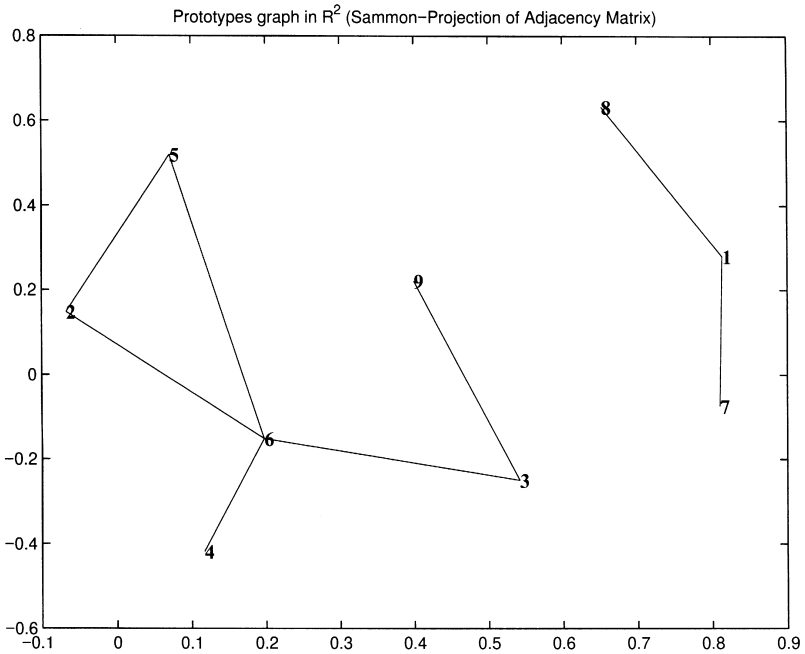

Fig. 7. Two-dimensional mapping of a pruned TRN adjacency graph.

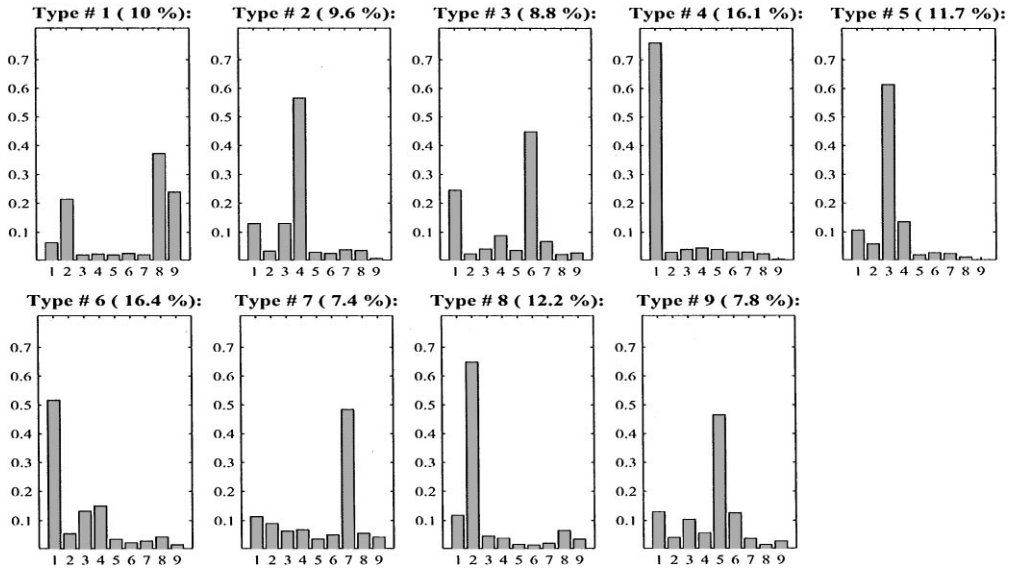

Fig. 8. TRN prototype profiles and relative segment sizes.

prototypes 1, 8, and 7. As the micro-level prototype profiles of segments 1 and 8 in Fig. 8 indicate, both represent store-loyal households with high preferences for private labels #2, #8, and #9. Note, that the connection between prototype 7 with an outstanding brand preference at the favor of

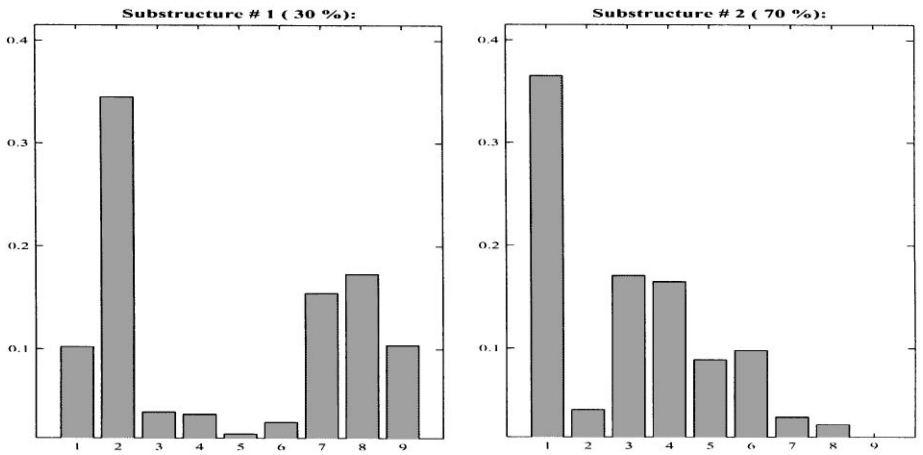

Fig. 9. Profiles of TRN substructure centroids.

the diet-brand #7 and prototype 1 is only weakly supported in terms of relative winner–co-winner frequencies (about 3.7% of the data) and therefore might also be regarded as a "stand-alone" prototype.

The resulting macro-level brand preference profiles for the two sub-structures can be examined by their respective centroid values as shown in Fig. 9. Sub-structure 2 represents about 70% of all households and is clearly dominated by the market leading brand #1. Finally, the shape of the branches of the adjacency graph itself provides another source of information about the similarities/dissimilarities of the various micro- or segment-level CMS. For example, prototypes 4 and 6 are much stronger connected in terms of winner–co-winner frequencies than prototypes 5 and 9 are, with the former representing a segment with high preferences for the baking margarine #3 and the latter for the fat-reduced margarine brand #5.

3.5. Comparison of results and validity issues

As illustrated above, SOM and TRN results differ from the joint space representation derived by the MULTICLUS model in several aspects:

First, in contrast to the more familiar geometric coordinate spaces, SOM and the more general TRN provide a discrete map of topologically ordered prototypes indicating the condensed competition among brands at a segment level. Thus, the concept of competitive "threat" has to be shifted from the usual interpretation in terms of spatial distances between brand positions and their (linear) segment-specific projections to an ordinal understanding of competitive relationship patterns. The combination (i.e., the relative differences) of final prototype values (see Figs. 6 and 8) indicates the intensity of brand competition within each of the segments. In addition to that, segments with more similar patterns of (revealed) brand substitutability are arranged adjacent to each other. Hence, CMS is a segment-specific concept in SOM and TRN analysis. Unlike the MULTICLUS results, segments derived by SOM/TRN do not share a common CMS with an unique configuration of the rival brands.

Second, the use of brand choice probabilities as input data makes segment-specific market shares directly accessible from SOM results. In contrast to the "fuzzy" clustering of preference data performed by MULTICLUS, SOM/TRN-based competitive maps provide a hard partition of the household mode. There is no need for any post hoc cluster assignment rules of households with respect to posterior probabilities. In SOM/TRN-like representations it is rather the brand mode that is portrayed in a "fuzzy" style; brands do not occupy single positions but assist in interpreting the prototypical patterns of purchase or brand-switching behavior in the various sub-markets.

Third, SOM/TRN-type analysis represents a more general approach for marketing data reduction. As previously shown by Mazanec [20] for three-way data, the non-parametric nature of SOM analysis also allows for compression of binary profile data. Finally, since only local information is required at each iteration, SOM training can be performed for data sets of unlimited size and/or "on-line" for continuously incoming data, such as retail scanner data.

In the above empirical applications all three models were tested with respect to stability of the resulting configurations. In contrast to the SOM and TRN partitions, MULTICLUS estimates are confronted with severe degeneration tendencies. For replicated SOM and TRN analyses using the total amount of data available, TRN delivers more stable configurations than the SOM model does (see columns 'A.Rand (total)' in Tables 2 and 3). This may be due to the rigid neighborhood updating concept applied in SOM analysis.

To validate the cluster solutions of the three models under study, a procedure proposed by Milligan [29] was applied for 30 split-half random samples each. As indicated in the last columns of Tables 2 and 3, TRN achieves slightly higher mean values for the used measure of partitions agreement (A.Rand: adjusted Rand index) as compared to SOM results. However, comparative validation of the MULTICLUS algorithm results in a mean adjusted Rand value of 0.516 for the above commented six dimensional nine clusters solution, which is significantly lower than the TRN nine prototype solution (mean A.Rand: 0.662).

4. Discussion

Our empirical analysis shows several benefits and shortcomings of the three methodologies under investigation, MULTICLUS, SOM, and TRN. As compared to MULTICLUS, we found the following advantages of using SOM or TRN for the task of combined market structuring and segmentation analysis: The non-parametric neural network approaches showed higher robustness against any kind of data preprocessing, a higher stability of the solutions in terms of a lower variance in the adjusted rand index of the estimation sample, and did not produce degenerated solutions. In terms of the cluster validation measure (adj. rand index for split half random samples) both TRN and SOM outperform MULTICLUS.

Comparing the SOM and TRN approach, we find advantages for TRN which uses a more flexible concept of adjacency structure. In TRN, no rigid grid of units must be specified. A further advantage of TRN lies in the possibility to exploit the information of the neighborhood graph which supports ex-post decisions about the segment configuration at both the micro and the macro level.

Of course, the SOM and TRN approaches have some drawbacks as compared to MULTICLUS. As a fully parametric model with known distributional properties, MULTICLUS configurations

are accessible to inferential statistics. MULTICLUS as a special case of the STUNMIX model family allows for a large variety of methodological options, such as the use of different types of preference models (ideal-point vs. ideal-vector unfolding models), internal vs. external analysis, simultaneous re-parameterization or property fitting (incorporation of brand descriptor variables) and simultaneous segment description. The latter two options are obtainable in TRN/SOM analysis in a two-stage procedure only (i.e., in a regression type analysis after completing training).

Our study indicates that non-parametric approaches such as TRN represent a useful exploratory tool for the marketing analyst — especially when distributional properties of the data analyzed are not known a priori.

Acknowledgements

The authors thank Kamel Jedidi for providing the MULTICLUS program and the guest co-editor, Jatinder Gupta, and the two anonymous referees for their helpful comments. Furthermore, we would like to express our gratitude to Josef Mazanec, Terry Elrod, Gilles Laurent and the discussants at the EMAC 1999 conference for comments on earlier versions of this manuscript.

References

[1] DeSarbo WS, Manrai AK, Manrai LA. Non-Spatial Tree Models for the Assessment of Competitive Market Structure: An Integrated Review of the Marketing and Psychometric Literature. In: Eliashberg J, Lilien GL, editors. Handbooks in operations research and management science, vol. 5, Marketing. Amsterdam: North-Holland, 1993. p. 193–257.
[2] Lilien GL, Kotler P, Moorthy SK. Marketing models. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[3] Kohonen T. Self-organized formation of topologically correct feature maps. Biological Cybernetics 1982;43:59–69.
[4] Martinetz T, Schulten K. Topology representing networks. Neural Networks 1994;7:507–22.
[5] Grover R, Srinivasan V. A simultaneous approach to market segmentation and market structuring. Journal of Marketing Research 1987;24:139–53.
[6] Wedel M, Kamakura WA. Market segmentation. Conceptual and methodological foundations. Dordrecht: Kluwer, 1998.
[7] Hruschka H. Market definition and segmentation using fuzzy clustering methods. International Journal of Research in Marketing 1986;3:117–34.
[8] Kamakura WA, Russell GJ. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 1989;26:379–90.
[9] Wedel M, Steenkamp J-BEM. A cluster-wise regression method for simultaneous fuzzy market structuring and benefit segmentation. Journal of Marketing Research 1991;28:385–96.
[10] Cooper LG, Klapper D, Inoue A. Competitive-component analysis: a new approach to calibrating asymmetric market-share models. Journal of Marketing Research 1996;33:224–38.
[11] DeSarbo WS, Manrai AK, Manrai LA. Latent class multidimensional scaling: a review of recent developments in the marketing and psychometric literature In: Bagozzi RP, editor. Advanced methods of marketing research Cambridge: Blackwell, 1994. pp. 190-222.
[12] Böckenholt I, Gaul W. Generalized latent class analysis: a new methodology for market structure analysis. In: Oppitz O, editor. Conceptual and numerical analysis of data. Berlin: Springer, 1989. pp. 367–76.
[13] DeSarbo WS, Howard DJ, Jedidi K. MULTICLUS: A new method for simultaneously performing multidimensional scaling and cluster analysis. Psychometrika 1991;56:121–36.

[14] Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B 1977;39:1–22.
[15] Wedel M, DeSarbo WS. An exponential-family multidimensional scaling mixture methodology. Journal of Business & Economic Statistics 1996;14:447–59.
[16] Allenby GM, Arora N, Ginter JL. On the heterogeneity of demand. Journal of Marketing Research 1998;35:384–9.
[17] Kohonen T. Self-organizing maps.. Berlin: Springer, 1995.
[18] Nour MA, Madey GR. Heuristic and optimization approaches to extending the Kohonen self organizing algorithm. European Journal of Operational Research 1996;93:428–48.
[19] Reutterer T. Competitive market structure and segmentation analysis with self-organizing feature maps. In: Andersson P, editor. Proceedings of the 27th EMAC Conference. Track 5: Marketing Research, 1998. p. 85–105.
[20] Mazanec JA. Positioning analysis with self-organizing maps: an exploratory study on luxury hotels. Cornell Hotel and Restaurant Administration Quarterly 1995;36:80–95.
[21] Bock HH. Simultaneous visualization and clustering methods as an alternative to Kohonen maps. In: Dell Riccia G, Kruse R, Lenz H-J, editors. Learning, networks and statistics. CISM Courses and Lectures no. 382. Vienna, New York: Springer, 1997. pp. 67–85.
[22] De Bodt E, Verleysen M, Cottrell M. Kohonen maps versus vector quantization for data analysis. In: Proceedings of the European Symposium on Artificial Neural Networks, Bruges, April 1997, p. 211–8.
[23] Mulier F, Cherkassky V. Self-organization as an iterative kernel smoothing process. Neural Computation 1995;7:1165–77.
[24] Martinetz T, Berkovich SG, Schulten K. "Neural gas" network for vector quantization and its application to time-series prediction.. IEEE Transactions on Neural Networks 1993;4:558–69.
[25] Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis: theory and practice. Cambridge: MIT Press, 1975.
[26] Green PE, Carmone FJ, Wachspress DP. On the analysis of qualitative data in marketing research. Journal of Marketing Research 1977;14:52–9.
[27] Sammon JW. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 1969;18(C):401–9.
[28] Hubert LJ, Arabie P. Comparing partitions. Journal of Classification 1985;2:193–218.
[29] Milligan GW. Clustering validation: results and implications for applied analysis. In: Arabie P, Hubert LJ, DeSoete G, editors. Clustering and classification. Singapore: World Scientific, 1996. pp. 341–75.

Thomas Reutterer is an Assistant Professor at the Department for Retailing and Marketing at the Vienna Univ. of Economics and Business Administration (VUEBA). In 2000 he is appointed Senior Lecturer at the Marketing Department of the University of Sydney. He holds a Master Degree and a Ph.D. in Marketing and Business Administration from VUEBA. His research interests include marketing and retailing research as well as business applications of marketing and management science methodology. He is a participant of the market segmentation and product positioning initiative of the interdisciplinary research project 'Adaptive Information Systems and Modeling in Economics and Management Science' (www.wu-wien.ac.at/am) sponsored by the Austrian Science Foundation (FWF) under grand SFB #10.

Martin Natter is an Associate Professor of Business Administration at the Vienna Univ. of Economics and Business Administration. He received a Masters Degree from the University of Vienna in Management Information Systems and a Ph.D. in Business Administration from the Vienna Univ. of Economics and Business Administration. His research interests are product development and marketing-operations interfaces. Some of his other research work has appeared in *Marketing Letters*, *Journal of Retailing and Consumer Services*, *European Journal of Operational Research*, *International Journal of Intelligent Systems in Accounting*, *Finance and Management* and *Lecture Notes in Economics and Mathematical Systems*.