Monografías del Seminario Matemático García de Galdeano **33**, 285–292 (2006)

ANALYSING A BICRITERIA CLUSTERING PROBLEM USING SOM

D. Lahoz, P. Mateo and F. Mallor

Abstract. In this work we consider a clustering problem, that is, grouping a large set of elements described by a large set of non-negative variables into homogeneous categories that are not predefined. The novelty is that besides the quantitative criterion to differentiate variables, as usual, there is also a qualitative criterion, that takes into account whether or not the variables take the value zero . This problem was presented and studied by means of a certain family of parametric transformations of the data set and also by means of Multiple Factor Analysis in [1]. In our work, we consider these parametric transformations of the data set and we use Self-Organizing Maps ([4]) as clustering technique. We compare the results with those obtained in [1]. This problem came up when the managers of a big telecommunications company wanted to know about the structure of their clients' portfolio identifying different consumption profiles. The proposed methods have been tested on artificial data sets with similar statistical properties to those found in the real one.

Keywords: SOM, K-means, clustering, data mining.

AMS classification: 82C32,68T10,91C20.

§1. Introduction

Clustering problems deal with grouping individuals or elements into homogeneous groups. The basic statement is as follows, a set of items, with characteristics of interest described by several variables, is available. Then, a Decision Maker wants to divide this set into a certain number of subgroups that share similar characteristics. The clustering problem that we address in this paper arose in a telecommunications company in which the managers wanted to know more details about their clients' portfolio. More specifically, they wanted to study the habits in relation to the consumption of their services in order to obtain “customers' profiles”.

Given that clustering problems play an important role in business, government, health, industry, among others, much research effort has been devoted to develop efficient algorithms (see a review of them in the Gordon's book, [3]). The novelty of our clustering problem is that the set of variables is simultaneously used to compare among the individuals in order to define their affinity to be integrated into the same cluster under two different criteria, of qualitative and quantitative nature. Since we have not found in the literature any analysis of a similar problem, a methodology based on prior transformation of the data was used and then followed, in one approach, by classical (statistical) clustering techniques and, in another approach, by neural networks techniques.

In [1] the authors presented the methodology and results obtained with the classical approach, being also compared with the results obtained with a methodology based in the Multiple Factor Analysis (MFA).

In this work, we address this clustering problem from the neural network point of view and compare the results with those obtained with the classical approaches exposed in [1].

The paper is organised as follows, in section 2 the problem is established. In section 3, a brief summary of the results obtained in [1] are related. Then, in section 4, the methodology of neural networks that we have used is presented and the results obtained are showed. Finally, the last section presents the conclusions and remarks.

§2. Problem statement and methodology

As we have said before, the problem came up when the managers of a big telecommunications company wanted to identify the different consumption profiles of their clients. Thus, the main objective of the study was to obtain the typology of the customers attending to their consumption pattern.

The company provided us with a big set of customers described by a large set of quantitative variables. For each customer i, $i = 1,\dots,N$, we knew his consumption level X_{ij} for service j, $j = 1,\dots,J$, during a certain period of time. For each client i, the values of X_{ij} constitute his complete consumption pattern.

The customer categories should be useful for identifying whether or not a client is a consumer of a given product, so customers in the same cluster should coincide in the type of products they consume. But it is also important to reflect their levels of consumption for the different services.

When a standard clustering technique is used to cluster consumers based on their consumption levels, only a quantitative criterion is applied. But, when customers are clustered according to a binary table obtained from the original data, representing the consumption or non-consumption of the services, only a qualitative criterion is applied. Neither of these solutions is satisfactory, if the aim is to consider both criteria simultaneously. Therefore, it was necessary to develop procedures to handle this clustering problem as a multicriteria one.

The simultaneous consideration of both criteria is handled in two different ways:

Prior transformation of data

This approach consists of transforming the original data using certain parametric family of functions. We have two objectives to be held: to separate the zero value from the others and to grade the different levels of consumption.

The proposed transformation is the following:

$$T(X) = X^{\lambda}, \quad \lambda \in [0,1]. \tag{1}$$

For $\lambda = 1$ the original data are maintained, for $\lambda = 0$ the binary use/non-use data are get. Any value between 1 and 0 grades the level of importance from the consumption level criterion until the consumption/non-consumption criterion.

Multiple Factorial Analysis

MFA works with continuous variables as well as with categorical ones (see [2]). In our case, the set of quantitative variables is the original one and the set of categorical variables is defined from the original one by coding each value of X_{ij} greater than or equal to 1 into 1. Thus, we have J new categorical variables, Y_j that indicate the consumption/non-consumption for each service. Each binary variable Y_j is associated with one of the variables X_j, and it is defined as

$$Y_{ij} = \begin{cases} 0, & \text{if } X_{ij} = 0, \\ 1, & \text{if } X_{ij} > 0. \end{cases}$$

MFA requires that each category of the new variables forms a separate column δ_1^j and δ_2^j in the table of data. This new table (including the quantitative and qualitative variables), is analysed with MFA to identify the factors and then, a cluster analysis is done on these factors.

The distance between individuals i and i' can be expressed by means of the weighted combination of two distances:

$$d^2\left(i,i'\right) = \alpha d_I^2\left(i,i'\right) + \beta d_{II}^2\left(i,i'\right),$$

where d_I is a distance in the space of the quantitative variables and d_{II} is a distance in the space of the qualitative variables. Then, these distances are a measure of similarity or proximity between customers, according to each one of the two criteria. α and β are weighting coefficients, depending on the highest axial inertia for the set of quantitative and qualitative variables, respectively, which represent the relative importance of each criterion. Thus, we have no option to choose these weights because they are automatically determined by the method.

§3. Results from the analysis by means of classical clustering techniques

Confidentiality requirements prevent us from revealing real data and conclusions, so, the proposed methods have been tested on artificial data sets with similar statistical properties to those found in the real data set. Therefore, in the simulation of the new variables the following facts were considered:

- The original data set included subsets of variables strongly correlated among them but with a low correlation, in general, with variables in the other subsets. Then, we have defined 2 subsets with 2 and 3 variables, respectively, according to this correlation pattern.

- The original variables were characterised by a high percentage of zero values (meaning non consumption of the associated service). We have simulated our variables with a similar percentages of zero values.

- The histograms of the original variables were, in general, very asymmetric, skewed on the right and including extreme values (in comparison with the mean value). We kept this characteristic in our simulated data.

CMA	MFA	$\lambda = 1$	$\lambda = 0.9$	$\lambda = 0.8$	$\lambda = 0.7$	$\lambda = 0.6$	$\lambda = 0.5$	$\lambda = 0.4$	$\lambda = 0.3$	$\lambda = 0.2$	$\lambda = 0.1$	$\lambda = 0$
a)	21.30	56.90	41.30	38.20	32.90	30.90	28.20	25.50	22.60	19.80	19.10	14.10
b)	26.80	20.20	21.70	22.40	25.50	27.00	25.90	26.40	25.80	23.50	24.00	17.40
c)	18.10	22.90	23.00	23.70	21.70	20.10	19.00	17.70	17.70	19.50	20.00	22.60
d)	33.80	0	14.00	15.70	19.90	22.00	26.90	30.40	33.90	37.20	36.90	45.90
B/W		0.848	1.8014	1.8745	1.8081	1.571	1.7255	2.1466	2.2121	2.1413	2.1556	1.417

Table 1: Clasical Multivariate Analysis (CMA), MFA and transformation data. B/W stands for Between/Within.

	C_{a9}	C_{b9}	C_{c9}	C_{d9}
C_{a10}	413	46	110	0
C_{b10}	0	171	1	30
C_{c10}	0	0	119	110

	C_{a3}	C_{b3}	C_{c3}	C_{d3}
C_{aMFA}	211	1	1	0
C_{bMFA}	10	170	1	0
C_{cMFA}	5	1	252	10
C_{dMFA}	0	5	4	329

Table 2: Common individuals between clusters. Left: $\lambda = 0.9$ and $\lambda = 1$. Right: $\lambda = 0.3$ and MFA.

In the MFA approach as well as in the data-transformation approach, a similar process is followed. First, an initial analysis is done in order to get the main factors (PCA, MCA,...). Then, a mixed clustering technique implemented in SPAD software ([8, 5]) is accomplished to get the groups of clients. In Table 1, the clusters obtained with MFA analysis and with data-prior transformation (values of $\lambda = 0, 0.1, 0.2, \ldots, 1$) are showed. Each row is associated with a different cluster. Each column corresponds to a different classification. The numbers in the body of the table are the percentages of individuals included in the cluster.

When MFA technique and the data transformation with $\lambda = 0, 0.1, \ldots, 0.9$ are applied, the number of optimal groups found is 4. And when the data transformation with $\lambda = 1$ is considered, the optimal number of groups is 3. The following two tables show the disaggregation from 3 to 4 groups that correspond to $\lambda = 1$ and $\lambda = 0.9$, respectively, and the number of coincidences between $\lambda = 0.3$ and MFA. MFA and $\lambda = 0.3$ generate the most similar groups, the percentage of coincidences reaches the value 96.6%.

§4. Analysis by means of neural networks techniques

SOM networks([4, 6, 7]) are used in order to get a reduction of the dimension of the data, building a two-dimensional map (hexagonal, 16×10), that represents the original data. After getting this map, a *k-means* algorithm ([10]) is used to build a classification of the elements in the map. This classification allows us to obtain the centroids of each group and then, each datum is associated with the group with the nearest centroid. This process gives us the classification of each datum. The operations have been accomplished with SOM toolbox for

SOM stand.	$\lambda=1$	$\lambda=0.9$	$\lambda=0.8$	$\lambda=0.7$	$\lambda=0.6$	$\lambda=0.5$	$\lambda=0.4$	$\lambda=0.3$	$\lambda=0.2$	$\lambda=0.1$	$\lambda=0$
a)	58.10	40.80	38.20	34.30	31.30	28.70	26.20	23.10	20.00	20.20	16.10
b)	19.90	20.80	22.40	23.00	27.60	26.60	25.90	26.60	23.20	22.20	23.00
c)	22.00	23.10	23.90	25.10	21.10	21.00	18.40	18.10	19.00	19.70	18.70
d)	0	15.30	15.50	17.60	20.00	23.70	29.50	32.20	37.80	37.90	42.20
B/W	0.9097	1.2951	1.3186	1.3659	1.4463	1.4948	1.5496	1.7004	1.760	1.6396	1.492

Table 3: Percentages of individuals in each group. Standardized measures

SOM no stand.	$\lambda=1$	$\lambda=0.9$	$\lambda=0.8$	$\lambda=0.7$	$\lambda=0.6$	$\lambda=0.5$	$\lambda=0.4$	$\lambda=0.3$	$\lambda=0.2$	$\lambda=0.1$	$\lambda=0$
a)	57.80	45.50	43.80	37.60	32.30	28.00	24.30	21.40	20.00	19.30	20.50
b)	25.80	18.00	18.30	20.20	21.30	21.70	21.20	19.70	19.00	20.20	21.30
c)	16.40	23.50	23.00	20.40	21.20	21.10	20.00	19.80	19.40	20.10	13.80
d)	0	13.00	14.90	21.80	25.20	29.20	34.50	39.10	41.60	40.40	44.40
B/W	2.097	3.4784	3.2543	2.4407	2.479	3.0568	2.9320	2.917	2.657	2.1380	1.372

Table 4: Percentages of individuals in each group. Unstandardized measures

Matlab 5 ([11]).

The above steps are accomplished for each set of transformed data considering standardised and non-standardised measures.

In Tables 3 and 4, the results obtained with the data-transformation approach and neural networks are showed. In the first table standardised measures are considered, and in the second one unstandardised measures are considered.

In these tables, it appears a row named B/W (*Between/Within*). It represents the ratio of the distance between groups and the distance of individuals within the groups. For these values, the greater the better. The Between distance is the Between-cluster sum of squares and its expression when considering k clusters is as follows ([9]):

$$B(k)=\sum_{l=1}^{k}\mathrm{dist}^2(c_l,m).$$

The expression for the Within distance, the Within-cluster sum of squares, is

$$W(k)=\sum_{l=1}^{k}\sum_{i=1}^{n_l}\mathrm{dist}^2(c_l,x_i^l),$$

where dist is the Euclidean distance, k is the number of clusters, n_l is the number of elements in cluster l, c_l is the centroid of cluster l, m is the global mean, and x_i^l is the element i of cluster l.

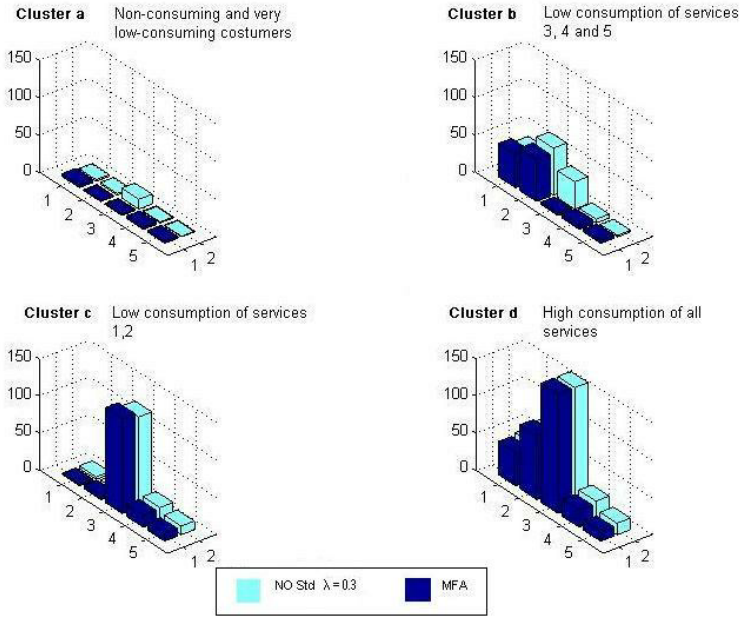

Figure 1: Patterns of consumption

Comparison of results

If we compare the results obtained with classical clustering techniques and with neural networks using standardised measures, we can observe that the classical clustering technique gets better results. On the other hand, when comparing classical clustering techniques to neural network techniques using unstandardised measures, the best results are obtained with the neural techniques (except for $\lambda = 0$, qualitative criterium, and $\lambda = 0.1$). Furthermore, the difference between ratios is bigger in this second comparison. Then, the neural network approach with unstandardised measures exhibits a higher capacity of separation between groups and of grouping individuals within the groups.

Finally, Figure 1 shows the pattern of consumption of the four groups obtained when we consider MFA method and neural network method with unstandardised measures and $\lambda = 0.3$ (the most similar to MFA). It can be observed that the groups are very similar. Slight differences appear in cluster b), where neural network approach includes individuals with a higher use of service 3.

§5. Conclusions

In this paper we have presented a real-world clustering problem under a bicriteria modelisation. The problem has been treated with Classical Statistical Multivariate Analysis as well as with Self- Organizing Maps with k-means methods. In both cases, they have showed to be valid tools

The development includes a systematic approach, MFA, and a family of prior transformations of the data that takes into consideration the two main aspects of the problem, qualitative and quantitative criteria. The parameter of the transformation can be interpreted as the relative importance given to the quantitative criterium against the qualitative criterium, ranging from 1 (only the quantitative criterium is taken into account) to 0 (only the qualitative criterium is considered).

The results obtained with the two methodologies are compared. This comparison shows that the behaviour of SOM with k-means is so good as the obtained with Classical Multivariate Analysis. Moreover, the results show that the use of unstandardised data with SOM with k-means leads us to a more discriminated clustering than the Classical Multivariate Analysis and MFA.

Currently, we are working in the development of interactive methods for the selection of the parameter λ in order to collect the preferences of the decision makers. Also, we are considering the use of different transformations.

References

[1] ABASCAL, E., GARCÍA-LAUTRE, I., AND MALLOR, F. Data mining in a bicriteria clustering problem. *European Journal of Operational Research* (in press).

[2] ESCOFIER, B., AND PAGÈS, J. *Analysis Factorielles Simples et Multiples*. Ed. Dunod. Paris, Paris, 1990.

[3] GORDON, A. D. *Classification*. Chapman & Hall/CRC, Second Edition, London, 1999.

[4] KOHONEN, T. *Self- Organizing Maps*. Springer, Berlin, 1997.

[5] LEBART, L., MORINEAU, A., LAMBERT P., AND PLEUVRET, P. *Manuel de Référence*. SPAD, CISIA-CERESTA, 1999.

[6] MARTÍN DEL BRIO, B., AND SANZ, A.. *Redes Neuronales y Sistemas Borrosos*. Ra-Ma, Madrid, 2001.

[7] OJA, E., AND KASKI, S. *Kohonen Maps*. Elsevier, Amsterdam, 1999.

[8] SPAD. *SPAD. Logiciel diffusé par CISIA*. 1 av. Herbillon 94160, Saint-Mandé, France.

[9] TIBSHIRANI, R., WALTHER, G., AND HASTIE, T. Estimating the number of cluster in a data set via the gap statistic. *J.R. Statist. Soc.B 63*, 2 (2001), 411–423.

[10] VESANTO, J., AND ALHONIEMI, J. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks 11*, 3 (May 2000), 586–600.

[11] VESANTO, J., HIMBERG, J., ALHONIEMI, E., AND PARHANKANGAS, J. SOM Toolbox for Matlab 5. Report A57, Espoo, Finland, 2000.

David Lahoz and Pedro Mateo
Departamento de Métodos Estadísticos
Universidad de Zaragoza C/ Pedro Cerbuna 12, 50009 Zaragoza, España
`davidla@unizar.es` and `mateo@unizar.es`

Fermín Mallor
Departamento de Estadística e Investigación Operativa
Universidad Pública de Navarra, Campus Arrosadia, 31006 Pamplona, España
`mallor@unavarra.es`