

Hierarchical clustering of modal ordinal symbolic data objects

Carmen Bravo*, José M. García-Santesmases**

* Universidad Complutense de Madrid, Servicio Informático de Apoyo al Usuario
- Investigación, Edificio Vicerrectorado de Alumnos, 28040 Madrid, Spain.
mcbravo@ucm.es

** Universidad Complutense de Madrid, Dpto. de Estadística e Inv. Operativa
Facultad de Ciencias Matemáticas, 28040 Madrid, Spain. josemgar@ucm.es

Abstract. The problem of analyzing the dispersion of a set of objects described by ordinal modal symbolic data is addressed in order to obtain homogeneous groups, which are evaluated by a consensus measure. Based on a generalized φ function a consensus measure for objects and for sets of objects described by modal ordinal data is defined. A variability measure for sets of subsets of objects based in the consensus measure of their members is proposed. A dissimilarity measure between objects and between set of objects based on this consensus variability measure is also given. It is proven that the Leik consensus measure is a φ function. An ascending hierarchical clustering algorithm is presented. The criterion to be minimized in each step is based on the decrease of the consensus variability. An example with modal ordinal data of 34 teachers that were evaluated by their students is presented.

Introduction

This paper proposes an ascending hierarchical clustering algorithm for modal ordinal symbolic data using a dissimilarity measure based on consensus variability. One common meaning of consensus is a general agreement among the members of a given group and can be seen as a function of shared team feelings towards an issue. A common way to analyse it is to use consensus measures to evaluate the strength of consensus in a class of individuals. As introduced by Leik (1966) the conception of consensus is simply a lack of dispersion, and a consensus measure provides a way to measure the dispersion in ordinal scales. In García-Santesmases and Bravo (2010) and García-Santesmases et al. (2010), three specific consensus measures for groups of individuals based on a single issue are given and are proved that satisfy the requirements given by Tastle (2005). They are extended to several issues and to symbolic data. In this paper our main contribution is to give a characterization of a consensus measure for a group of individuals based on a single issue through the introduction of a φ function with some properties. This characterization covers all the requirements given by Tastle (2005) for a consensus measure to be considered viable. It can be used to build consensus measures for symbolic data objects and sets, variability consensus measures and distances based on these variabilities. The