

**Lauro, N. Carlo; Verde, R.; Palumbo, F.**

**Factorial discriminant analysis on symbolic objects.** (English) Zbl 0977.62070

Bock, Hans-Hermann (ed.) et al., Analysis of symbolic data. Exploratory methods for extracting statistical information from complex data. Berlin: Springer. Studies in Classification, Data Analysis, and Knowledge Organization. 212-233 (2000).

Discriminant analysis refers to a set of techniques aiming at describing the relations between a set of  $p$  quantitative variables and a categorical variable with  $m$  labels, i.e., a classificatory variable that defines a partition of the population under study into  $m$  classes. It consists of two main aspects: (i) selection of the best sub-set of the original predictors (selection aspect); (ii) construction of a decision rule (classification rule) in order to classify statistical units into one of the  $m$  classes (classification aspect).

A generalization of Factorial Discriminant Analysis (FDA) to symbolic objects (SOs) is proposed which consists in a symbolic-numerical-symbolic procedure. It is based on a numerical analysis of transformed SOs and a symbolic interpretation of its results. Similarly as in FDA, two sets of SOs for evaluation purposes are considered: a training set and a test set. SO-FDA proceeds first by a numerical transformation of the SOs' descriptors, consisting of a suitable coding which depends on the type of variables, and of an optimal quantification of the coded categorical variables by means of a nonsymmetrical factorial approach. Then nominal variables are coded according to the classical disjunctive coding system, whereas numerical categorized variables are coded according to a fuzzy coding system. Furthermore, it is proposed to codify modal variables also in a fuzzy way assuming as coding values the weights associated to each of their categories. In order to allow for logical relationships among the descriptors, a mathematical structure is created by the Cartesian product of the coding values associated with each variable. In this way, we perform a correspondence between the numerical and the geometrical description of a SO: each combination of the descriptor values is interpreted as the coordinate vector of a vertex of the hypercube that visualizes the SO.

In the FDA framework for qualitative predictors, the quantification step is achieved by carrying out a Multiple Correspondence Analysis (MCA). The geometrical results of SO-FDA are visualized in the factorial plane by maximum covering area rectangles. Analogously, the classes are visualized as rectangles including the SOs belonging to the class. For classification purposes two geometrical classification rules are defined on the basis of the proximities among classes represented on the factorial plane by their elements, and each SO of the test set. In particular, it is considered a generalized Minkowski metric and a proximity measure based on the minimum description potential increase. The similarity between pairs of symbolic objects of the training set are evaluated on the basis of their size and shape and respective positions.

For the entire collection see [\[Zbl 1039.62501\]](#).

Reviewer: [M.P.Moklyachuk \(Kyiv\)](#)

**MSC:**

[62H30](#) Classification and discrimination; cluster analysis (statistical aspects)  
[62H25](#) Factor analysis and principal components; correspondence analysis

Cited in **10** Documents

**Keywords:**

[symbolic objects](#); [discriminant analysis](#); [fuzzy coding system](#)