

# Hierarchical Mixed Topological Maps

Ndèye Niang\*, Mory Ouattara \*,\*\*

\*Laboratoire CEDRIC, CNAM 292, rue Saint Martin, 75141 Paris Cedex 03, France,  
n-deye.niang\_keita@cnam.fr,  
<http://cedric.cnam.fr>

\*\*Centre Scientifique et Technique du Bâtiment, 84 Avenue Jean Jaurès, 77420  
Champs-sur-Marne  
mory.ouattara@cstb.fr,  
<http://cedric.cnam.fr>

**Abstract.** We address the problem of clustering individuals described with several mixed variables divided in homogeneous blocks. We propose a hierarchical method with two levels to partition the individuals. The method is based on two successive steps using mixed topological maps combined with agglomerative hierarchical clustering. The proposed approach allows to take into account simultaneously qualitative and quantitative variables as well as the variable blocking. A real example on indoor air quality illustrates the proposed method.

## 1 Introduction

Clustering analysis is probably one of the most widely used statistical tools in data mining. In various applications in engineering, biology, business and social sciences classical clustering methods as well as new approaches are more and more used to reduce huge transactional and experimental data. These data are often column partitioned, as the variables are divided in several homogeneous and meaningful blocks. For example in biology or chemistry experimental analysis, a data block may consist of a set of measurement variables which refer to the same type of instrument or method used for the analysis; alternatively, a block may contain variables having biological similarity. In the indoor air quality study which illustrates the method we propose, several questionnaires related to different possible causes of pollution have been fulfilled separately. The resulting data sets contains a combination of categorical and continuous variables, we refer to them as mixed data sets.

The general problem addressed in this paper is discovering unknown homogenous groups of observations keeping into account the multiblock structure of the data. We propose a hierarchical approach with two levels of clustering using agglomerative hierarchical clustering combined with mixed topological maps (*MTM*) (Lebbah et al., 2005), a modified version of self organizing map (*SOM*) (Kohonen, 1982, 1995, 1998) for mixed variables. The first step clusters each initial block of variables separately providing local partitions (ie according to each block). In the second step, based on the results of the former one, the different local partitions are combined into a single global one. The aim of this approach, called hierarchical mixed topological map (*HMTM*) is interpreting block-specific patterns of heterogeneity