A Growing Self-Organizing Neural Network for Lifestyle Segmentation

Reinhold Decker
Bielefeld University

Abstract: Lifestyles can be used to explain existent and to anticipate future consumer behavior, both in a geographical and a temporal context. Basing market segmentations on consumer lifestyles enables the development of purposeful advertising strategies and the design of new products meeting future demands. The present paper introduces a new growing self-organizing neural network which identifies lifestyles, or rather consumer types, in survey data largely autonomously. Before applying the algorithm to real marketing data we are going to demonstrate its general performance and adaptability by means of synthetic 2D data featuring distinct heterogeneity with respect to the arrangement of the individual data points.

Key words: Consumer typology, growing self-organizing neural network, lifestyle segmentation, marketing.

1. Introduction

1.1 Lifestyles and consumer typologies

Lifestyle is an important consumer characteristic and a determinant of individual purchasing behavior (Assael 1995; Engel, Blackwell, and Miniard 1995; Jobber and Lancaster 2003). It is represented by individuals' activities, interests and opinions. Cognitions from lifestyle segmentations influence marketing decisions in several ways, e.g. in media planning/selection and new product development for mass markets. The latter field of application can be motivated by the fact that many products and services are only successful in an economical respect, when they are purchased by thousands or even millions of people. Thus the focus of marketing researchers and marketers is on the identification of broad trends and patterns corresponding to the consumers' daily life, leisure behavior, and spending habits. Referring to plausible lifestyle segments proves to be more promising in marketing planning than the isolated use of demo- and psychographic data (Wedel and Kamakura 2000).

However, it should be recognized that lifestyles are continuously changing (Mowen and Minor 1998) and depend on cultural and sociological developments. New lifestyles develop from time to time. In this respect existing consumer typologies or segmentations, such as AIO ('activities, interests, and opinions') by Wells and Tigert (1971) or VALS ('values and life-styles') by the Stanford Research Institute (Mitchell 1983) are always a compromise between universality and specificity. The VALS typology, e.g., results from a market segmentation based on the individual's resources, mainly income and education, and self-orientation, i.e. attitude towards oneself, one's inspirations and the things one does to communicate and achieve them (Brassington and Pettitt 2005). In many cases an additional analysis of current primary data is indispensable to complete standardized information services such as VALS. This particularly applies if special consumer areas in daily life, such as nutrition behavior, are in the center of interest. Advances in automatic data production, e.g. by electronic micro test markets like IRI BehaviorScan or online panels, enable vivid portraits of different consumer types.

The reliable detection of meaningful lifestyle patterns and the empirical determination of consumer typologies from survey data is an elementary task of marketing research and contributes to a deeper understanding of existing relationships between products/services and consumers. Following Brassington and Pettitt (2005, p. 119) lifestyle segmentation methods 'can open the door to a better-tailored, more subtle offering to the customer on all aspects of the marketing mix'. Therefore different approaches, most of which apply traditional cluster or factor analysis, have been discussed in the past to address this topic. In both cases the analyst has to control the grouping process to a large extent. In hierarchical cluster analysis the selection of the similarity measure (e.g. the Tanimoto coefficient, which presents itself in case of binary coded consumer attributes), the fusion method (e.g. average linkage, where the distance between two clusters or segments is assumed to equal the average distance between pairs of observations, one in each cluster), and the criterion for determining an adequate number of clusters or consumer types (e.g. cubic clustering criterion according to Milligan and Cooper (1985)), influence the final structure of the resulting typology. Partitioning algorithms such as k-means, on the other hand, necessitate a pre-specification of the number of clusters (consumer types) to be considered. Similar problems arise from applying factor analysis as well, e.g. regarding the number of factors (consumer types) to be extracted or assumed, the communality estimation and the factor rotation. The adequate use of the respective algorithms requires extensive knowledge about the internal structure of the data to be analyzed. But the more intervention an algorithm requires from outside, the more subjectivity is involved regarding the final typology. Therefore, a methodology is desirable which

enables the determination of meaningful consumer typologies as autonomously as possible. In this respect self-organizing neural networks applying unsupervised learning techniques seem to be a promising alternative. The so-called prototypes (see subsection 2.1) resulting from the application of this kind of neural networks to appropriate data provide an almost 'natural' and easy-to-interpret basis for consumer typologies.

1.2 Related methodical work

Self-organizing neural networks have been an integral part of the data analytical instruments of natural and social sciences for several years. The spectrum of applications ranges from automatic image, text, and speech processing (Kohonen 2001) through the analysis of gas chromatographic patterns (Questier, Guo, Walczak, Massart, Boucon, and de Jong 2002), to financial data and industry analysis (DeBoeck and Kohonen 1998; Simula, Vasara, Vesanto, and Helminen 1999), qualification analysis in business administration (Wagner 2004), as well as market segmentation in e-commerce (Vellido, Lisboa, and Meehan 1999) and market basket analysis (Decker and Monien 2003).

As a consequence thereof this class of neural networks is the subject of continuing efforts for improvement. Corresponding research interests are devoted to both their individual design for specific areas of application and the elimination of existing methodological problems. Some of these problems, e.g. the flexible determination of network structure and size, have been solved or at least significantly reduced by algorithms such as the neural gas network (Martinetz and Schulten 1991), the growing neural gas network (Fritzke 1995), the growing self-organizing map (Villmann and Bauer 1998), the growing hierarchical self-organizing map (Dittenbach, Rauber, and Merkl 2002), and the grow-when-required network (Marsland, Shapiro, and Nehmzow 2002). In addition proposals with regard to a speeding-up of the adaptation process are already available for particular neural networks, e.g. for neural gas networks (Atukurale and Suganthan 2000). However, the efficient determination of adequate parameter settings continues to be a crucial practical problem, which has to be solved for each data set by a more or less troublesome trial-and-error process, if the internal structure of the data is unknown. This applies to neural networks with an a priori defined topology as well as to growing ones.

1.3 Motivation of the new approach

In contrast to natural and technical sciences, where unsupervised pattern detection has been an established subject of current research for a number of years, corresponding work is still in its infancy in marketing and consumer research.

This, among other things, results from the fact that the application of most of the available self-organizing neural networks requires considerable experience regarding the adequate specification of the relevant control parameters. Experiences of this kind are rarely available on the part of marketing practitioners.

Against this background we are going to introduce a self-organizing neural network approach, where the number of parameters to be preset by the user at the beginning of the data analysis process is limited to a minimum. The Growing Neural Network with Autonomous Parameter Specification (GNNAPS) presented in this paper determines the majority of the control parameters required for an adequate network adaptation more or less autonomously. To our knowledge the present study is the first to use a growing self-organizing neural network within the scope of lifestyle segmentation and provides an impression of the insights attainable in this way.

The remainder of the paper is organized as follows: In section 2 we briefly reflect some basic methodological aspects and outline the GNNAPS algorithm. Its performance and adaptability is demonstrated by means of a synthetic 2D data set in section 3 before real survey data collected from a German household panel are analyzed in section 4. The paper concludes with a brief discussion and an outlook on future research.

2. A Growing Neural Network with Autonomous Parameter Specification (GNNAPS)

2.1 Preliminary remarks

The methodology underlying the GNNAPS algorithm is vector quantization, the basic idea of which is to represent J K-dimensional input vectors (consumer profiles) $\boldsymbol{t}_j = (t_{j1}, \ldots, t_{jk}, \ldots, t_{jK})$, with $j \in \{1, \ldots, J\}$, by an adequate number H of weight vectors $\boldsymbol{\eta}_h = (\eta_{h1}, \ldots, \eta_{hk}, \ldots, \eta_{hK})$, with $h \in \{1, \ldots, H\}$. Each weight vector $\boldsymbol{\eta}_h$ defines one node ('neuron') of a neural network and will be interpreted in the following as the prototypical representative of a lifestyle segment. In the course of network adaptation each input vector (consumer profile) $\boldsymbol{t}_j$ is assigned to one weight vector or prototype $\boldsymbol{\eta}_h$ so that the distance between them is minimal. For $H < J$, which is the default, the assignment of J input vectors to H prototypes equals a compression of the data set considered (Gersho and Gray 1992). At the end of the network adaptation process the whole set of consumer profiles is represented by a set of 'prototypes' characterizing the individual lifestyle segments.

Striving for an algorithm, which requires as little as possible prior knowledge concerning the structure of the data to be analyzed, special attention is devoted to the autonomous or rather data-driven determination of network parameters. In

the present case only two control parameters have to be preset by the user, namely the so-called compression level CL and the maximum number of adaptation steps L considered to be necessary to adequately represent the available data. The former determines the extent of data reduction, whereas the latter influences the accuracy of data representation. To receive acceptable results L should be significantly larger than the number of observations J.

Setting CL close to 0 is tantamount to the objective of only slightly compressing the available data set and the permission of representing the relevant patterns by a comparatively large number of prototypes. If a compression level close to 1 is selected, only a few prototypes are generated, which involves a high level of generalization and the acceptance of comparatively large distances between the prototypes (weight vectors), if the data features heterogeneous patterns. In any case $0 < CL < 1$ must apply.

2.2 The algorithm

Self-initialization:

At the beginning of the adaptation process, i.e. in adaptation step $l = 1$, the neural network, or rather the associated set of nodes ('neurons') $\mathcal{U}$, contains only two non-connected nodes u_1 and u_2, i.e. $\mathcal{U} = \{u_1, u_2\}$ applies, with associated K-dimensional weight vectors $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$, both initialized with positive random values. The set of edges ('connections') $\mathcal{C}$ between nodes is empty. In the final neural network or rather its graphical visualization the edges are connecting those nodes which represent similar lifestyle segments. Both sets together determine the topological structure of the initial neural network. The size of these sets, and therewith the size of the whole network, grows in the course of network adaptation. This process comprises two aspects, namely the addition of new nodes and the adaptation of the weight vectors of already existing nodes. The latter is frequently equated with the term 'learning'.

The decision as to whether a new node has to be added to the current network or not in network adaptation step l depends on the extent to which the best matching node $u_{h_{Best}}$ fits the current input vector. The best matching node is the one whose weight vector has the smallest Euclidean distance to the current input vector. According to Marsland, Shapiro, and Nehmzow (2002) we call this the activity $v_{h_{Best}}$ of the best matching node. A formal definition is given later. The smaller the aforesaid distance, the higher the activity is. If the activity of the best matching node falls below threshold

$$v_{Thres} = \exp\left(\frac{S_{Max}}{2}\left((1 - CL)^{\frac{1}{4}} - 1\right)\right),$$

we can take this as a hint at an insufficient fit between the respective weight vector

and the input vector considered. Depending on the compression level v_{Thres} can have values between $\exp(-\frac{S_{Max}}{2})$, for $CL \rightarrow 1$, and 1, for $CL \rightarrow 0$, where S_{Max} denotes the maximum Euclidean distance between two input vectors. A simple but also quite rough approximation is $S_{Max} = \sqrt{\sum_{k=1}^{K}(\max_j t_{jk} - \min_j t_{jk})^2}$, where $\max_j t_{jk}$ and $\min_j t_{jk}$ are the maximum and the minimum value of all input vectors with respect to dimension k. A methodologically more elegant way of approximating S_{Max} is to calculate a random sample of Euclidean distances $D = \{d_{j_1}, \ldots, d_{j_i}, \ldots, d_{j_n}\}$, with $d_{j_i} = \sqrt{\sum_{k=1}^{K}(t_{j_i k} - t_{j_{i'} k})^2}$, for $j_i, j_{i'} \in \{1, \ldots, J\}$, $j_i \neq j_{i'}$, and $n < J$, from the available data set and to define $S_{Max} = \max\{d_{j_1}, \ldots, d_{j_n}\}$. In doing so the distribution of D can be found and an estimation of the standard deviation becomes possible. The approximation of S_{Max} spares us the time-consuming calculation of $\binom{J}{2}$ distances required for determining the true maximum Euclidean distance.

The addition of new nodes is internally controlled by two further variables, namely the firing counter $y_{h_{Best}}$ and the training requirement $w_{h_{Best}} \in [0, 1]$ of the best matching node. The former allows us to take into account how often $u_{h_{Best}}$ has been the best matching node in the network adaptation process so far. The more often a particular node matches the current input best, indicated by its firing counter, the lower its training requirement is. Both variables are initialized as follows: $y_1 = y_2 = 0$ and $w_1 = w_2 = 1$.

Similar to the addition of new nodes the learning process, i.e. the adaptation of the existent nodes (weight vectors) to the current input, is controlled by special variables as well. The learning rates ϵ_{Best} and ϵ_{Second} determine the degree to which the best and the second best matching node are adapted to the input. Restricting this process to two nodes is not unusual in the relevant literature (cf. e.g. Fritzke 1995). The higher the learning rates, the stronger the weight vectors concerned are adapted to the present data. For the best matching node we define:

$$\epsilon_{Best} = 0.1 + 0.5 \left(1 - \exp\left(-\frac{CL \cdot S_{Max}}{\sqrt{L}} \right) \right).$$

Learning rate ϵ_{Best} is positively correlated with compression level CL and maximum distance S_{Max} and it has an anchor point at 0.1. In that way we refer to the prerequisite that, as a rule, the larger the range of data, the faster the adaptation of weights should be. ϵ_{Best} decreases with an increasing maximum number of adaptation steps L, which equals a smoother and more moderate adaptation of the weight vectors. The learning rate of $u_{h_{Second}}$ corresponds directly to that of $u_{h_{Best}}$:

$$\epsilon_{Second} = \frac{\epsilon_{Best}}{\sqrt{L}}.$$

The smaller L, the more we have to extend the adaptation to the second best

matching node and vice versa. A small maximum number of adaptation steps causes the algorithm to adapt the topological neighbors of the best matching node to a high degree in order to achieve a fast generation of the neural network. A large L, on the other hand, causes a rather moderate adaptation of weights.

Starting from these presets one network adaptation step of the algorithm comprises the following sub-steps S1 - S10:

(S1) Input selection and determination of the best matching nodes:

An input vector $\boldsymbol{t}_j = (t_{j1}, \ldots, t_{jK})$, with $j \in \{1, \ldots, J\}$, is randomly selected from the data base, and the Euclidean distances to all nodes of the current neural network are calculated:

$$dist(\boldsymbol{t}_j, \boldsymbol{\eta}_h) = \|\boldsymbol{t}_j - \boldsymbol{\eta}_h\| = \sqrt{\sum_{k=1}^{K}(t_{jk} - \eta_{hk})^2} \qquad \forall\, h \text{ with } u_h \in \mathcal{U}.$$

The smallest distance of all determines the best matching node $u_{h_{Best}}$:

$$h_{Best} = \arg\min_{h \in \{1, \ldots, H\}} dist(\boldsymbol{t}_j, \boldsymbol{\eta}_h).$$

The second best matching node $u_{h_{Second}}$ follows from:

$$h_{Second} = \arg\min_{h \in \{1, \ldots, H\} \backslash h_{Best}} dist(\boldsymbol{t}_j, \boldsymbol{\eta}_h).$$

(S2) Calculation of the activity of the best matching node:

The activity of the best matching node is a function of its (Euclidean) distance to the current input vector:

$$v_{h_{Best}} = \exp\left(-dist(\boldsymbol{t}_j, \boldsymbol{\eta}_{h_{Best}})\right).$$

(S3) Calculation of the threshold for the training requirement:

Due to the fact that a node can adequately represent a subset of the data only after a certain number of adaptations, the training requirement has to fall below threshold

$$w_{Thres} = \epsilon_{Best} \cdot \left(1 - \frac{1}{\sqrt{|\mathcal{U}|}}\right)$$

before a new node is inserted. The larger the learning rate ϵ_{Best}, the higher this threshold is and the less intensively $u_{h_{Best}}$ has to be adapted before a new node may be added. The threshold increases with the number of nodes $|\mathcal{U}|$ in the network as well. That is to say the more nodes included in the network, the less intensive each of them has to be trained before a new node may be added.

The motivation of this is simple: The more advanced the adaptation process, the better the data set is represented by the available weight vectors. Due to the fact that new nodes are always inserted in the neighborhood of already existing ones, the former have to be trained the less intensively, the later they are added to the neural network.

(S4) Generation of connections between $u_{h_{Best}}$ and $u_{h_{Second}}$:

The nodes $u_{h_{Best}}$ and $u_{h_{Second}}$ will be connected, if either the distance between the respective weight vectors is so small that no further node can be inserted between them or at least one of both is not yet adapted sufficiently. In this case the age $a_{h_{Best}h_{Second}}$ of the concerning connection is set to 0:

$$\text{If } \exp\left(-\frac{dist(\boldsymbol{\eta}_{h_{Best}}, \boldsymbol{\eta}_{h_{Second}})}{2}\right) \geq v_{Thres} \text{ or } w_{h_{Best}} \geq w_{Thres} \text{ or } w_{h_{Second}} \geq w_{Thres}$$
$$\quad \text{then do}$$
$$\qquad \text{If } (h_{Best}, h_{Second}) \notin \mathcal{C} \text{ then do } \mathcal{C} = \mathcal{C} \cup \{(h_{Best}, h_{Second})\} \text{ end,}$$
$$\qquad a_{h_{Best}h_{Second}} = 0,$$
$$\quad \text{end.}$$

The above condition guarantees the generation of correct masked Voronoi polyhedra (Martinetz and Schulten 1994).

(S5) Addition of a new node:

If both the activity and the training requirement of the best matching node fall below the respective thresholds and if also the current number of nodes is smaller than the number of input vectors to be represented, then a new node or weight vector is added to the network. Formally, this procedure looks as follows:

$$\text{If } v_{h_{Best}} < v_{Thres} \text{ and } w_{h_{Best}} < w_{Thres} \text{ and } |\mathcal{U}| < J$$
$$\quad \text{then do}$$
$$\qquad \mathcal{U} = \mathcal{U} \cup \{u_{h_{New}}\},$$
$$\qquad \boldsymbol{\eta}_{h_{New}} = \tfrac{1}{2} \cdot (\boldsymbol{\eta}_{h_{Best}} + \boldsymbol{t}_j),$$
$$\qquad \mathcal{C} = \mathcal{C} \setminus (h_{Best}, h_{Second}),$$
$$\qquad \mathcal{C} = \mathcal{C} \cup \{(h_{New}, h_{Best}), (h_{New}, h_{Second})\},$$
$$\qquad y_{h_{New}} = 0, \quad w_{h_{New}} = 1, \quad y_{h_{Best}} = 0, \quad w_{h_{Best}} = 1,$$
$$\qquad \text{go to sub-step S7,}$$
$$\quad \text{end.}$$

By inserting the new node between the current input vector $\boldsymbol{t}_j$ and the best matching node $\boldsymbol{\eta}_{h_{Best}}$ the concerning Voronoi polyhedron is divided, which causes an improvement of data representation. The connections of nodes are changed

in such a way that the Delaunay triangulation (Martinetz and Schulten 1994) is preserved. The firing counter $y_{h_{New}}$ and the training requirement $w_{h_{New}}$ of the new node are initialized, whereas the corresponding control variables of the best matching node are reset.

(S6) Adaptation of weight vectors:

If no new node is added in sub-step S5 the weight vector of the best matching node as well as those of its topological neighbors are adapted by using learning rates ϵ_{Best} and ϵ_{Second}. For the best matching node we define:

$$\boldsymbol{\eta}_{h_{Best}} = \boldsymbol{\eta}_{h_{Best}} + \Delta\boldsymbol{\eta}_{h_{Best}} \quad \text{with } \Delta\boldsymbol{\eta}_{h_{Best}} = \epsilon_{Best} \cdot w_{h_{Best}}^{\frac{\ln(l+\exp(1))}{\ln(L+\exp(1))}} \cdot (\boldsymbol{t}_j - \boldsymbol{\eta}_{h_{Best}}).$$

With exponent $\ln(l + \exp(1))/\ln(L + \exp(1))$ the training requirement is related to the current state of network adaptation. At the beginning of the training process the strength of adaptation declines slowly, but the speed increases the more the algorithm approaches the maximum number of adaptation steps L. So the nodes have more time to fit the given data structure. Weighting the extent of adaptation resulting from $(\boldsymbol{t}_j - \boldsymbol{\eta}_{h_{Best}})$ with the training requirement provokes that a node is adapted the less intensively, the more frequently this took place in the past. For the neighboring nodes the basic form is sufficient:

$$\boldsymbol{\eta}_{h_i} = \boldsymbol{\eta}_{h_i} + \Delta\boldsymbol{\eta}_{h_i} \quad \text{with } \Delta\boldsymbol{\eta}_{h_i} = \epsilon_{Second} \cdot w_{h_i} \cdot (\boldsymbol{t}_j - \boldsymbol{\eta}_{h_i}) \quad \forall i \text{ with } (h_{Best}, h_i) \in \mathcal{C}.$$

Accordingly the extent of adaptation declines faster for the neighboring nodes than for the best matching one.

(S7) Update of control variables:

The firing counter and the training requirement of the best matching node are updated as follows:

$$y_{h_{Best}} = y_{h_{Best}} + 1 \qquad \text{and} \qquad w_{h_{Best}} = \frac{1}{y_{h_{Best}} + 1}.$$

(S8) Removal of old connections:

All connections between the best matching node and its topological neighbors are aged by 1:

$$a_{h_{Best}h_i} = a_{h_{Best}h_i} + 1 \qquad \forall (h_{Best}, h_i) \in \mathcal{C}.$$

The current maximum age of a connection is $a_{Max} = |\mathcal{U}| - 1$ and equals the maximum number of edges emanating from a node u_h, if all the other nodes are topological neighbors of u_h. The latter results, if u_h is always the best or second

best matching node, and if the probability of becoming the best or second best matching node is equal for all the other nodes. All connections exceeding the maximum age, and ergo being dispensable, are removed:

$$\mathcal{C} = \mathcal{C} \setminus (h_i, h_{i'}) \qquad \forall\, h_i, h_{i'} \text{ with } (h_i, h_{i'}) \in \mathcal{C}, \text{ and } a_{h_i h_{i'}} > a_{Max}.$$

(S9) Removal of nodes:

All the nodes without a connection to any other node and whose contribution to the goodness of data representation is negligible are removed:

$$\mathcal{U} = \mathcal{U} \setminus u_{h_i} \qquad \forall\, h_i \text{ with } (h_i, h_{i'}) \notin \mathcal{C}, \; h_i \neq h_{i'} \text{ and } y_{h_i} < |\mathcal{U}|.$$

(S10) Check of the stopping criterion:

After having updated the adaptation step counter according to $l = l + 1$ the stopping criterion $l > L$ is checked. If this holds, the algorithm stops and the connection matrix $\boldsymbol{C} = (c_{h_1 h_2})_{h_1, h_2 = 1, \ldots, H = |\mathcal{U}|}$ is generated with

$$c_{h_1 h_2} = \begin{cases} 1, & \text{if } (h_1, h_2) \in \mathcal{C} \\ 0, & \text{otherwise.} \end{cases}$$

Otherwise it continues with sub-step S1.

2.3 Measures of network performance

At the end of the adaptation process each node u_h represents a non-empty subset of data points or rather a lifestyle segment. The goodness of data representation within a segment can be assessed using the maximum distance MD_h between the data points (consumer profiles) concerned and the associated weight vector (prototype) $\boldsymbol{\eta}_h$. Considering the maximum of maxima across all segments leads to the so-called maximum distance error

$$MDE = \max_h \{MD_h\} \qquad \text{with } MD_h = \max_{j \mid m_j = h} dist(\boldsymbol{t}_j, \boldsymbol{\eta}_{m_j}) \qquad \forall\, h,$$

where $m_j = \arg\min_{h \in \{1, \ldots, H\}} dist(\boldsymbol{t}_j, \boldsymbol{\eta}_h)$ is the subscript of the weight vector which best matches input vector $\boldsymbol{t}_j$. So MDE refers to the worst match of a consumer profile to its associated prototype and assesses the balance of segmentation. A very popular performance measure is the quantization error (Kohonen 2001)

$$QE = \sum_{j=1}^{J} dist(\boldsymbol{t}_j, \boldsymbol{\eta}_{m_j}),$$

which assesses the distortion due to data compression and the effectiveness of node placement in the neural network (Yerramalla, Cukic, and Fuller 2003). Both measures should take lowest possible values.

In lifestyle segmentation, as can be seen in the empirical part of this paper, we are not only interested in good data representation but also in an adequate network topology. The latter is related to the number of nodes and the lengths of the connections in the graph by which the neural network is represented. The fewer nodes this graph has and the shorter the paths between individual nodes (expressed by the edges or connecting lines), the more compact the neural network is. In other words neighboring nodes should represent similar lifestyle segments and should therefore be connected directly by an edge, whereas those nodes, which are located further away from each other, should not. Thus, in the ideal case, a well-trained neural network represents both the existent lifestyle patterns and the similarity of these patterns. We are going to demonstrate this with the synthetic 2D data in section 3. Referring to Marsland, Shapiro, and Nehmzow (2002) the following simple measure of compactness can be defined

$$C1 = \sum_{h_1=1}^{H} \sum_{h_2=h_1+1}^{H} c_{h_1 h_2} \cdot dist(\boldsymbol{\eta}_{h_1}, \boldsymbol{\eta}_{h_2}) \qquad \text{with } dist(\boldsymbol{\eta}_{h_1}, \boldsymbol{\eta}_{h_2}) = \|\boldsymbol{\eta}_{h_1} - \boldsymbol{\eta}_{h_2}\|,$$

where $c_{h_1 h_2}$ indicates whether nodes u_{h_1} and u_{h_2} are connected ($c_{h_1 h_2} = 1$) or not ($c_{h_1 h_2} = 0$). To facilitate comparisons of different adaptation runs regarding the compactness of the respective neural networks $C1$ can be related to the number of connections:

$$C2 = \frac{1}{|\mathcal{C}|} \cdot C1.$$

Again, both measures should take lowest possible values.

For more global comparisons we can additionally consider the geometric mean GM of the above performance measures. Due to the congruence of the orientation of the individual measures GM should be as small as possible, too.

3. Performance Study with 2D Synthetic Data

The performance and adaptability of the GNNAPS approach is analyzed by means of a 2D synthetic data set, which is similar to that used by Martinetz and Schulten (1991) as well as Fritzke (1995). It contains $J = 19\,160$ data points (input vectors), which define manifold graphical objects. Figure 1 illustrates the whole data set. Each object is characterized by the number J_i (with $J = \sum_{i=1}^{7} J_i$) of its data points.

$J_1 = 800$

$J_2 = 1\,600$ $J_3 = 8\,000 = 5 \cdot 1\,600$ $J_4 = 7\,260$

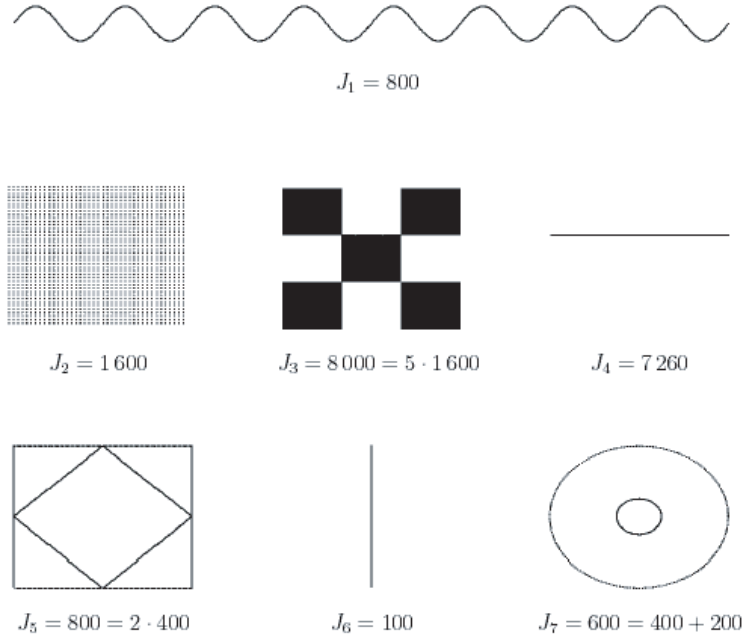$J_5 = 800 = 2 \cdot 400$ $J_6 = 100$ $J_7 = 600 = 400 + 200$

Figure 1: Visualization of the 2D synthetic data set

The two single lines in particular are worth closer consideration. The significantly different numbers of data points ($J_4 = 7\,260$ vs. $J_6 = 100$) result from the varying 'density' of the horizontal line on the right hand side. The more one goes to the right on this line, the more data points are represented by the respective section of the line. In contrast to this each of the five black rectangles in the 'chessboard pattern', as well as the dotted rectangle on the left hand side, have been generated with $1\,600$ data points only. This topological heterogeneity of the input data is a special challenge to the algorithm's adaptability.

If we apply the GNNAPS algorithm to this data with $CL = 0.15$ and if we restrict the maximum number of adaptation steps to $L = 10^7$, the neural network with $H = 181$ nodes depicted in Figure 2 results. The corresponding performance measures are $QE = 17\,343.50$, $MDE = 2.49$, $C1 = 801.62$, and $C2 = 3.09$, which leads to a geometric mean $GM = 101.70$. The dots and lines denote the nodes and edges (connections) of the neural network. The larger a node is drawn, the more input vectors are represented by the respective weight vector. Obviously all graphical objects, including the horizontal 'density line' on the right hand side, are represented adequately by the available set of nodes.
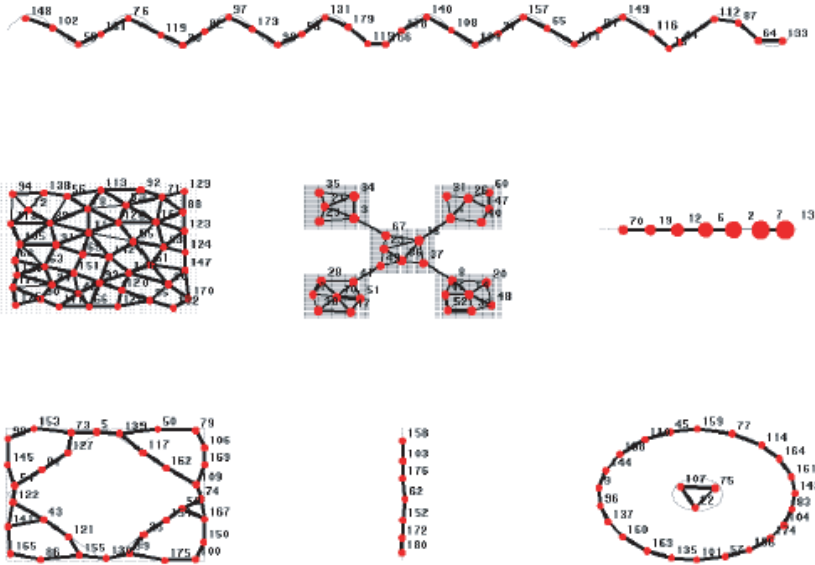
Figure 2: Data representation with GNNAPS ($CL = 0.15$ and $L = 10^7$)

To draw comparisons we additionally applied the growing neural gas network (GNGN) suggested by Fritzke (1995) to the present data. The GNGN algorithm is a powerful benchmark because of its impressive results in earlier studies, e.g. in a comprehensive comparison with k-means, growing k-means, and the original neural gas network published by Daszykowski, Walczak, and Massart (2002). Initializing the GNGN algorithm with control parameters similar to those used by Fritzke (1995), and fixing the step width for adding new nodes equal to $55\,249$ in combination with a maximum number of adaptation steps $L = 10^7$, results in a pattern representation with performance values $QE = 12\,422.49$, $MDE = 3.83$, $C1 = 870.44$, $C2 = 2.89$, and $GM = 104.60$. The above-mentioned step width ensures the comparability of results by causing the GNGN algorithm to generate $H = 182$ nodes on the whole. The corresponding neural network is depicted in Figure 3.

Again the attained representation looks adequate. In particular the representation of the 'chessboard pattern' is without doubt convincing. But the data pattern underlying the horizontal 'density line' claims a comparatively large number of nodes, which slightly hampers the representation of the graphical objects in the lower right and left hand corner. The representation of the large rectangle equals the GNNAPS solution. The visual conformity of both representations is reflected in the individual performance measures as well. The superiority of GNGN over GNNAPS regarding QE and $C2$ (with values $12\,422.49$ vs. $17\,343.50$

and 2.89 vs. 3.09) is opposed to the superiority of GNNAPS over GNGN regarding MDE and $C1$ (with values 2.49 vs. 3.83 and 801.62 vs. 870.44). Considering the geometric mean one might conclude a slight superiority of GNNAPS over GNGN, which results from the fact that the balance in pattern representation turns out to be somewhat better with GNNAPS.
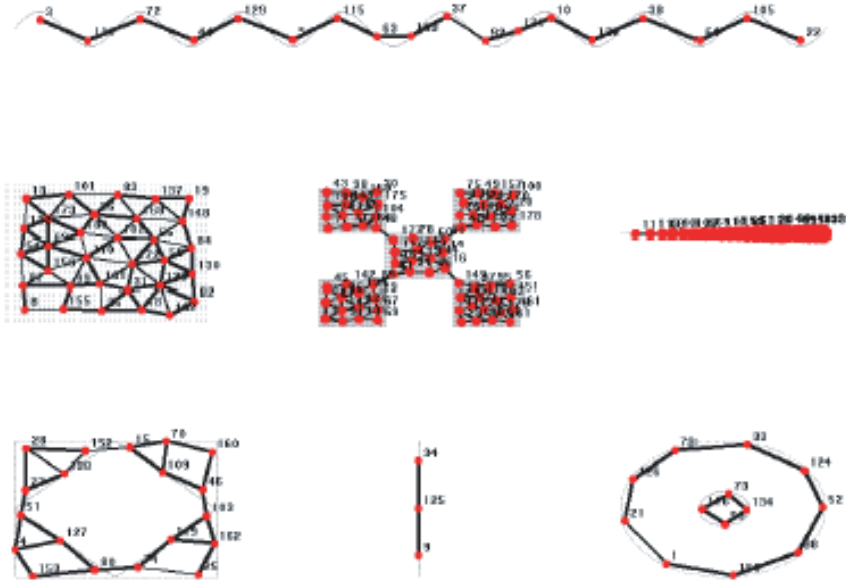

Figure 3: Data representation with GNGN ($L = 10^7$)

To get a closer impression of how GNNAPS learns the given topological relations, the performance measures for an increasing number of adaptation steps with compression level $CL = 0.15$ are shown in Table 1. The H values in the second column quickly converge to their final level of 181. However GNNAPS does not increase the number of nodes (starting with $H = |\mathcal{U}| = 2$) monotonously to achieve an acceptable representation of the data, but also reduces them, if this is justifiable with the current stage of adaptation. The GNGN algorithm lacks this flexibility per definition. At the same time the GM value of GNNAPS declines continuously. As few as 10^6 (i.e. $\approx 50 \cdot J$) adaptation steps are sufficient to approach a relatively stable representation of the relevant patterns. Larger numbers of adaptation steps primarily cause a kind of fine-tuning.

Table 1: Performance of GNNAPS and GNGN (in brackets) on the 2D data

L	H	QE	MDE	$C1$	$C2$	GM
10^3	71	29 605.96	10.70	970.15	8.51	226.14
	(73)	(57 254.56)	(29.43)	(809.16)	(6.73)	(309.50)
10^4	172	19 386.00	4.47	1 441.52	4.85	156.89
	(174)	(17 607.17)	(12.32)	(1 250.63)	(4.22)	(183.94)
10^5	187	18 394.51	2.49	1 139.28	3.91	119.52
	(186)	(12 958.07)	(4.60)	(896.91)	(2.98)	(112.35)
10^6	178	18 155.80	2.55	822.49	3.30	105.88
	(179)	(12 691.01)	(3.73)	(866.46)	(2.96)	(104.97)
$5{\cdot}10^6$	179	18 762.75	2.50	803.48	3.13	104.22
	(180)	(12 427.32)	(3.85)	(846.39)	(2.89)	(104.01)
10^7	181	17 343.50	2.49	801.62	3.09	101.70
	(182)	(12 422.49)	(3.83)	(870.44)	(2.89)	(104.60)

For comparison purposes the performance measures of GNGN are given in brackets. For each L the step width for adding new nodes was selected in such a way that the GNGN solution is comparable to that of GNNAPS, as far as possible. Once again the parametrization of GNGN was aligned with suggestions by Fritzke (1995). The impressive performance of GNGN with respect to the quantization error QE is opposed to the superiority of GNNAPS regarding the maximum distance error MDE. The more advanced the adaptation process, the better the measure of compactness $C1$ turns out to be with GNNAPS, whereas $C2$ levels off close to 3 for both algorithms.

The differences between both algorithms with respect to QE and MDE can easily be motivated. The GNGN algorithm minimizes the quantization error, i.e. feature spaces showing a high density of data points are represented very well (cf. the 'chessboard pattern'), whereas the contrary applies to feature spaces with a comparatively low density of data points (cf. the graphical objects in the lower left and right hand corner). The GNNAPS algorithm, on the other hand, also minimizes the maximum distance error MDE. Thus, it is less susceptible to data heterogeneity and attains a more balanced pattern representation. Moreover, since the number of nodes generated by the GNGN algorithm directly depends on the number of input signals and the selected step width for adding new nodes, its use requires some experience regarding the definition of the latter. Otherwise the parametrization may degenerate into a troublesome trial-and-error process, if the internal structure of the data is unknown. In this respect the adaptation of GNNAPS is self-controlled to a high degree, which eases its use by non-specialists, e.g. in marketing and consumer research.

4. Application to Real Survey Data

4.1 The data

In this section we are going to use the new algorithm to empirically determine a consumer typology from real survey data. The data underlying the following considerations was made accessible by the German ZUMA Institute and is part of a sub-sample of the 1995 GfK ConsumerScan Household Panel Data. For a detailed description of this data set see Papastefanou, Schmidt, Börsch-Supan, Lüdtke, and Oltersdorf (1999). It contains socio-economic and demographic characteristics of several thousands of households/consumers as well as individual attitudes towards nutrition (e.g. slimness orientation, plain fare, and brand products), aspects of daily life (e.g. traditional living, convenience-orientated cooking, and mistrust towards new products), environment (e.g. ecological awareness, mobility, and industry), and shopping (e.g. tendency to purchase new products, price consciousness, and preference for small retail stores). A considerable number of the respective statements or items are more or less concerned with individual nutrition behavior (e.g.: *'Multivitamin juices are an important supplement to daily nutrition.'*). The special relevance of food-related lifestyle analysis has already been emphasized by Grunert, Brunsø, and Bisp (1997), de Boer, McCarthy, Cowan, and Ryan (2004) and others. Hollensen (2004), in particular, states that food consumption habits may even be used as a general indicator of lifestyle in international or global marketing. To identify existing lifestyle patterns we consider the attitudes of $J = 4\,266$ households/consumers measured by means of $K = 81$, mostly Likert-scaled items (with $1 \equiv$ 'I definitely disagree.', ..., $5 \equiv$ 'I definitely agree.'). The scale is assumed to be equidistant, and therefore the data can be treated as metric.

4.2 Selected results

From a practical point of view the given task requires a comparatively high degree of generalization to get an easy-to-grasp set of prototypes representing reasonable groups of consumers. Therefore GNNAPS was applied with a high compression level, namely $CL = 0.98$. Together with a maximum number of adaptation steps $L = 10^7$ we get $H = 14$ prototypes involving performance values $QE = 36\,355.28$, $MDE = 16.67$, $C1 = 193.71$, and $C2 = 5.87$. Fixing CL close to 1 causes a strong compression of the data, and ergo, provides a small number of prototypes or consumer types respectively.

The numbers of consumers represented by weight vectors $\boldsymbol{\eta}_1 = (\eta_{11}, \ldots, \eta_{1,81})$, ..., $\boldsymbol{\eta}_{14} = (\eta_{14,1}, \ldots, \eta_{14,81})$ range between 129 and 514. Each 'lifestyle prototype' can be described by considering its specific attitude profile. As to be expected

Table 2: Profiles of selected 'lifestyle prototypes'

Items considered in the survey	η_1	η_5	η_{10}
I like to have company.	4.54^{++}	3.96	3.49^{--}
I spend my spare time very actively.	4.20^{++}	3.69	2.52^{--}
Instead of saving one should spend money to improve life.	3.65^{++}	2.99	1.95^{--}
I want to enjoy my life to the full.	3.60^{+}	3.21	1.74^{--}
I like the atmosphere in small shops.	4.21^{+}	3.50	3.12^{-}
I like to try new products.	3.87^{+}	3.57	2.58^{-}
Many products that I buy are still unknown to other consumers.	3.10^{+}	2.60	1.67^{--}
I'm always on the lookout for new products which suit my requirements better.	3.25^{++}	2.57	1.59^{--}
I am not willing to do without individual service when shopping.	4.10^{++}	3.05	2.80^{-}
I often buy products that are new on the market sooner than my friends.	2.93^{+}	2.47	1.52^{--}
I watch my figure when eating and drinking.	4.35^{++}	3.88	3.18
I would characterize myself as being very slimness-orientated.	3.56^{++}	2.99	2.05^{-}
Multivitamin juices are an important supplement to daily nutrition.	4.24^{++}	3.75	3.29
I frequently use vitamin and mineral products to keep fit.	3.73^{++}	2.84	1.79
I only eat vegetarian food.	1.73	1.41	1.09^{--}
I often treat myself to a delicacy.	3.64^{+}	2.68	2.11
If one believed in all nutrition tips one couldn't eat or drink anything.	4.63^{++}	4.02	4.23

some of the respective weights differ more across the available prototypes than others. To create expressive consumer types it is advisable to focus on those items which show the clearest differences across all prototypes. In Table 2 the prototypes of three exemplary groups of consumers are depicted. Each prototype is represented by its weight vector, i.e. η_1, η_5 and η_{10}. Item weights equaling the maximum or minimum across all prototypes are marked with $++$ and $--$, whereas signs $+$ and $-$ denote the second largest and smallest values. The 'extreme values' indicate a clear approval or disapproval regarding the corresponding

statement. For the sake of clarity only those items have been listed which have an 'extreme value' either for $\boldsymbol{\eta}_1$ or $\boldsymbol{\eta}_{10}$.

Prototype $\boldsymbol{\eta}_1$ represents about 7.15 % (= 305) consumers with a rather hedonistic way of life. Weight $\eta_{13} = 4.54^{++}$ in the first row of the table, for example, indicates their clear agreement, on average, with the statement '*I like to have company.*'. Consumers of this type like to try new and unknown things and are willing to take care of their body by attaching great importance to healthy nutrition. However they do not show a disposition to the extreme such as pure vegetarian diet. This applies to all groups with the exception of number 9 where a distinct refusal of 'normal' consumer behavior can be observed. Members of this group have maximum weights on items such as '*I regard most new products on the market as unnecessary.*' ($\eta_{91} = 3.95^{++}$) and '*I am very distrustful of advertising messages.*' ($\eta_{9,11} = 4.43^{++}$). They strictly disagree with the statement '*There is too much fuss about diets.*'. Members of the first group are called the 'hedonists' in the following.

Those 166 consumers ($\equiv$ 3.89 %) that are represented by prototype $\boldsymbol{\eta}_{10}$ may be characterized as conservative and less health-orientated with a frugal diet and lifestyle. They show little interest for new products and strictly decline extreme forms of nutrition such as vegetarianism. We label this group the 'conservatives' in the following. Likewise the 'hedonists' they tend to question whether it helps much trying to take into account all nutrition tips (see last row of Table 2). In this respect both consumer groups stand out against those represented by prototype $\boldsymbol{\eta}_5$. Due to the fact that this prototype represents the largest group of consumers, namely 514, we can use it as a reference object. This becomes apparent by the average level of the individual weights, which lie mostly between those of prototype $\boldsymbol{\eta}_1$ and prototype $\boldsymbol{\eta}_{10}$. The comparatively low value of group 5 regarding the last item results from the less distinct profile of these 'average consumers'. Moreover, none of the 81 weights of prototype $\boldsymbol{\eta}_5$ equals the maximum or minimum across all groups.

Further items which strongly discriminate between the 14 prototypes are, e.g., the relevance of product quality in purchase decisions, the preference for brand products, the time spent to prepare a meal, the attention paid to mild/non-irritant foods, the purchase of products without additives, and the preference for domestic goods. The straight interpretability of the individual weights is a considerable advantage over traditional segmentation tools such as hierarchical cluster analysis and multidimensional scaling.

To get a global impression of the relations between the individual prototypes we can look at the 2D projection of the neural network. The connection graph in Figure 4 visualizes the topological structure of the network underlying the lifestyle segmentation. The relative size of the nodes indicates the number of consumers

being represented by the respective prototype. The age of the connections existing in adaptation step L is represented by the thickness of the corresponding edges in the connection graph. By this means the strength of the relations between the prototypes or lifestyle segments is depicted. The thicker an edge is drawn, the stronger the respective relation is.
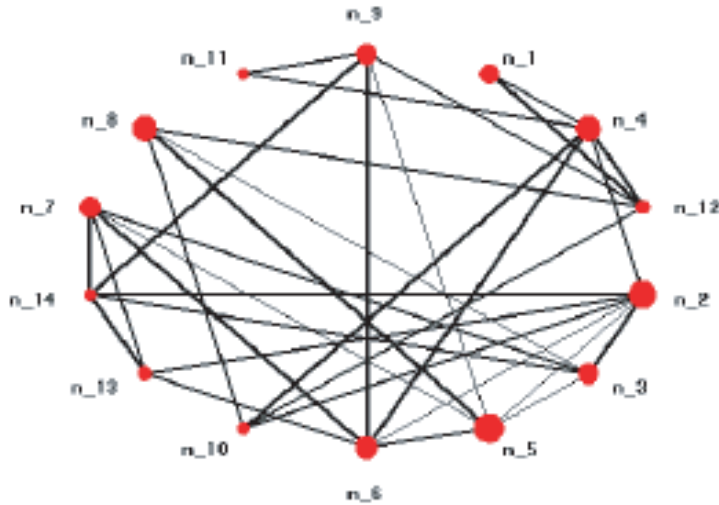

Figure 4: Connection graph (with $|\mathcal{C}| = 33$)

From Figure 4 we learn that there is no direct connection (edge) between prototypes $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_{10}$ which is in accordance with the dissimilarities between both groups of consumers regarding their lifestyles. The missing connection to the reference object $\boldsymbol{\eta}_5$ in both cases is plausible as well. The connection graph as a whole expresses the complexity of the inter-segment relationships in the considered population.

4.3 Managerial implications

Lifestyle analyses by means of GNNAPS are useful for decision making in fields like product positioning and promotion strategy development (Mowen and Minor 1998). The empirical results motivate, for instance, the creation of new top quality food products with a distinct health-orientation. Top quality should also include high standard taste requirements. This is conformable with Jago (2000) who argues that health is one of the most important drivers of new product development in food industry. Pronounced associations with vegetarian food

should however be avoided here. Products of this type would precisely meet the preferences of a considerable 7.15 % of the population represented by the available data. The remaining consumer types or lifestyles can be analyzed in an analogous manner with respect to their implications for new product development or advertising planning for instance.

5. Conclusions

In this paper we have introduced a new algorithm which has proven its ability to detect heterogeneous data patterns with a comparatively small number of parameters to be controlled by the user. In contrast to the established GNGN approach, which features significantly more parameters to be preset, GNNAPS works highly satisfactorily with two external control parameters (L and CL) only. The process of determining the size and the structure of the neural network is quite autonomous in this respect. The comparatively small number of external parameters facilitates the use of the new algorithm in practice, e.g. in exploratory data analysis and marketing research, and reduces the total time required for network adjustment.

In the empirical study we were able to show that the principle of computing prototypes from consumer behavior data corresponds directly to classification tasks in marketing research, and particularly in lifestyle segmentation. The poor requirements of the algorithm regarding the prior knowledge about the internal structure of the survey data enabled an exploratory analysis in the primary sense. In the present example a straight interpretation of the individual weights towards future food preferences and appropriate managerial implications was possible.

Acknowledgement

References

Assael, H. (1995). *Consumer Behavior and Marketing Action* (5th ed.). South-Western College Publishing.

Atukorale, A. and Suganthan, N. (2000). Hierarchical overlapped neural gas network with application to pattern classification. *Neurocomputing* **35**, 165-176.

Brassington, F. and Pettitt, S. (2005). *Essentials of Marketing*. Prentice Hall.

Daszykowksi, M., Walczak, B., and Massart, D. L. (2002). On the optimal partitioning of data with k-means, growing k-means, neural gas, and growing neural gas. *Journal of Chemical Information and Computer Sciences* **42**, 1378-1389.

De Boer, M., McCarthy, M., Cowan, C., and Ryan, I. (2004). The influence of lifestyle characteristics and beliefs about convenience food on the demand for convenience foods in the Irish market. *Food Quality and Preference* **15**, 155-156.

DeBoeck, G. and Kohonen, T. (1998). *Visual Explorations in Finance.* Springer.

Decker, R. and Monien, K. (2003). Market basket analysis with neural gas networks and self-organizing maps. *Journal of Targeting, Measurement and Analysis for Marketing* **11**, 373-386.

Dittenbach, M., Rauber, A., and Merkl, D. (2002). Uncovering hierarchical structure in data using the growing hierarchical self-organizing map. *Neurocomputing* **48**, 199-216.

Engel, J. F., Blackwell, R. D., and Miniard, P. W. (1995). *Consumer Behavior* (8th ed.). Dryden Press.

Fritzke, B. (1995). A growing neural gas network learns topologies. In *Advances in Neural Information Processing Systems 7* (Edited by G. Tesauro, D. S. Touretzky, and T. K. Leen). MIT Press, 625-632.

Gersho, A. and Gray, R. M. (1992). *Vector quantization and signal compression.* Kluwer.

Grunert, K. G., Brunsø, K., and Bisp, S. (1997). Food-related lifestyle: Development of a cross-culturally valid instrument for market surveillance. In *Values, Lifestyles and Psychographics* (Edited by L. Kahle and L. Chiagouris). Erlbaum, 337-354.

Hollensen, S. (2004). *Global Marketing: A Decision-oriented Approach* (3rd ed.). Prentice Hall.

Jago, D. (2000). A healthy outlook. *Prepared Food* **169**, 29.

Jobber, D. and Lancaster, G. (2003). *Selling and Sales Management* (6th ed.). Prentice Hall.

Kohonen, T. (2001). *Self-Organizing Maps* (3rd. ed.). Springer.

Marsland, S., Shapiro, J., and Nehmzow, U. (2002). A self-organising network that grows when required, *Neural Networks* **15**, 1041-1058.

Martinetz, T. and Schulten, K. (1991). A 'neural gas' network learns topologies. In *Artificial Neural Networks* (Edited by T. Kohonen, K. Maekisara, O. Simula, and J. Kangas). North Holland, 397-402.

Martinetz, T. and Schulten, K. (1994). Topology representing networks, *Neural Networks* **7**, 507-522.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set, *Psychometrica* **50**, 159-179.

Mitchell, A. (1983). *The nine American lifestyles: Who we are and where we're going.* Macmillan.

Mowen, J. C. and Minor, M. (1998). *Consumer Behavior* (5th ed.). Prentice Hall.

Papastefanou, G., Schmidt, P., Börsch-Supan, A., Lüdtke, H., and Oltersdorf, U. (1999). Social and economic research with consumer panel data. *Proceedings of the First ZUMA Symposium on Consumer Panel Data*, GESIS.

Questier, F., Guo, Q., Walczak, B., Massart, D. L., Boucon, C., and de Jong, S. (2002). The neural-gas network for classifying analytical data. *Chemometrics and Intelligent Laboratory Systems* **61**, 105-121.

Simula, O., Vasara, P., Vesanto, J., and Helminen, R. R. (1999). The self-organizing map in industry analysis. In *Industrial Applications of Neural Networks* (Edited by L. C. Jain and V. R. Vemuri). CRC Press.

Vellido, A., Lisboa, P. J. G., and Meehan, K. (1999). Segmentation of the on-line shopping market using neural networks. *Expert Systems with Applications* **17**, 303-314.

Villmann, T. and Bauer, H.-U. (1998). Applications of the growing self-organizing map. *Neurocomputing* **21**, 91-100.

Wagner, R. (2004). Mining promising qualification patterns. In *Innovations in Classification, Data Science, and Information Systems* (Edited by D. Baier and K.-D. Wernecke). Springer, 249-256.

Wedel, M. and Kamakura, W. A. (2000). *Market Segmentation: Conceptional and Methodological Foundations* (2nd ed.). Kluwer Academic Publishers.

Wells, W. D. and Tigert, D. J. (1971). Activities, interests and opinions. *Journal of Advertising Research* **11**, 27-35.

Yerramalla, S., Cukic, B., and Fuller, E. (2003). Lyapunov stability analysis of the quantization error for DCS neural networks. To appear in: *Proceedings of the 2003 International Joint Conference on Neural Networks*, IEEE Press.

Reinhold Decker

Department of Business Administration and Economics

Bielefeld University, P.O. Box 100131

D-33501 Bielefeld, Germany

rdecker@wiwi.uni-bielefeld.de