UNIVERSITÄT LEIPZIG

Prof. Dr. Helge Löbler
Lehrstuhl für BWL, insb. Marketing

Arbeitspapiere

Herausgeber:
Univ.-Professor
Dr. Helge Löbler

Helge Löbler/Petra Buchholz/Helge Petersohn

Neural Networks as Competitors for methods
Of data reduction and classification in SPSS

Arbeitspapier Nr. 2
Leipzig, Juni 1996

Universität Leipzig

Wirtschaftswissenschaftliche Fakultät
Lehstuhl für BWL, insb. Marketing

Marschnerstr. 31
04109 Leipzig
Tel.: 0341-9733750/751
Fax: 0341-9733759
Email: loebler@hpswifa.wifa.uni-leipzig.de

NEURAL NETWORKS AS COMPETITORS FOR METHODS OF DATA REDUCTION AND CLASSIFICATION IN SPSS

Löbler, H., Buchholz, P., Petersohn, H.

Prof. Dr. Helge Löbler

Lehrstuhl für Betriebswirtschaftslehre, insb. Marketing,

Universität Leipzig, D - 04109 Leipzig, Marschnerstraße 31, Germany

Dipl.Ök. Petra Buchholz

Lehrstuhl für Betriebswirtschaftslehre, insb. Marketing,

Universität Leipzig, D - 04109 Leipzig, Marschnerstraße 31, Germany

Dr. Helge Petersohn

Institut für Wirtschaftsinformatik

Universität Leipzig, D - 04109 Leipzig, Marschnerstraße 31, Germany

NEURAL NETWORKS AS COMPETITORS FOR METHODS OF DATA REDUCTION AND CLASSIFICATION IN SPSS

Abstract: The main purpose of this paper is to demonstrate the data reduction technique of self-organizing maps and to compare it with data reduction techniques in SPSS. Especially, factor analysis and multidimensional scaling (MDS) are chosen. Subsequent to data reduction a cluster analysis was conducted. Due to taking the same cluster algorithm on the base of different data reduction approaches we can compare the final outputs of the cluster algorithm in respect to a target criterion. This is the homogeneity within the groups compared to the homogeneity between the groups. The application example is taken from literature (Backhaus et. al. 1994).

1. INTRODUCTION

In contemporary research neural networks (NNW) are of growing interest for marketers. They are used as alternatives to or in combination with classical multivariate methods. Prior research was focused mainly on classification, positioning or segmentation problems (Hruschka/Natter 1993; 1995). Most of the application examples dealt with supervised learning techniques like backpropagation and the confrontation with discriminant analysis or regression analysis (Hruschka 1991; Hruschka/Natter 1995; Jung/Wiedmann 1994; Rittinghaus-Mayer 1993). In this kind of analysis a "teacher" is used to find out wether the output created by the NNW is a good one or not. Particularly, the system minimizes the error between the predicted output of the NNW and the real objects or a prior classification.

An application example of unsupervised learning was presentated by Mazanec (1995a; 1995b). For identifying competition among European cities he used an unsupervised NNW. Mazanec tried to figure out similarities and differences of different European cities using a self-organizing map. One problem that occurs by choosing unsupervised NNWs and comparing it with classical factor or cluster analysis is to find out which output among all approaches is the best or the most useful one. If one compares an unsupervised neural network with for example factor analysis there normally is a lack of a benchmark the output of NNWs as well as of factor analysis can be compared with. Therefore the function of the "teacher" is obsolete. The same is true, if we compare outputs of neural networks with cluster analysis output. There is nothing like an objective classification or an a-priori-classification. Therefore, Mazanec consequently did not value his results better or worse compared with classical methods.

As already mentioned implicitly, we can use neural networks for at least two purposes: one is to reduce the dimensions of the original data matrix and the second is to build groups or classes of objects (Rehkugler et. al. 1995; Schöneburg et. al. 1990). The first purpose is similar to classical factor analysis or multidimensional scaling. The second purpose is comparable with classical cluster analysis.

Data reduction means that an input data vector of higher dimensionality is transformed into a vector of lower dimensionality. That poses three consequences: a reduction of information, the necessity to interpret the latent dimensions behind the data and the increase of graphicness, which is important for practical application problems. Especially, for purposes of visualization it is often usefull to reduce original data matrices in a two-dimensional space. After the reduction of the original data matrix into the tranformed data matrix with two columns it is possible to use this reduced data matrix for the cluster analysis.

Classification means to bundle objects in that way that they are homogeneous within a class and heterogeneous between classes. In Marketing classification can be used for multiple purposes. For instance, to position rival brands due to statements of consumers about the perception of that brands. If marketers try to figure out market segments they usally use cluster analysis for building groups for products or consumers. For an easier interpretation of these segments or clusters marketers often base this cluster analysis on a reduced data matrix and for the reduction of dimensions they normally use factor analysis or multidimensional scaling.

The main purpose of this paper is to demonstrate the data reduction technique of self-organizing maps (SOM) and to compare it with data reduction techniques in SPSS. Especially, factor analysis and multidimensional scaling (MDS) are choosen. Since data reduction is often used as a base for classification or for building classes and not for it's own purpose we will evaluate the output of data reduction in using it as an input for clustering the original objects and measuring the quality of these clusters in this paper. Due to taking the same cluster algorithm on the base of different data reduction approaches we can compare the final output of the cluster algorithm in respect to a target criterion. This is the homogeneity within the groups compared to the homogeneity between the groups. Although we have a target criterion for evaluating the cluster results based on the different data reduction approaches, it is not possible to use a supervised neural network for this purpose because we only have a criterion, but not a real solution the generated solution can be compared with.

2. RESEARCH DESIGN

For analysis we used the application example of Backhaus et. al. (1994) as input data. The input data matrix is shown in table 1:

SF	PR	HB	UF	BE	GE	KG	TF	VG	NK	
4,500	4,000	4,375	3,875	3,250	3,750	4,000	2,000	4,625	4,125	SANELLA
5,167	4,250	3,833	3,833	2,167	3,750	3,273	1,857	3,750	3,417	HOMA
5,059	3,824	4,765	3,438	4,235	4,471	3,765	1,923	3,529	3,529	SB
3,800	5,400	3,800	2,400	5,000	5,000	5,000	4,000	4,000	4,600	DELICADO
3,444	5,056	3,778	3,765	3,944	5,389	5,056	5,615	4,222	5,278	HOLLBUTT
3,500	3,500	3,875	4,000	4,625	5,250	5,500	6,000	4,750	5,375	WEIHBUTT
5,250	3,417	4,583	3,917	4,333	4,417	4,667	3,250	4,500	3,583	DUDARFST
5,857	4,429	4,929	3,857	4,071	5,071	2,929	2,091	4,571	3,786	BECEL
5,083	4,083	4,667	4,000	4,000	4,250	3,818	1,545	3,750	4,167	BOTTERAM
5,273	3,600	3,909	4,091	4,091	4,091	4,545	1,600	3,909	3,818	FLORA
4,500	4,000	4,200	3,900	3,700	3,900	3,600	1,500	3,500	3,700	RAMA

Table 1: Input data matrix

The example reflects the evaluation of eleven margarine and butter brands (objects) by ten features (variables) as means of thirty respondents. The features like price (PR), durability (HB) or calorie content (KG) were measured on a 7-point-rating scale from 1 "low" to 7 "high".

Figure 1 demonstrates the research design:

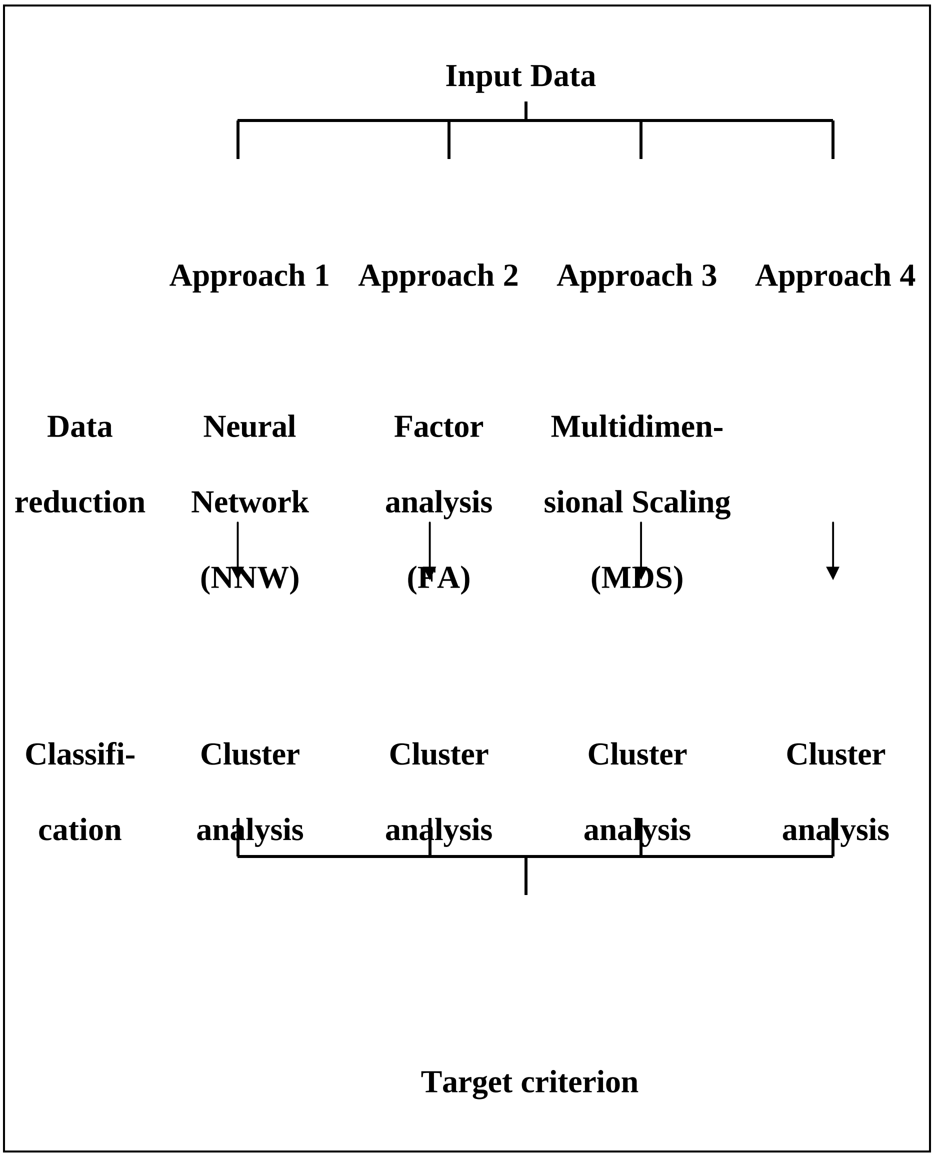

Figure 1: Research design

In a first step (data reduction) we confronted a neural network of the Kohonen type (approach 1) with factor analysis (approach 2) and multidimensional scaling (approach 3). Here the step of data reduction was performed and the output was used as an input for cluster analysis. To make sure that data reduction doesn't present a distorted picture of the input data, we have added approach 4 in a control function where we dropped the data reduction and used the original data matrix as an input for cluster analysis. In the second step we conducted a cluster analysis with the Ward procedure.

Because of the absence of a grade or target criterion for data reduction techniques the quality of the results are measured after cluster analysis. As a grade criterion we used the quotient of homogeneity within groups and homogeneity between groups:

$$g(R) = \frac{h_o}{h_e} \quad \text{with}$$

$$h_o = \sum_{p=1}^{r} \sum_{x_i, x_j \in K_p} d(x_i, x_j) \qquad i \neq j$$

$$h_e = \sum_{k=1}^{r} \sum_{p>k}^{r} \sum_{\substack{xj \in K_k \\ > x_i \in K_p}} d(x_i, x_j) \qquad i \neq j$$

$$d(x_{i,} x_j) = \sqrt{\sum_{l=1}^{m} \left(x_{ie} - x_{je}\right)^2} \qquad i \neq j$$

h_o = sum of all distances within a class, added over all classes

h_e = sum of all distances between objects belonging to different classes, added over all class pairs

$d(x_i,x_j)$ = Euklidean distance between objects x_i and x_j

r = number of classes of a Classification R

K_p = Class p

m = number of variables

l = index of variables

i,j = indices of objects

As neural network we choosed the concept of self-organizing maps (SOM) of the Kohonen type, provided by NeuralWorks. To permit deeper insights into this concept, a description follows below:

At first, during the learning phase for each neuron u, the Euklidean distance between the weight vector and the normalized object vector is calculated (net_u).

$$net_u = \sqrt{\sum_{l=1}^{m}\left(w_{ul} - o_{il}\right)^2} \text{ with} \qquad (1)$$

$$o_{il} = \frac{x_{il}}{\sqrt{\sum_{l=1}^{m} x_{il}^2}}$$

w_{ul} = components of the weight vector to neuron u

o_{il}= value of variable l for the normalized vector of object i

Second, during training the node with the smallest distance and the neighbour nodes adjust their weights to be closer to the values of the input data. More similar vectors are closer to each other than less similar ones.

$$\Delta w_{ul} = \sigma(o_{il} - w_{ul})$$

$$w_{ul\ new} = w_{ul\ old} - \Delta w_{ul}$$

σ = learning rate

The steps 1 and 2 are repeated until approximately the thirtyfold of the number of objects is achieved. In this example we have 330 learning steps.

Third, during the recall phase, the weight matrix generated during step 1 and 2 is used to calculate a representation function for each object as mentioned in (1). The smallest neuron net_{ui} of object i as well as two of its neighbours are selected for the topological representation on the map.

Forth, the topological representation of the objects on the map (two-dimensional) allows the transformation into a two-dimensional grid, because neighbourhood nodes or neurons represent similar objects. Thus all objects were represented in a two-dimensional space which was used as input for cluster analysis.

3. RESULTS AND DISCUSSION

We choosed all options within the factor analysis provided by SPSS. For comparability the number of factors in factor analysis was fixed on two. To be "fair" in this comparison we selected the option with the best result which was confrontated with the best net-work. For factor analysis the principal component analysis with varimax rotation has provided best results. For MDS the default options were taken. Serveral neural network configurations were tested. The number of nodes or neurons varied from 2 x 2 to 10 x 10 with learning rates of 0.5 and 0.8. The neural network with 5 x 5 neurons and a learning rate of 0.8 has succeeded in respect to our target criterion. This network was selected for the comparison. Because the number of classes is unknown, 2 to 5 classes-solutions were generated.

Results of the approaches, shown in figure 1, are presented below:

	Approach 1	**Approach 2**	**Approach 3**	**Approach 4**
Number of classes	g (R)	g (R)	g (R)	g (R)
5	0,0823499	0,169249	0,16148	0,116782
4	0,132797	0,189062	0,356912	0,199896
3	0,243243	0,23288	0,414738	0,481082
2	0,488132	0,558449	0,632062	0,55449

Figure 2: Homogeneity within classes and between classes for 2 to 5 classes

According to our research design the approach with the lowest value concerning target criterion g (R) is the best solution. Stressing this target criterion g (R), the neural network has succeeded for all cluster-solutions, except the 3-cluster solution. Here, the factor analysis is the "winner". Obviously, the neural network doesn't provide best results for all cases. It may surprise that the cluster analysis based on the original data matrix is not the best because first no information reduction took place and second the Ward procedure uses the squared distance. The Ward procedure does not guarantee the best result because of the agglomeration procedure.

4. LIMITATIONS OF THE STUDY AND RESEARCH OUTLOOK

On a simple and well known application example we have shown that neural networks can provide better results as classical multivariate methods for the purpose of data reduction and classification measured on a selected target criterion. Although this single application example doesn't allow any generalization, it gives hints for future research: We use the neural network for data reduction but it also can be used for building classes on the base of an original data matrix. Also it is possible to combine neural networks in reducting the data matrix and in building classes out of these reduced data matrix. One also can choose very different target criteria for measuring the results of the unsupervised learning. In former research it was shown that supervised neural networks can be very useful in market and marketing research. Our purpose was to use unsupervised neural networks in comparison with different multivariate analysis methods. As Mazanec (Mazanec 1995 a/b) we use the neural networks for data reduction, but in addition to this we used from a very practical point of view a criterion to find out which kind of data reduction performs best. To do so we had to add a classification procedure as we did. Especially for practical purposes this seems to be very helpful. At least we hope that this article motivates other researchers to use unsupervised neural networks in marketing research.

REFERENCES:

Backhaus, K./Erichson, B./Plinke, W./Weiber, R. (1994): *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 7th Edition, Berlin et.al. 1994.

Hruschka, H. (1991): *Einsatz künstlicher neuraler Netzwerke zur Datenanalyse im Marketing*. In: Marketing ZFP, Vol. 13 (1991), No. 4, p. 217 - 225.

Hruschka, H./Natter, M. (1993): *Analyse von Marktsegmenten mit Hilfe konnexionistischer Modelle*. In: Zeitschrift für Betriebswirtschaft, Vol. 63. (1993), No. 5, p. 42 - 442.

Hruschka, H./Natter, M. (1995): *Clusterorientierte Marktsegmentierung mit Hilfe künstlicher Neuronaler Netzwerke*. In: Marketing ZFP, Vol. 15 (1995), No. 4, p. 249 - 254.

Jung, H. H./Wiedmann, K. P. (1994): *Konnektionistische Prognosemodelle. Möglichkeiten und Probleme des Einsatzes neuronaler Netze im marketingorientierten Kontext*. Working paper No. 101, University of Mannheim 1994.

Mazanec, J. A. (1992): *Classifying Tourists into Market Segments: A Neural Network Approach*. In: Journal of Travel & Torism Marketing, Vol. 1, No. 1, 1992, p. 39 - 59.

Mazanec, J. A. (1995a): *Competition among European tourist cities: a comparative analysis with multidimensional scaling and self-organizing maps*. In: Tourism Economics, 1995, No. 1, p. 283 - 302.

Mazanec, J .A. (1995b): *Positioning Analysis with Self-Organizing Maps. An exploratory Study on Luxury Hotels*. In: Cornell Hotel and Restaurant Administration Quarterly. Vol. 36, No. 6, p. 80 - 95.

Rehkugler, H./Zimmermann, H. G. (Eds.) (1994*): Neuronale Netze in der Ökonomie*. München 1994.

Rittinghaus-Mayer, D. (1993*): Die Anwendung von Neuronalen Netzen in der Marketingforschung*. München 1993

SCHöneburg, E./Hansen, N./Gawelzyk, A. (1990): *Neuronale Netzwerke*. Haar/ München, 1990.

Keywords:

Neural Networks - Classification