Market basket analysis with neural gas networks and self-organising maps

Received (in revised form): 3rd March, 2003

Reinhold Decker

is Professor and Head of Marketing at the University of Bielefeld, Germany and a visiting professor in marketing at universities in Austria, Russia and Bulgaria. He is author and co-author of numerous publications in journals, conference proceedings and compilations focussing on marketing research and data analysis as well as the author and co-editor of academic books. He is a member of several academic societies and serves as a referee for different journals.

Katharina Monien

received her diploma in mathematics in 2001 and is a lecturer in marketing at the Department of Economics and Business Administration, University of Bielefeld. Her research interest is the application of neural networks and machine learning in marketing.

Abstract Market basket analysis has been an elementary part of quantitative decision support in retail marketing for many years and it is regularly cited as a prime application area of data mining. In this paper two competitive neural network approaches are presented and discussed with respect to their suitability for purchase interdependence analysis on the product category level. Particular attention is paid to the user-oriented representation or visualisation of cross-category dependences. Both approaches are applied to point of sales scanner data provided by a German retail chain to check how far they are able to uncover presumed purchase interdependences.

INTRODUCTION

Market basket analysis has not only been an important topic of traditional retail marketing for more than 25 years¹⁻³ but has also gained an increasing relevance for electronic retailing.⁴ According to Russell and Petersen⁵ market basket analysis focuses on the decision process in which a consumer selects items from a given set of product categories on the same shopping trip. Correspondingly, the term 'purchase interdependence' is used in the following pages to refer to interrelations between different elements of a retail assortment resulting from purchases that have already been carried out.

The analysis of market basket data has experienced a renaissance as a result of various publications on data mining and knowledge discovery in databases since 1993.⁶⁻⁹ In these papers market basket analysis is partly regarded as a typical field of application for data mining and the main emphasis is put on association rule-based approaches. In Decker and Schimmelpfennig¹⁰ the traditional association coefficient-based approach (measuring purchase interdependence by means of cluster analysis and multidimensional scaling) is compared with a rule-based approach reflecting on electronic retailing. In the present paper both a self-organising map and a neural

Katharina Monien
Department of Economics
and Business
Administration, University of
Bielefeld, PO Box 10 01 31,
33501 Bielefeld, Germany.
Tel: +49 (0)521 106 4844;
Fax: +49 (0)521 106 6456;
e-mail: kmonien@
wiwi.uni-bielefeld.de

gas network approach are investigated with respect to their suitability for purchase interdependence analysis on the product category level.

The practical benefit of such approaches strongly depends on their ability to uncover 'relevant' interrelations in the assortment to be analysed and the extent to which one succeeds in adequately documenting or visualising the same for decision support in retail marketing. It may be that existing asymmetric interdependences between individual product categories and the inevitable occurrence of 'random take away effects'¹¹ are special challenges in this context. At the same time purchase interdependence analysis is confronted with a continuously growing database resulting from point of sales (POS) scanning. All these aspects have to be kept in mind when a new approach is discussed.

The rest of the paper is organised as follows. In the following section a description of the self-organising map as well as the less well-known neural gas network approach are presented as tools for purchase interdependence analysis. Then easy-to-interpret representations of purchase interdependences are proposed in the next section. Finally, a reality-based application of both approaches is presented. The paper closes with some conclusions and a short outlook.

PURCHASE INTERDEPENDENCE ANALYSIS BY MEANS OF SELF-ORGANISING MAPS AND NEURAL GAS NETWORKS

Both approaches introduced for purchase interdependence analysis start with two basic assumptions. First, the measurement of purchase interdependence is meaningful only on the multicategory level. This means that all product categories of interest have to be

considered simultaneously. Secondly, interrelations of the interesting kind usually lead to similar 'market basket patterns'. Two market baskets have a similar pattern if they are characterised by a more or less identical combination of product categories.

To identify those interrelations the authors adapted the neural gas network (NGN) approach introduced by Martinetz and Schulten¹² and the well-known self-organising map (SOM) approach introduced by Kohonen¹³ to the problem on hand. In both cases the relevant patterns are represented by 'neurons' (units). The formal counterpart of such a unit is a weight vector. When the training of the neural network is finished, that is to say after the net weights have been fitted to the given data, the resulting final weight vectors are called prototypes.

Before starting the methodological considerations some symbols have to be introduced. Let n be the number of interesting product categories and m the unrestricted number of individual market baskets to be analysed. Then an individual market basket can be defined by binary transaction vector $\mathbf{t}_j = (t_{j1}, t_{j2}, \dots, t_{jn})'$ with $j \in \{1, \dots, m\}$, where $t_{jk} = 1$ if market basket j contains at least one item of product category $k \in \{1, \dots, n\}$ and $t_{jk} = 0$ otherwise.

Essentials of the SOM approach

In most applications the units of an SOM are organised on a two dimensional grid where the individual positions reflect the interrelations between the respective units.¹⁴ In the following, each unit u_h ($h \in \{1, \dots, p\}$) is represented by a weight vector $\boldsymbol{\eta}_{ab} = (\eta_{ab1}, \dots, \eta_{abk}, \dots, \eta_{abn})'$ where a and b refer to the position of the unit within a rectangular grid and $0 \leq \eta_{abk} \leq 1$ holds true.

Applying the SOM approach to POS scanner data in order to identify existing purchase interdependences means carrying out two different tasks simultaneously: finding an optimal set of prototypes representing similar market baskets and ensuring the optimal topological arrangement of these prototypes. In the resulting map similar weight vectors (representing similar market basket patterns) are located close together. How to interpret such a map in detail with respect to purchase interdependence will be shown in the empirical part of the paper.

The net training process looks like this: at the beginning, all weights η_{abk} have to be initialised in a suitable way. In the following iterations $l \in \{1, \dots, l_{\max}\}$ the distances between each weight vector η_{ab} and a randomly chosen input vector (market basket) $\mathbf{t}_j$ are computed according to

$$\text{dist}(\mathbf{t}_j, \eta_{ab}) = \|\mathbf{t}_j - \eta_{ab}\|^2 = \sum_{k=1}^n (t_{jk} - \eta_{abk})^2 \quad (1)$$

to determine the winning unit $u_{c_d c_b}$ with minimum distance. Then each weight vector has to be updated as follows:

$$\eta_{ab}(l+1) = \eta_{ab}(l) + \alpha(l) \cdot nhf_{c_d c_b}(a, b) \cdot (\mathbf{t}_j - \eta_{ab})$$

where learning rate $\alpha(l) = \alpha(0)(1 - l/l_{\max})$ is a decreasing function of iteration l . The extent of this adaptation can be controlled via neighbourhood function

$$nhf_{c_d c_b}(a, b) = \exp\left(-\frac{(a - c_a)^2 + (b - c_b)^2}{2\sigma(l)^2}\right)$$

where $\sigma(l) = \sigma(0)(1 - l/l_{\max})$ determines the scope of the neighbourhood kernel. The whole procedure is repeated until the maximum number of iterations $l_{\max}$ or any other *a priori* defined stopping criterion is reached. The minimisation of

distances finally leads to the optimal prototype system $\{\eta_{c_d c_b}\}$.

To optimise the topological structure of the whole map the accumulated neighbourhood distance

$$\begin{aligned} \sum_{(c_a, c_b)} \text{dist}(\eta_{c_a c_b}, \eta_{\tilde{c}_a \tilde{c}_b}) &= \sum_{(c_a, c_b)} \sum_{(\tilde{c}_a, \tilde{c}_b)} \|\eta_{c_a c_b} - \eta_{\tilde{c}_a \tilde{c}_b}\|^2 \\ &= \sum_{(c_a, c_b)} \sum_{(\tilde{c}_a, \tilde{c}_b)} \sum_{k=1}^n (\eta_{c_a c_b k} - \eta_{\tilde{c}_a \tilde{c}_b k})^2 \end{aligned} \quad (2)$$

with $u_{\tilde{c}_a \tilde{c}_b} \in N_{c_a c_b}$ has to be minimised as well. Neighbourhood set $N_{c_a c_b}$ contains all units $u_{\tilde{c}_a \tilde{c}_b}$ that are topological neighbours of the winning unit $u_{c_a c_b}$ within the rectangular grid.

In the basic SOM approach the number of units p has to be fixed in advance which can be restrictive in some cases. To avoid this disadvantage Alahakoon *et al.*¹⁵ have proposed a so-called growing self-organising map (GSOM) approach where both the size and the shape of the network are determined dynamically during the training process. Because of the fundamental nature of this investigation this aspect is not elaborated on at this point.

Essentials of the NGN approach

Again, each unit u_h ($h \in \{1, \dots, s\}$) is represented by a weight vector

$\eta_h = (\eta_{h1}, \dots, \eta_{hk}, \dots, \eta_{hn})' \in [0; 1]^n$. The dimensionality of these vectors is equal to the number of interesting product categories n . In contrast to the SOM approach, there is no *a priori* defined grid to be fitted.¹⁶

The similarity of both approaches also becomes apparent in the training process. After initialisation the distance between each weight vector η_h and a randomly chosen binary transaction vector $\mathbf{t}_j$ is computed in accordance to equation 1. In respect of these distances all units u_h are arranged in such a way that the

winning unit with weight vector $\boldsymbol{\eta}_{win}$ takes rank $r(\mathbf{t}_j, \boldsymbol{\eta}_{win}) = 0$, the co-winning unit takes rank 1 and so on. This ranking explicitly determines the strength of the adaptation of the individual weight vectors in each iteration. It is

$$\Delta \boldsymbol{\eta}_h(l) = \varepsilon(l) \cdot nhf_{\lambda(l)}(r(\mathbf{t}_j, \boldsymbol{\eta}_h)) \cdot (\mathbf{t}_j - \boldsymbol{\eta}_h)$$

with neighbourhood function

$$nhf_{\lambda(l)}(r(\mathbf{t}_j, \boldsymbol{\eta}_h)) = \exp\left(-\frac{r(\mathbf{t}_j, \boldsymbol{\eta}_h)}{\lambda(l)}\right)$$

and learning rates

$$\varepsilon(l) = \varepsilon_{init} \cdot \left(\frac{\varepsilon_{End}}{\varepsilon_{init}}\right)^{\frac{l}{l_{max}}}$$

and

$$\lambda(l) = \lambda_{init} \cdot \left(\frac{\lambda_{End}}{\lambda_{init}}\right)^{\frac{l}{l_{max}}} \text{ respectively.}$$

Obviously, the numerical value of neighbourhood function $nhf_{\lambda(l)}(r(\mathbf{t}_j, \boldsymbol{\eta}_h))$ decreases, other things being equal, with an increasing rank of the respective unit. Therefore, only those weight vectors that are close to the input signal $\mathbf{t}_j$ are changed significantly according to:

$$\boldsymbol{\eta}_h(l+1) = \boldsymbol{\eta}_h(l) + \Delta \boldsymbol{\eta}_h(l).$$

The basic NGN approach can be extended in different ways. For example, following Martinetz and Schulten¹⁷ a data-driven topological structure can be added to show the interrelations between units. For this purpose a connectivity matrix $\mathbf{C} = (c_{h_1 h_2})_{h_1, h_2 = 1, \dots, s}$ has to be introduced with

$$c_{h_1 h_2} = \begin{cases} 1, & \text{if unit } u_{h_1} \text{ and } u_{h_2} \text{ are connected} \\ 0, & \text{otherwise.} \end{cases}$$

Each time unit u_h becomes the winning unit indicator variable $c_{h\bar{h}}$ is set to 1 for the co-winning unit $u_{\bar{h}}$. To take into account the strength of the connection

between two units u_h and $u_{\bar{h}}$ an additional controlling variable $age_{h\bar{h}}$ is introduced. This variable is set to 0 if $c_{h\bar{h}}$ is set to 1. The ‘ages’ of all the other connections to winning unit u_h are raised by 1. If the ‘age’ of a connection exceeds the dynamically computed maximum

$$age_{max}(l) = age_{init} \cdot \left(\frac{age_{End}}{age_{init}}\right)^{\frac{l}{l_{max}}}$$

this connection is removed from the network. Threshold $age_{max}(l)$ depends on pre-defined initial and final values age_{init} and age_{End} as well as on iteration l .

Current extensions of the NGN approach focus on the speeding up of the training process or on the dynamic determination of the number of units to be included in the network. Atukorale and Suganthan,¹⁸ for example, have published a so-called implicit ranking scheme. The basic idea of this proposal is to redefine the neighbourhood function using a kind of normalised distance

$$q(\mathbf{t}_j, \boldsymbol{\eta}_h) = \frac{dist(\mathbf{t}_j, \boldsymbol{\eta}_h) - dist_{min}}{dist_{max} - dist_{min}}$$

instead of the original rank order to reduce training time. The maximum and minimum distances $dist_{max}$ and $dist_{min}$ between the current input signal $\mathbf{t}_j$ and each weight vector have to be defined adequately. For the winning unit $q(\mathbf{t}_j, \boldsymbol{\eta}_{win})$ is equal to 0. Another modification suggested by the same authors is the so-called truncated update rule, where only those weight vectors are updated with normalised distances smaller than a dynamically computed threshold.

In an own sensitivity study it became apparent that there are only negligible differences in the running times of these modifications, if $s \leq 20$. Therefore, a detailed description was dispensed with and a decision made in favour of the original NGN approach for the investigations.

Table 1: Elements of the NGN output

Unit	Frequency	Ranking	Weights
u_1	$P(u_1)$	$PC_{k_1^1} PC_{k_2^1} \dots PC_{k_n^1}$	$\eta_{1k_1^1} \eta_{1k_2^1} \dots \eta_{1k_n^1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_s	$P(u_s)$	$PC_{k_1^s} PC_{k_2^s} \dots PC_{k_n^s}$	$\eta_{sk_1^s} \eta_{sk_2^s} \dots \eta_{sk_n^s}$

Similarly to the GSOM approach, there is also a dynamic extension of the NGN approach called growing neural gas network. For a detailed description of this see Fritzke.¹⁹

REPRESENTATION OF INTERRELATIONS

As already emphasised at the beginning of the paper, an essential aspect of purchase interdependence analysis is the adequate documentation or visualisation of the uncovered interrelations. The availability of comprehensible representations (eg as an essential part of regular sales reporting) makes it easier to apply this information to marketing decisions, for instance, with respect to product placement and promotion pricing. In this section some procedures are proposed that can be applied to both the SOM and the NGN approach. For demonstration purposes the focus is on the latter.

Table 1 gives a general impression of the NGN output underlying the following considerations. After having finished the training process the weights of each unit u_h can be arranged in descending order: $1 \geq \eta_{hk_1^h} \geq \dots \geq \eta_{hk_n^h} \geq \dots \geq \eta_{hk_n^h} \geq 0$. The corresponding 'ranking' of the individual product categories (abbreviated by PC) is displayed in the centre of the table whereas the share of market baskets $P(u_h)$ (with $\sum_{h=1}^s P(u_h) = 1$) which fit prototype h is given in column 'Frequency'.

Starting from this it is possible to visualise interesting interrelations both on the product category and the market basket level by means of a graph.

Visualising purchase interdependences

To visualise possible interdependences on the product category level

$P(PC_{k_i^h} | u_h) = \eta_{hk_i^h} \forall h, k_i^h$ (with $k_i^h \in \{1, \dots, n\}$) is interpreted as the probability of observing product category $PC_{k_i^h}$ in a market basket that corresponds to prototype h . Then this probability is combined with the observed frequency of each prototype to get the compound probability $P(PC_{k_i^h} \cap u_h) = P(u_h) \cdot P(PC_{k_i^h} | u_h) \forall h, k_i^h$. To decide whether two product categories have to be connected in a relating graph because of their interdependence a user-defined threshold d can be introduced. Product categories k_1 and k_2 are connected if both probabilities $P(PC_{k_1})$ and $P(PC_{k_2})$ are greater than d for at least one h . Alternatively, a suitable elbow criterion can be applied. Figure 1 shows what such a graph might look like.

Product categories 6 and 7 are interdependent in the given sense but without having any other meaningful interrelation, whereas product category 2, for example, is part of a much more complex net. In the empirical application section such a graph will be created from real POS scanner data with presumed interrelations between individual product categories.

Figure 1 Possible structure of a simple purchase interdependence graph

Heterogeneity and asymmetry in market basket patterns

Using connectivity matrix $\mathbf{C}$ a graph that represents existing similarities between individual prototypes is created. Two units are ‘neighbouring’ if they are represented by similar weight vectors. Each edge of the graph is weighted by the reciprocal age $1/age_{h_1h_2}$. The longer a connection has not been confirmed during network training the lower the weight of this edge. If a continuously growing number of market baskets have to be represented (eg as a result of daily POS scanning in retail stores) the current relevance of a connection can be expressed dynamically in this way. A high number of edges with comparatively low weights point to distinctive differences (heterogeneity) with respect to the underlying buying patterns.

Furthermore, using the formula of Bayes, asymmetries on the market basket level can be described. Let

$$\begin{aligned} P(u_h | PC_k) &= \frac{P(PC_k \cap u_h)}{P(PC_k)} \\ &= \frac{P(u_h) \cdot P(PC_k | u_h)}{\sum_{h=1}^s P(u_{h_i}) \cdot P(PC_k | u_{h_i})} \quad (3) \end{aligned}$$

be the probability of observing pattern

(prototype) h when product category PC_k is already an element of an arising market basket. The computation of these probabilities can provide valuable hints about those product categories which induce the purchase of items of the remaining categories. From a sales-oriented point of view those product categories are of particular interest to the retail management which induce market baskets containing ‘profitable’ product categories with high probability. Information of this kind is useful primarily for periodical promotion planning. Figure 2 illustrates the general structure of an individual vertex of the graph. The given probabilities result from equation 3.

EMPIRICAL APPLICATION TO POS SCANNER DATA

Data description

To demonstrate their general suitability both neural network approaches were applied to POS scanner data collected by a German retail chain in the mid-1990s. For illustration purposes 25 product categories from the chemist’s assortment were selected. Accompanying investigations of the available data have

$PC_{k_1^h}$	$PC_{k_2^h}$	$\dots$	$PC_{k_n^h}$	u_h
$P(u_h PC_{k_1^h})$	$P(u_h PC_{k_2^h})$	$\dots$	$P(u_h PC_{k_n^h})$	
$P(PC_{k_1^h})$	$P(PC_{k_2^h})$	$\dots$	$P(PC_{k_n^h})$	

Figure 2 General structure of vertices

shown that the following considerations can also be transferred without restrictions to other categories of products in everyday use. In this context, it should be mentioned that in this implementation, using SAS Release 8.02, neither the SOM nor the NGN approach is restricted with respect to the maximum number of market baskets to be analysed. But, of course, an upper bound may result from the storage capacity of the employed hardware.

The data referred to in Table 2 result

from 2,079 market baskets. Product categories with possible purchase interdependences have been put together in separate fields. Each market basket is coded with a binary vector where 1 indicates the occurrence and 0 the non-occurrence of the respective product category. The total number of items of a product category in a basket is not considered. In doing this it is possible to abstract from biasing stock-keeping effects. The transformation of the original data (containing, for example, information on price, time and date of purchase) into binary vectors was realised with standard data management facilities of SAS.

The product categories 8 to 12, for example, can be assumed to be interrelated in the relevant sense because they are at least partially complementary. The same, but in a somewhat more pronounced way, seems to be valid for product category 17 and 18. Items of these two groups can only be used jointly. Finally, product categories 21 and 22 represent so-called random take away products, that are often placed spontaneously in the basket at the checkout. A distinct and causally motivated relation to any other product category listed in the table is not discernible.

Results received from SOM

Checking several possible alternatives an 8×8 SOM layer was found to produce

Table 2: Profile of the POS scanner data

No.	Product category	Occurrence
1	Shampoos	732
2	Hair conditioners	263
3	Hair lotions	206
4	Tampons	221
5	Sanitary napkins	248
6	Cat food	174
7	Rewards for cats/dogs	127
8	Juices for babies	147
9	Desserts for babies	260
10	Vegetables for babies	120
11	Childrens food	98
12	Childrens menus	93
13	Denture cleansing agents	76
14	Denture fixer	50
15	Sun protectors/blockers	21
16	After sun lotions	10
17	Shaving soaps/creams	181
18	Razor blades	233
19	Slim/diet food	56
20	Functional/health food	132
21	Cough drops	473
22	Chewing gum	362
23	Heart and nerve tonics	56
24	Eyeshadow	234
25	Lipsticks	174

<i>a/b</i>	1	2	3	4	5	6	7	8
1	1 2	2 21	2	17	17 18	18	8 10	9 10
2	1 2 3	3	2 4	4 17	5 18	8	8 9	10 11
3	1 3	1 3	1 4	1 4 5	4 5	4	9	9 11
4	1	1 18	1 17	1 5	5 21		9 12	11 12
5	1 21	1	1	2 5	5	6	13 14	20
6	21 24	1 24	1 22	2 22	6 22	6 7	7	20 21
7	24	22 24	22	22	21 22	6 21	4 21	21
8	24 25	22 25	5 22	20 22	21 22	21 22	23	21

Figure 3 SOM output after $I_{\max} = 100,000$ iterations

the best results with respect to heterogeneity (cf. for this equation 1)

$$het\{\boldsymbol{\eta}_{c_a c_b}\} = \frac{1}{m} \sum_{j=1}^m dist(\mathbf{t}_j, \boldsymbol{\eta}_{c_a c_b}) = 1.183,$$

and simplicity (cf. equation 2)

$$simpl\{\boldsymbol{\eta}_{c_a c_b}\} = \sum_{(c_a, c_b)} dist(\boldsymbol{\eta}_{c_a c_b}, \boldsymbol{\eta}_{c_a \hat{c}_b}) = 2.390$$

of the prototype system as well as the interpretability in content. The solution accompanied by the initial learning rate $\alpha(0) = 0.7$ and $\sigma(0) = 3$ for the neighbourhood kernel is shown in Figure 3.

Each of the 64 fields of the map represents one unit or prototype. To make interpretations easier, however, only those product categories with weights greater than 0.8 have been displayed. That is why three product categories, 15, 16 and 19, do not appear in the map. Regarding the first two product categories (none of them have weights greater than 0.25), this is not too surprising because of the comparatively low frequency of occurrence (cf. Table 2). Product category 19, with maximum weight 0.73, however, only narrowly misses its inclusion in the map at the 'expected' place (row 5, column 8).

All in all, the different 'clouds' in Figure 3 conform to a great extent to the presumed interrelations. The hair care products of categories 1, 2 and 3, for instance, define a cloud where product category 1 plays a seemingly special role. Taking into account the basic function of the items of this category (shampoos) within the entire hair care process this seems to be quite plausible. In the same way the product categories 8 to 12 define a fairly compact cloud in the upper right-hand corner of the map. This could be rated as a hint at some stronger relations between these product categories. The interrelation of product categories 24 and 25, to mention another nice example, is also undoubtedly understandable.

Furthermore, the suggested type of representation is useful with respect to the detection of so-called random take away products. A simple and plausible indicator for this phenomenon is the extent to which a product category 'scatters' across the map. In the present case this seems to be valid for product category 21 (cough drops) which appears in several, but not necessarily neighbouring, units and which is displayed together with different product categories within one unit. For the other

Table 3: NGN output after $l_{\max} = 20,000$ iterations

h	$P(u_h)$	$PC_{K_1^h}$	$PC_{K_2^h}$	$PC_{K_3^h}$	$PC_{K_4^h}$	$PC_{K_5^h}$	$\eta_{hk_1^h}$	$\eta_{hk_2^h}$	$\eta_{hk_3^h}$	$\eta_{hk_4^h}$	$\eta_{hk_5^h}$
1	0.071	24	25	1	3	18	0.925	0.497	0.279	0.143	0.095
2	0.131	22	21	1	24	3	1.000	0.341	0.275	0.106	0.066
3	0.091	5	1	4	22	6	1.000	0.363	0.226	0.121	0.074
4	0.125	9	10	8	11	12	0.858	0.373	0.362	0.331	0.242
5	0.058	20	19	21	22	1	0.983	0.300	0.275	0.167	0.133
6	0.055	6	7	13	14	3	0.772	0.605	0.158	0.158	0.079
7	0.143	21	1	5	24	4	1.000	0.232	0.117	0.084	0.077
8	0.089	18	17	1	25	13	0.773	0.681	0.314	0.065	0.059
9	0.170	1	2	3	25	6	0.926	0.554	0.304	0.088	0.051
10	0.067	4	1	22	2	24	1.000	0.371	0.171	0.100	0.100

presumed random take away product category 22 (chewing gum) unfortunately no such clear picture is obtained. In a countermove this seems to apply for product category 1.

Results received from NGN

Table 3 contains the output of the original NGN approach using 10 units. Even this parsimonious specification turned out to produce acceptable results with respect to heterogeneity

$$het\{\boldsymbol{\eta}_h\} = \frac{1}{m} \sum_{j=1}^m dist(\mathbf{t}_j, \boldsymbol{\eta}_h) = 1.090$$

and interpretability. The initial and final values of the learning rates ($\varepsilon_{Init} = 0.7$, $\varepsilon_{End} = 0.005$, and $\lambda_{Init} = 1.0$, $\lambda_{End} = 0.01$) as well as those of the 'age' ($age_{Init} = 10$, $age_{End} = 100$) are similar to proposals made in Martinetz and Schulten,²⁰ Martinetz *et al.*²¹ and Fritzke.²²

Because of space restrictions only five product categories at a time (starting with the highest weight) have been displayed. Obviously, the product categories for children are clearly dominating prototype $h = 4$ and are strongly interdependent. Even product category 12 at rank 5 has a weight greater than 0.2. But there are also some other very plausible purchase interdependences indicated by comparatively high weights within one

prototype. This applies for example to the hair care product categories 1, 2 and 3 that are dominating prototype $h = 9$. At the same time it emerges that product category 1 (shampoos) seemingly contains random take away products because of its co-occurrence in several other market basket prototypes. Another very nice interrelation is uncovered by prototype $h = 6$ where the dental care products appear together with cat food and rewards for cats and dogs. Obviously, these products for small pets are predominantly bought by an older clientele. Insights of this kind can be used excellently for customer oriented sales promotions. Finally, prototype $h = 1$ contains face and hair care products typically bought by female consumers.

All in all the NGN results are very similar to those produced by the SOM. The NGN approach proves, however, to be more parsimonious both regarding training time and the required number of units. In this respect, at least in this investigation, the NGN approach slightly dominates the SOM approach.

Nevertheless, a final assessment needs more comparisons on different data sets.

Taking into account that it makes a great difference whether a prototype represents a frequent or a scarce market basket pattern the compound probabilities $P(PC_{k_i^h} \cap u_h)$ can be computed using the available weights

Figure 4 Interdependences on the product category level

and frequencies. Additionally, defining a threshold $d = 0.025$, for example, finally leads to the graphical representation of purchase interdependences on the product category level depicted in Figure 4.

The present visualisation of purchase interdependences is remarkable in two respects. First, the individual subgraphs reflect, as expected, the most important interrelations from a data analytical point of view. The starlike subgraph on the left-hand side, for instance, confirms once again the special role of product category 1 which has already been mentioned and which results from the at least partly existing random take away effects. Secondly, graphical representations like this are an elegant way of enabling visualisation of both direct and indirect interdependences. Product categories 2 and 3, for example, are directly interdependent. But there is also an indirect relation to product category 4 and 5 via product category 1. In contrast to this the apparently strongly interdependent product categories 6 and 7 seem to occupy a solitary position. This point will be considered later.

Figure 5 shows the result of an application of the Bayes approach (cf. equation 3) to the given data. To simplify the graph only those connections (and, as a result, prototypes) have been depicted that exceed an *a priori* defined probability to appear. The probability of an edge connecting unit u_{h_1} and unit u_{h_2} is equal to the number of iterations where these units were the winning and the co-winning unit divided by the total numbers of iterations. Fixing the threshold for this probability to $1/25$, for instance, results in the graph on hand. Unit u_3 and u_5 are missing because of their 'weak' connections to the other units in the present sense. Each edge of the graph is additionally weighted by its reciprocal age.

The small weight of the edge connecting unit u_2 and u_7 ($1/18$), for example, indicates that the seeming similarity of both prototypes could not be confirmed for a longer time. The opposite holds for the edge connecting unit u_9 and u_{10} . Obviously, the latest input signal (transaction vector) has confirmed the common ground of both prototypes. The reader might be

Figure 5 Interdependences on the market basket level

astonished about the connection of unit u_2 and u_4 although both seem to be represented by different prototypes. In fact, this edge is primarily determined by product categories ranked 6 to 25, which are not displayed. Therefore, if the NGN is continuously (eg daily) adapted to new POS scanner data the changes of weights in the course of time provide valuable hints at an emerging alignment or differentiation of buying patterns. A corresponding challenge to future research might be the development of a measurement framework to monitor dynamically movements in the observed purchase interdependences. Additionally, it would be possible to focus on the respective influence of changing sales promotion activities.

According to Figure 2 each vertex of the graph in Figure 5 contains the rank

order of product categories (first row), the Bayesian probabilities $P(u_i | PC_k^h)$ (second row), and the product purchase probabilities $P(PC_k^h)$ (third row). Looking at the last two rows some interesting asymmetries are detectable. For instance, in unit u_9 , the probability of observing the respective market basket pattern is greater with product category 2 on hand instead of product category 1 (0.741 versus 0.445). The contrary is valid for the probabilities of buying items of these two product categories (0.127 versus 0.352). An example of a more or less symmetric relation is given by product categories 9 and 11 (0.858 versus 0.878) in unit u_4 . But the fact that items of product category 9 appear in a market basket with a higher probability (0.125 versus 0.047) makes this one more interesting for promotional activities.

Figure 6 MDS representation of the NGN results

It is necessary to point out that asymmetries of the present kind are only valid for individual prototypes. Product categories 24 and 1, for example, are characterised by an extremely asymmetric relation with respect to prototype $h = 1$, whereas the same relation looks nearly symmetric for prototype $h = 10$. This is caused by the fact that product category 24 (together with 25) dominates the profile of prototype $h = 1$. Information about those ‘dominations’ can be used to force cross-sellings within the assortment under consideration.

Last but not least, the Bayesian probabilities can be used to generate rules like this: ‘Product category k determines the occurrence of market basket type h ’. Transforming the conditional probabilities into verbal rules eases the communication between the analyst and the decision maker.²³ The interesting nature of such a rule can be determined using the lift, a measure that is well-known from data mining.²⁴

$$\begin{aligned} \text{lift}(PC_k \Rightarrow u_h) &= \frac{\text{conf}(PC_k \Rightarrow u_h)}{\sup(u_h)} \\ &= \frac{P(u_h | PC_k)}{P(u_h)} \end{aligned}$$

If only product categories with a lift greater than 1.0 are considered another graph will result which is very similar to that depicted in Figure 4. The corresponding probabilities are in bold face in Figure 5. This time product categories 6 and 7 as well as 13 and 14 would be connected within the concerning subgraph. In this way the abovementioned interesting nature of this interrelation is confirmed from a methodological point of view as well.

Visualising purchase interdependences by means of NGN-based multidimensional scaling

To be able to compare the results of the previous subsection to traditional approaches of purchase interdependence analysis, for instance to those starting from association coefficients like Tanimoto,²⁵ it seems to be helpful to carry out a simple transformation of the NGN output. In the present case product categories k_1 and k_2 are assumed to be interrelated if they have similar probabilities $P(PC_{k_1} \cap u_h)$ and $P(PC_{k_2} \cap u_h)$ with respect to all prototypes h . Analysing the corresponding similarities

by means of multidimensional scaling (MDS) finally leads to Figure 6. Because of the high conformity of this representation (produced with SAS PROC MDS) with the assumptions above a more intensive interpretation is not necessary.

CONCLUSIONS AND OUTLOOK

This paper is concerned with the presentation and discussion of two alternative neural network approaches for purchase interdependence analysis. With the empirical investigation it could be shown that both approaches are powerful tools providing outputs that can be processed in different ways to extract information. Both can isolate random take away effects to a certain degree and can be extended regarding the detection of asymmetries at the market basket level. An important advantage of both approaches is the absence of a methodologically motivated restriction of the maximum number of market baskets to be analysed. The adaptability of NGN makes this approach a useful instrument for dynamic POS scanner data analysis. Considering the fact that, at least for the data here — the NGN approach is superior to the SOM approach with respect to both the training time and the required number of weights the former is worth a more thorough investigation in the present context. On the other hand, in contrast to NGN, implementations of the basic SOM methodology are available in several commercial or academic tools for data analysis which makes its application significantly easier.

Future research should concentrate on the development of meaningful quality measures for application-oriented market basket analysis and the identification of possible differences between weekly shopping baskets and those of 'top-up' shopping trips. Beyond this the reliable

and comprehensive isolation of random take away effects still requires considerably greater effort.

The authors are preparing a further application of both approaches to a large data set provided by another retail chain — once again concerning everyday products, but different from the chemist's assortment. Those who are interested in the results (which are scheduled to be available in summer 2003) are invited to write to the authors.

Acknowledgment

The authors would like to thank two anonymous reviewers for their helpful hints concerning an earlier draft of the paper.

References

- 1 Böcker, F. (1978) 'Die Bestimmung der Kaufverbundenheit von Produkten', Duncker & Humblot, Berlin.
- 2 Hruschka, H. (1991) 'Bestimmung der Kaufverbundenheit mit Hilfe eines probabilistischen Meßmodells', *Zeitschrift für betriebswirtschaftliche Forschung*, Vol. 43, No. 5, pp. 418–434.
- 3 Merkle, E. (1981) 'Die Erfassung und Nutzung von Informationen über den Sortimentsverbund in Handelsbetrieben', Duncker & Humblot, Berlin.
- 4 Hao, M. C., Dayal, U., Hsu, M., Sprenger, T. and Gross, M. H. (2001) 'Visualization of directed associations in e-commerce transaction data', Hewlett Packard Research Laboratories, Palo Alto.
- 5 Russell, G. J. and Petersen, A. (2000) 'Analysis of cross category dependence in market basket selection', *Journal of Retailing*, Vol. 76, No. 3, pp. 367–392.
- 6 Agrawal, R., Imielinski, T. and Swami, A. (1993) 'Mining association rules between sets of items in large databases', in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington, pp. 207–216.
- 7 Brin, S., Motwani, R., Ullman, J. D. and Tsur, S. (1997) 'Dynamic itemset counting and implication rules for market basket data', in *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, Tucson, pp. 255–264.
- 8 Hu, Z., Chin, W.-N. and Takeichi, M. (2000) 'Calculating a new data mining algorithm for market basket analysis', in Pontelli, E. and Santos Costa, V. (eds) 'Practical aspects of declarative languages', *Lecture Notes in Computer Science*, No. 1753, Springer, Berlin, pp. 169–184.
- 9 Hao *et al.* (2001) *op cit.*
- 10 Decker, R. and Schimmelpfennig, H. (2002)

- 'Alternative Ansätze zur datengestützten Verbundmessung im Electronic Retailing', in Ahlert, D., Olbrich, R. and Schröder, H. (eds) 'Jahrbuch Handelsmanagement 2002 — Electronic Retailing', Deutscher Fachverlag, Frankfurt, pp. 193–212.
- 11 Schmalen, H., Pechtl, H. and Schweitzer, W. (1996) 'Sonderangebotspolitik im Lebensmittel-Einzelhandel', Schäffer-Poeschel, Stuttgart.
- 12 Martinetz, T. and Schulten, K. (1991) 'A "neural gas" network learns topologies', in Kohonen, T., Mäkisara, K., Simula, O. and Kangas, J. (eds) 'Artificial neural networks', North Holland, Amsterdam, pp. 397–402.
- 13 Kohonen, T. (1982) 'Self-organized formation of topologically correct feature maps', *Biological Cybernetics*, Vol. 43, pp. 59–69.
- 14 Kohonen, T. (2001) 'Self-organizing maps', 3rd edn, Springer, Berlin.
- 15 Alahakoon, D., Halgamuge, S. K. and Srinivasan, B. (2000) 'Dynamic self-organizing maps with controlled growth for knowledge discovery', *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, pp. 601–614.
- 16 Martinetz and Schulten (1991) *op cit.*
- 17 *Ibid.*
- 18 Atukorale, A. and Suganthan, N. (2000) 'Hierarchical overlapped neural-gas network with application to pattern classification', *Neurocomputing*, Vol. 35, No. 1–4, pp. 165–176.
- 19 Fritzke, B. (1995) 'A growing neural gas network learns topologies', in Tesauro, G., Touretzky, D. S. and Leen, T. K. (eds) 'Advances in neural information processing systems 7', MIT Press, Cambridge, pp. 625–632.
- 20 Martinetz and Schulten (1991) *op cit.*
- 21 Martinetz, T., Berkovich, S. G. and Schulten, K. (1993) '"Neural-gas" network for vector quantization and its application to time-series prediction', *IEEE Transactions on Neural Networks*, Vol. 4, No. 4, pp. 558–569.
- 22 Fritzke (1995) *op cit.*
- 23 Pedrycz, W. (2001) 'Granular computing in data mining', in Kandel, A., Last, M. and Bunke, H. (eds) 'Data mining and computational intelligence', Physica, Heidelberg, pp. 37–61.
- 24 Brin *et al.* (1997) *op cit.*
- 25 Merkle (1981) *op cit.*