

Title:

Variable Selection for Latent Class Analysis

Abstract:

A stepwise variable selection method for latent class cluster analysis of multivariate categorical data is developed. Given a set of already selected clustering variables, the usefulness of a variable is assessed comparing two models. In one model the variable adds further information about the clusters, and in the other model it does not but it can be related to the clustering variables. A backward-stepwise greedy algorithm is used to conduct the search over the model space and the comparison between models is performed using BIC. The method is demonstrated on simulated data, where it selects the correct clustering variables. Furthermore an application on real data related to low back pain disorders is considered and our method recovers the true group structure with a small number of variables. This is joint work with Michael Fop.