AN INTER-MODELS DISTANCE FOR CLUSTERING UTILITY FUNCTIONS[1]

Elvira Romano, Carlo Lauro

Dipartimento di Matematica e Statistica, Università di Napoli Federico II, Complesso Universitario di Monte S. Angelo, Via Cinthia, I - 80126 Napoli - Italy
elvroman@unina.it; clauro@unina.it

Giuseppe Giordano

Dipartimento di Scienze Economiche e Statistiche Università di Salerno
Via Ponte Don Melillo, 84084 Fisciano (Sa), Italy
ggiordan@unisa.it

Abstract

Conjoint Analysis is one of the most widely used techniques in the assessment of the consumer's behaviors. This method allows to estimate the partial utility coefficients according to a statistical model linking the overall note of preference with the attribute levels describing the stimuli. Conjoint analysis results are useful in new-product positioning and market segmentation. In this paper a cluster-based segmentation strategy based on a new metric has been proposed. The introduced distance is based on a convex linear combination of two Euclidean distances em bedding information both on the estimated parameters and on the model fitting. Market segments can be then defined according to the proximity of the part-worth coecients and to the explicative power of the estimated models.

Key words: Multiattribute Preference Data, Conjoint Analysis, Cluster Analysis, Market Segmentation.

---

[1] *This paper was financially supported by MIUR grants: "Multivariate Statistical and Visualization Methods to Analyze, to Summarize, to Evaluate Performance Indicators", coordinated by prof. M.R. D'Esposito and "Models for Designing and Measuring Customer Satisfaction", coordinated by prof. C.N. Lauro.*

1. INTRODUCTION

Conjoint analysis is a statistical technique useful to explain and predict consumer preferences (Green *et al*., 1990). In this framework, the overall notes of preference given to a set of attributes, are modeled through *utility functions*. They represent the analytical formulation that defines the relationship among the attribute-levels and allow to determine the part-worth coefficients for each consumer. Conjoint results are widely used in the field of Market Segmentation to identify consumers group sharing similar taste and behavior with respect to different products or services. A thorough discussion of several issues in the aggregation of utilities in market simulation can be found in Gustafsson (2001).

The attributes can be modeled either as linear, quadratic function, or as main effects in a regression model with dummy independent variables (part-worth model When the part-worth model is assumed, the preference y_j toward the $j-th$ stimulus is expressed as linear combination of the coded attribute-levels and the part-worth parameters. A known drawback is the unreliability of the estimated coefficients when using high reduced factorial designs. Too few degrees of freedom available in the estimation phase, at individual level, affect data reliability and may lead to classification errors (Vriens *et al*., 1996). This paper examines the effect of introducing the information about the explicative power of the models in a market segmentation strategy. At this aim, in section 2, we propose a new strategy for discriminating among estimated models. It is based on different kinds of information in the definition of the model distance. We introduce an Inter-Models (IM) distance which is able to take into account both the analytical structure of the models − through the difference between the estimated parameters − and the information about the model fitting through the difference between the adjusted $Adj-R^2$ indexes related to each pair of models. From this, in section 3, a suitable classification strategy highlighting the distance properties is showed. Finally, the main results of the proposed approach are illustrated in section 4 by comparing simulated datasets.

2. THE DATA STRUCTURE AND THE INTER-MODELS DISTANCE

Let consider the collection of utility models $M = \left\{m^1, \ldots, m^j, \ldots, m^J\right\}$, where each entry m^j is a K-dimensional vector defined as:

$$m^j = \left(w_1^j, \ldots, w_k^j, \ldots, w_K^j\right) \tag{1}$$

where the value of w_k^j is the information related to the j-th model. The first $(K-1)$ values are the estimated model parameters, the K-th value is the information related to the model fitting. For each of the J fitted utility models the part-worth coefficients $\left(w_1^j,\ldots,w_k^j,\ldots,w_{K-1}^j\right)$ are assumed to be estimated by Ordinary Least Square *(OLS)*. As with the classical cluster based segmentation strategies we aggregate the estimated parameters to define market segments. Moreover, we use a statistical index of model fitting (i.e. the $Adj-R^2$), as supplementary information about the utility functions, in order to exploit the actual predictive power of the models. Thus, the data collection M (Table 1) consists of two kinds of information: the analytical terms and the statistical model fitting.

Tab.1: The data collection.

Utility models	Coefficients	Model fitting
Model 1	$w_1^1,\ldots,w_k^1,\ldots,w_{K-1}^1$	w_K^1
…	…	…
Model j	$w_1^j,\ldots,w_k^j,\ldots,w_{K-1}^j$	w_K^j
…	…	…
Model J	$w_1^J,\ldots,w_k^J,\ldots,w_{K-1}^J$	w_K^J

The two information are combined to define the following measure:

$$IM(m^j,m^{j'}|\lambda) = \lambda IM_p + (1-\lambda)IM_r \qquad (2)$$

with $\lambda \in (0,1]$. The IM measure is a convex combination of two quantities IM_p and IM_r, where IM_p is the L_2-norm between the estimated parameters:

$$IM_p = \left[\sum_{k=1}^{K-1}\left(w_k^j - w_k^{j'}\right)^2\right]^{\frac{1}{2}} \ (j \neq j') \qquad (3)$$

and IM_r is the L_1-norm between the $Adj-R^2$.

$$IM_r = \left|w_K^j - w_K^{j'}\right| \ (j \neq j') \qquad (4)$$

Let us consider J models m^j from an arbitrary input space Ω, the function

$$IM(m^j,m^{j'}|\lambda) : \Omega \times \Omega \rightarrow R^+$$

satisfies the following conditions:

1. $IM(m^j,m^{j'}|\lambda) \geq 0$ *and* $IM(m^j,m^{j'}|\lambda) = 0 \ \forall \ m^j = m^{j'} \in \Omega$

2. $IM(m^j, m^{j'}|\lambda)$ *is symmetric, i.e.* $IM(m^j, m^{j'}|\lambda) = IM(m^{j'}, m^j|\lambda)$
3. $IM(m^j, m^{j'}|\lambda) \leq IM(m^j, m^{j*}|\lambda) + IM(m^{j^*}, m^{j'}|\lambda)\ \forall\ m^j, m^{j'}, m^{j^*} \in \Omega$

The properties are a trivial consequence of the *IM* definition as a convex combination of two distances. In particular, the third property can be proved as follows. In the definition of the *IM* distance (2), for the first term we have for each $j \neq j' \neq j^*$:

$$\lambda \left[\sum_{k=1}^{K-1} \left(w_k^j - w_k^{j'}\right)^2\right]^{\frac{1}{2}} \leq \lambda \left\{ \left[\sum_{k=1}^{K-1} \left(w_k^j - w_k^{j^*}\right)^2\right]^{\frac{1}{2}} + \left[\sum_{k=1}^{K-1} \left(w_k^{j^*} - w_k^{j'}\right)^2\right]^{\frac{1}{2}} \right\} \tag{5}$$

while, for the second adding term we have:

$$(1-\lambda)\left|w_K^j - w_K^{j'}\right| \leq (1-\lambda)\left\{\left|w_K^j - w_K^{j^*}\right| + \left|w_K^{j^*} - w_K^{j'}\right|\right\} \tag{6}$$

then from (5) and (6) we find:

$$\begin{aligned} ID(m^j, m^{j'}|\lambda) \quad \leq \quad & \lambda\left[ID_p(m^j, m^{j^*}|\lambda) + ID_p(m^{j^*}, m^{j'}|\lambda)\right] + \\ & +(1-\lambda)\left[ID_r(m^j, m^{j^*}|\lambda) + ID_r(m^{j^*}, m^{j'}|\lambda)\right] \end{aligned} \tag{7}$$

that is

$$IM(m^j, m^{j'}|\lambda) \leq IM(m^j, m^{j^*}|\lambda) + IM(m^{j^*}, m^{j'}|\lambda).$$

It follows that the defined function $IM(m^j, m^{j'}|\lambda)$ is a distance. The trimming parameter λ plays the role of a merging weight of the two components IM_p and IM_r. In the trivial case when $\lambda = 1$ the distance $IM(m^j, m^{j'}|\lambda = 1)$ is defined as a function of the part-worth coefficients (it is the usual approach in cluster based segmentation). However, it may happens that for some consumers the estimated utility function does not fit adequately. As a consequence, the derived segmentation could not be representative of the actual consumers' behavior. In this case, values of $\lambda \neq 1$ allow to recover such kind of information. We looks for a λ-value for the set of models, taking into account the explicative power of their theoretical preference models.

The definition of the *IM* distance (2) takes into account the different explicative power of each models so that, two models with similar estimated coefficients are mainly differentiated for their model fitting values. Of course, if two models have different coefficient values they should not be moved closer because of a similar fitting measure. For this reason, the trimmer value of λ should not be less than a minimum level. As a rule of thumbs, this cut off level is chosen as a function of the number of elements involved in the two part of the IM. For instance, if we estimated $K-1$ coefficients, the trimmering λ-value is settled not less than $\frac{1}{K-1}$. The cut off level protects against trivial classification results, since the IM_p is constrained to play the most important role in the whole distance.

3. CLUSTERING UTILITY FUNCTIONS

The market segmentation phase is based on a clustering method in the framework of multidimensional data analysis.

A hierarchical classification technique based on the *IM* distance and the Ward's aggregating criterion is carried out. The choice of the linkage criterion strongly influences the hierarchical tree structure, the Ward's method allows to gather minimum intra-classes variance and maximum inter-classes variance, it is the best suited for the proposed *IM* distance. The definition of the *IM* distance implies the choice of a suitable value for the trimmering parameter λ. The correct parameterization of the *IM* distance allows to put the right emphasis on the model fitting adequacy. Setting $\lambda < 1$ allows the IM_r part to enter in the definition of the inter-model distance and separate those models that even showing similar estimated coefficients, manifest varying accuracy due to different fitting. The λ-value is a parameter to be optimized for each given set M of estimated utility functions. The choice of the λ^* optimum is based on the capability of the $ID(m^j, m^{j'}|\lambda)$ to realize the best tree structure in the sense of the Cophenetic Coefficient, defined as follows:

$$Coph(m^j, m^{j'}|\lambda) = \frac{\sum_{m^j<m^{j'}} \left(IM_{m^j,m^{j'}} - \overline{IM}\right)\left(\widetilde{IM}_{m^j,m^{j'}} - \overline{\widetilde{IM}}\right)}{\left[\sum_{m^j<m^{j'}} \left(IM_{m^j,m^{j'}} - \overline{IM}\right)^2 \sum_{m^j<m^{j'}} \left(\widetilde{IM}_{m^j,m^{j'}} \overline{\widetilde{IM}}\right)^2\right]^{\frac{1}{2}}} \quad (8)$$

where $IM_{m^j,m^{j'}}$ is the distance between each pairs of rows in the matrix $M_{(J,K)}$ and $\widetilde{IM}_{m^j,m^{j'}}$ corresponds to the linkage distances between the objects paired in the clusters. Finally $\overline{IM}$, and $\overline{\widetilde{IM}}$ are, respectively, the averages of $IM_{m^j,m^{j'}}$ and $\widetilde{IM}_{m^j,m^{j'}}$.

This coefficient measures the linear correlation between the original *IM* distances and the linkage distance provided by the tree structure, it allows to measure how the data fits into the hierarchical classification tree. The steps of the algorithm, on which the strategy is based, can be summarized as follows :

1. *initialization phase*: the computation of IM_p and IM_r is carried out, and a grid of $\lambda \in \left[\frac{1}{K-1}, 1\right]$ is settled with a user defined granularity;
2. *classification phase*: for each λ the IM is computed, and the clustering algorithm is performed according to the Ward's linkage criterion. The linkage distances related to each tree structure (dendrogram) are recorded;
3. *optimization phase*: the linkage distances are compared with the original distance by means of the Cophenetic Coefficient. Therefore, the value of λ^* is selected so that the cophenetic coefficient is maximum;

The final partition is settled according to the distance:

$$IM(m^j, m^{j'}|\lambda^*) = \lambda^* IM_p + (1-\lambda^*) IM_r \quad (9)$$

As usual, the number of clusters is chosen by exploring the hierarchical tree structure derived from the (9).
In the next section a simulation study is carried out to show the main advantages of the proposed strategy.

4. SIMULATIONS STUDY

The proposed approach is illustrated by simulating several datasets under different conditions. The aim is to show how in presence of different model fitting, traditional clustering methods based on the estimated coefficients lead to unreliable clustering structure. Indeed, in this case the coefficients may give a biased image of the original preference data.
Two simulated data set are used to provide some useful insight into the procedure and the distance properties.

- *Models with different coefficients but similar fitting values.*
  In the first simulation study, we start considering the case of well fitted models. Three classes of preference functions are generated from a parent model:

$$y_i = w_0 + \sum_{k=1}^{K=4} w_i x_i + e_i \quad (i = 1, \ldots, K) \quad (10)$$

Each class is characterized by different coefficient values according to the following scheme:

Tab. 2: Simulation plan for the three classes of models with dierent coecients and similar fitting values.

	N	w_0	w_1	w_2	w_3	w_4	e_i
Class A	30	6.50	-1.33	1.00	1.25	-1.83	$N(0,1)$
Class B	30	6.50	0.50	-1.50	-2.25	0.33	$N(0,1)$
Class C	30	6.50	-0.50	-0.25	3.00	0.00	$N(0,1)$

In order to generate three sets of models with a quite good approximation, we build the global preference ratings according to the model (10) and with coecients given in Table 2. They are used as dependent variables in a multivariate multiple regression model with dummy explicative variables defined by the orthogonal experimental design in Table 3.

Tab. 3: Experimental design.

	Intercept	x_1	x_2	x_3	x_4
1	1	1	-1	-1	1
2	1	-1	-1	-1	1
3	1	1	1	0	1
4	1	-1	1	0	1
5	1	1	0	1	1
6	1	-1	0	1	1
7	1	1	-1	-1	-1
8	1	-1	-1	-1	-1
9	1	1	1	0	-1
10	1	-1	1	0	-1
11	1	1	0	1	-1
12	1	-1	0	1	-1

The estimated models give raise to three sets of similar coefficients with very good model fitting. In figure, 1 the box-plots of the $adj-R^2$ are reported, the intercepts and the four sets of estimated coefficients distributions.
The next step of the simulation study consists in:

1. setting a grid of h values for the trimmer value λ, with the constraint $0.2 < \lambda < 1$ and increments of 0.1 (i.e. $h = 9$);

2. computing, for each λ_i $i = 1,\dots,h$, the $IM_{(i)}$ distance among all pairs of the 90 models;

3. building the h tree structures according to the Ward's criterion;

4. obtaining, for each tree structure, the value of the cophenetic coefficient $Coph_i$;

5. choosing the tree structure for the largest value of λ such that $\max_\lambda (Coph_i \; i = 1, \ldots, h)$.

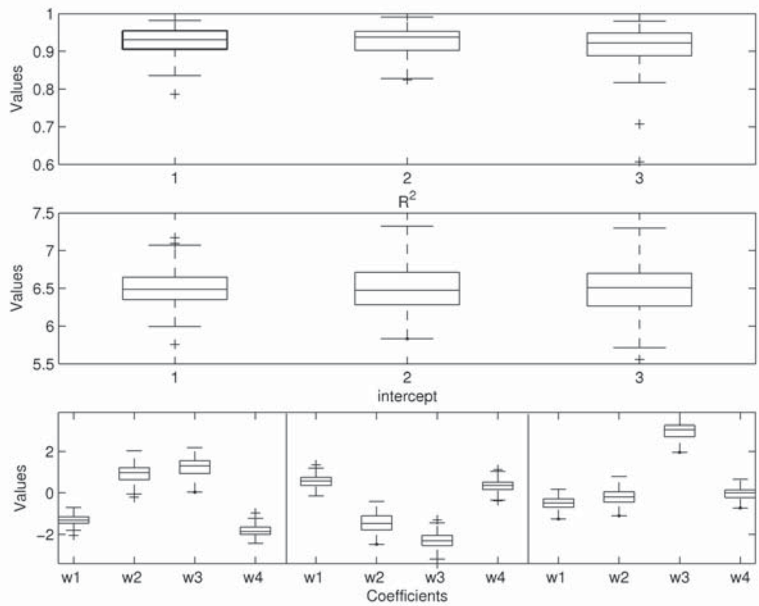

Fig: 1. Box-plot of the *adj* - *R*[2], the intercept and the four estimated coecients.

We obtain, at the optimum, for our models, the value of $Coph_i = 0.987$ for $\lambda = 0.7$. The related tree structure is shown in figure 2. The value of λ at the optimum indicates that the best classification result is obtained considering only the 70% of the distance due to the analytical part and 30% due to the fitting information. Of course, we expected a value of λ close to 1, because of the strong structure of the analytical part and the homogeneous values of goodness of fit (see above, figure 1). To validate this result, we have replicated the simulation study one hundred times under the same conditions.
Figure 3 shows the distribution of the values of λ at the optimum.

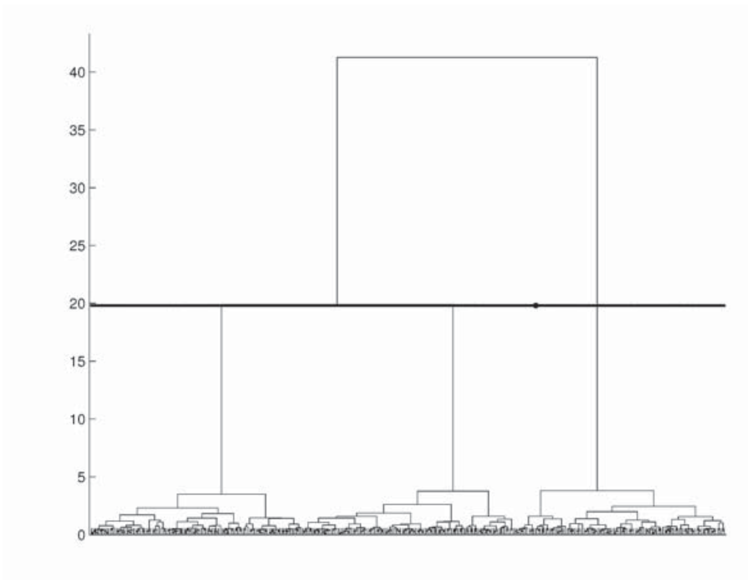

Fig. 2. The tree structure of the simulated models.

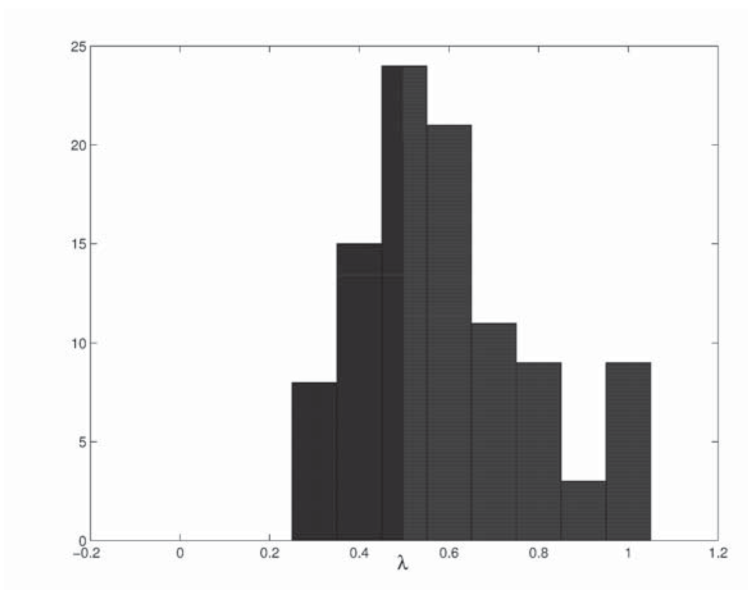

Fig. 3. Distribution of the optimum λ -values.

Note that almost the 50% of the tree structures have been obtained setting a values of λ less or equal to 0.5. The related cophenetic coefficients computed for every tree structures and for each λ_{ij} $(i = 1, \ldots, 9; j = 1, \ldots, 100)$ range from 0.96 to 0.98 indicating that good classification results could be obtained with a value of λ greater than 0.3.

- *Models with different coefficients and different fitting values.*

When there are several models with almost similar coefficients it is likely to consider them as belonging to the same market segment. However, the occurrence of different fitting could mask the presence of hidden structures of preference. In order to analyze the behavior of the defined models distance *IM* in this circumstance, a cluster structure has been generated varying the coefficient values according to the scheme in Table 4. Looking at this table, there are two classes which share

Tab. 4: Simulation plan involving a hidden tree structure.

	N.	w_0	w_1	w_2	w_3	w_4	e_i
Class A	20	6.50	-1.33	1.00	1.25	-1.83	$N(0, 1)$
Class B	20	6.50	-1.33	1.00	1.25	-1.83	$N(0, 3)$
Class C	20	6.50	-0.50	-0.25	3.00	0.00	$N(0, 1)$

the same coefficients but they have different fitting values due to error terms, and a third class representing a well separated segment. In this case the aim is to find the third structure that could not appear if we just look to the coefficient values. As in the previous case, we replicate the steps $1-5$ of the procedure one hundred times and record the value of λ related to the maximum value of the Cophenetic Coefficient. The results are shown in figure 4. As expected, the most part of the trimmering value λ is equal to 0.2. It corresponds to the maximum cophenetic coefficient value found in all replications. Finally, we have compared the two tree structures obtained for the *ID* distance with $\lambda = 0.2$ and $\lambda = 1$, figures 5 and 6 illustrate the differences. In particular, it is clear how the *ID* distance (with$\lambda = 0.2$) enhance the visualization of a three clusters structure in the dendogram.

5. CONCLUDING REMARKS

This article addresses the important issue of market segmentation in Conjoint Analysis. A new distance based both on the difference between the estimated

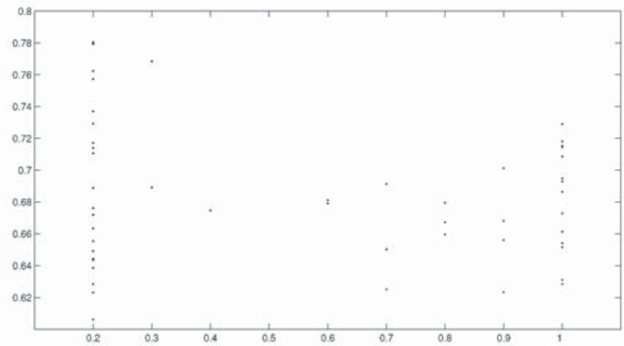

Fig. 4: Distribution of the λ -values corresponding to the maximum cophenetic coecient in 100 replications of the clustering.

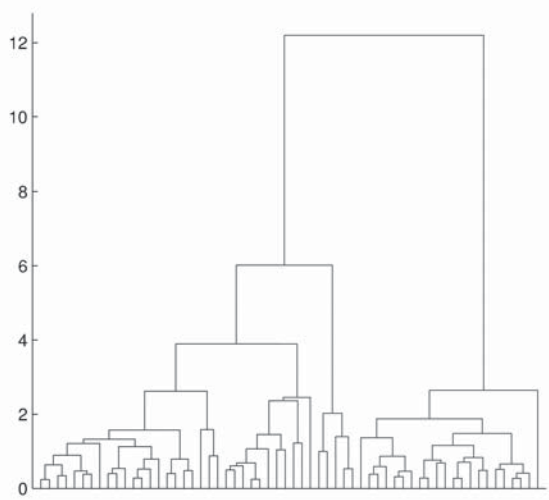

Fig. 5: Dendrogram of the models. λ = 1:the bad fitted models are hidden in the two cluster structure.

parameters and the information about the model fitting has been proposed. It has the advantage to consider not only the trend but also the actual explicative power

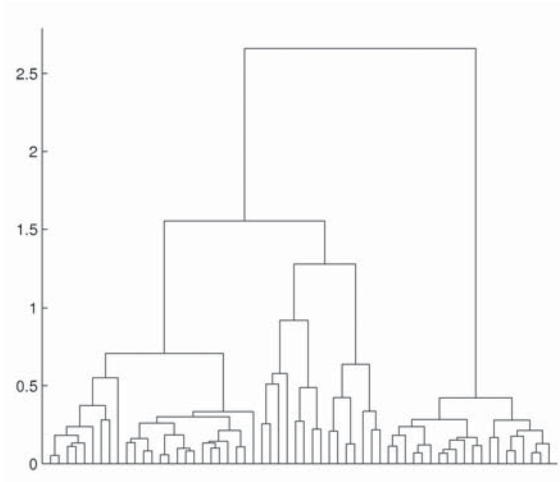

Fig. 6: Dendrogram of the models. $\lambda = 0.2$:the cophenetic coecient is maximum, a three cluster structure is more evident.

of the models. Simulations results seem to support the proposed strategy. As further research real applicative fields need to be explored.

REFERENCES

GREEN P.E., SRINIVASAN V. (1990), Conjoint Analysis in marketing: New Developments With Implication for Research and Practice, *Journal of Marketing*, 25: 3–19.

GUSTAFSSON A., HERMANN A., HUBER F. (2000), Conjoint Measurement. Methods and Application. Berlin, Heidelberg: *Springer Verlag*.

HENNING C. (2000), Identifiability of Models for Clusterwise Linear Regression, *Journal of Classification*, 17: 273–296.

LAURO C., SCEPI G., GIORDANO G., (2002) Cluster Based Conjoint Analysis, in *Proceedings of Sixth International Conference on Social Science Methodology*, RC Logic & 33 Methodology, August 17-20, 2004 Amsterdam.

PLAIA A. (2003), Constrained Clusterwise Linear Regression, in *New Developments in Classification and Data Analysis*, M. Vichi P. Monari, S. Mignani, S. Montanarini eds., *Springer*, Bologna 79–86.

SPAETH H. (1979), Clusterwise Linear Regression, *Computing*, 22: 367–373.

TAKANE Y., DE LEW J., YOUNG F.W. (1990), Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features, *Psycometrika*, 41(4): 505–529.

VRIENS M., WEDEL M. and WILMS T. (1996), Metric Conjoint Segmentation Methods: a Monte Carlo comparison, *Journal of Marketing Research*, 33: 73–85.

CLASSIFICAZIONE DI FUNZIONI DI UTILITÀ ATTRAVERSO UNA DISTANZA TRA MODELLI

Riassunto

La Conjoint Analysis è una delle tecniche maggiormente utilizzate nella valutazione del comportamento dei consumatori. Questa metodologia consente di stimare i coefficienti di utilità parziale in base ad un modello statistico che lega la valutazione globale di preferenza alle caratteristiche descrittive degli stimoli (prodotti o servizi). I risultati della Conjoint Analysis trovano vasta applicazione nella segmentazione del mercato.

In questo lavoro viene proposta una strategia di classificazione basata su una nuova metrica. La distanza introdotta è definita come combinazione convessa di due distanze. Essa consente di tener conto di una duplice qualità dell'informazione relativa al modello: il valore dei coefficienti stimati e la bontà di adattamento. Di conseguenza, la differenziazione tra segmenti di mercato è ottenuta considerando la prossimità dei modelli di utilità individuali stimati e la capacità predittiva degli stessi.